



2019.12.2-12.6, Kenting, Taiwan

27th International Conference on Computers in Education

Conference Proceedings Volume 1



27th International Conference on Computers in Education
Conference Proceedings Volume 1

Editors: Maiga CHANG, Hyo-Jeong SO, Lung-Hsiang WONG, Fu-Yun YU, Ju-Ling SHIH

Publisher: Asia-Pacific Society for Computers in Education (APSCE)

Address: Center for Science and Technology for Learning,
National Central University,
No.300, Jhongda Road, Jhongli City,
Taoyuan County 32001, Taiwan

Telephone: +886-3-4227151 ext. 35407

FAX: +886-3-4261931

Email: service@apsce.net

Website: <http://www.apsce.net>

Date of publication:

ISBN:

Copyright 2019 Asia-Pacific Society for Computers in Education.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, transmitted, in any forms or any means, without the prior permission of the Asia-Pacific Society for Computers in Education. Individual papers may be uploaded on to institutional repositories or other academic sites for self-archival purposes.

- TABLE OF CONTENTS -

EDITORS	I
ASSOCIATE EDITORS	II
MESSAGE FROM THE CONFERENCE CHAIR.....	III
MESSAGE FROM THE LOCAL ORGANIZING COMMITTEE CHAIRS.....	V
MESSAGE FROM THE INTERNATIONAL PROGRAM COORDINATION CHAIRS.....	VI
ORGANIZATION	IX
APSCE FELLOWS PROGRAM	XIX
EARLY CAREER RESEARCHER AWARD WINNER (2019)	XXIV
LAST TEN YEARS' DISTINGUISHED RESEARCHER AWARD WINNERS	XXV
LAST FIVE YEARS' EARLY CAREER RESEARCHER AWARD WINNERS	XXVI
KEYNOTE SPEAKERS.....	XXVII
THEME-BASED INVITED SPEAKERS.....	XXXII
PANELS	1
MAIN CONFERENCE C1 - ARTIFICIAL INTELLIGENCE IN EDUCATION/INTELLIGENT TUTORING SYSTEM (AIED/ITS) AND ADAPTIVE LEARNING	11
MAIN CONFERENCE C2 - COMPUTER-SUPPORTED COLLABORATIVE LEARNING (CSCL) AND LEARNING SCIENCES..	99
MAIN CONFERENCE C3 - ADVANCED LEARNING TECHNOLOGIES (ALT) LEARNING ANALYTICS AND DIGITAL INFRASTRUCTURE	212
MAIN CONFERENCE C4 - CLASSROOM, UBIQUITOUS, AND MOBILE TECHNOLOGIES ENHANCED LEARNING (CUMTEL).....	375
MAIN CONFERENCE C5 - EDUCATIONAL GAMIFICATION AND GAME-BASED LEARNING (EGG)	493
MAIN CONFERENCE C6 - TECHNOLOGY ENHANCED LANGUAGE LEARNING (TELL)	606
MAIN CONFERENCE C7 - PRACTICE-DRIVEN RESEARCH, TEACHER PROFESSIONAL DEVELOPMENT AND POLICY OF ICT IN EDUCATION (PTP)	709
AUTHOR INDEX.....	780

EDITORS

Maiga CHANG, Athabasca University, Canada

Hyo-Jeong SO, Eawha Womans University, South Korea

Lung-Hsiang WONG, Nanyang Technological University, Singapore

Fu-Yun YU, National Cheng Kung University, Taiwan

Ju-Ling SHIH, National Central University, Taiwan

ASSOCIATE EDITORS

(in alphabetical order)

Ivica BOTIČKI, University of Zagreb, Croatia
Ming-Puu CHEN, National Taiwan Normal University, Taiwan
Ali DEWAN, Athabasca University, Canada
Stian HAKLEV, École Polytechnique Fédérale de Lausanne, Switzerland
Elizabeth KOH, Nanyang Technological University, Singapore
Tomoko KOJIRI, Kansai University, Japan
Kuo-Chen LI, Chung-Yuan Christian University, Taiwan
Daner SUN, The Education University of Hong Kong, Hong Kong
Yun WEN, Nanyang Technological University, Singapore

MESSAGE FROM THE CONFERENCE CHAIR



Lung-Hsiang WONG

Conference Chair
Nanyang Technological University
Singapore

On behalf of the organizing committee, I would like to extend my warm welcome to all delegates of the 27th International Conference on Computers in Education (ICCE) 2019, the flagship conference series of the Asia-Pacific Society for Computers in Education (APSCE). ICCE is no stranger to Taiwan, given that this conference series (a biennial event in its first three editions) originated from Taipei, Taiwan back in 1989—exactly three decades ago. Since then, Taiwan has hosted five more ICCE editions including the one this year, the first time it is being held in Kenting, the southern tip of the island. The conference theme this year, “Intelligent Computing and Convergent Education”, signifies the importance of striking a concerted and harmonious relationship between learners and the technology, where learners may proactively construct knowledge networks and social networks of their own with the mediation of intelligent computing, which further interconnect with each other and build up as a whole intellectual universe.

Four outstanding keynote speakers will share their insights across varying areas in the field of computers in education. They are (1) Mina C. JOHNSON from Arizona State University, USA, who will address the design issues of VR, Augmented and Mixed Realities for STEM and game-based learning; (2) Hiroaki OGATA from Kyoto University, Japan, who will explicate how evidence-based education can be accomplished through learning analytics; (3) J. M. Joke VOOGT from University of Amsterdam, the Netherlands, who will offer alternative views on the how to prepare learners for living and working in the digital age; and (4) Stephen J.-H. YANG from National Central University, Taiwan, who will share with us his insights on the new challenges of precision education imposed on the research on AI in education.

Furthermore, there are three equally inspiring theme-based invited speeches – (1) “Creating conditions for Knowledge Building for the public good” by Bodong CHEN from University of Minnesota, USA; (2) “Embrace data intelligence in designing for the e-Schoolbag: Addressing the 1:1 challenges” by Xiaoqing GU from East China Normal University, China; and (3) “Flipped-classroom course model with ICT support to activate discussion in foreign language classrooms” by Yuichi ONO from University of Tsukuba, Japan.

In one way or another, these presentations connect with the essence of the aforesaid conference theme and will stimulate us to reflect upon the roles of learning theories, pedagogical strategies, and technological tools in reshaping the education landscape against the backdrop of knowledge society.

Indeed, organizing such a large-scale conference requires the orchestrated efforts and

steadfast support from the conference organizing committee members and conference paper reviewers. To recognize these noteworthy individuals, their names are listed in the proceedings. I would also like to express my heartfelt appreciation to all the kind individuals who have rendered their help in every possible way to make this conference a reality. I am also grateful to all the paper authors and registered participants for their exciting academic contributions to the fruitful intellectual exchange in this conference. I hope all delegates will have opportunities to renew friendships, forge new friendships and professional collaborations. I hope that you will have a fruitful and fun-filled time at this very special conference and leave Kenting—the ocean paradise of Formosa—with magnificent, affectionate memories.

Thank you!

MESSAGE FROM THE LOCAL ORGANIZING COMMITTEE CHAIRS



Fu-Yun YU

Local Organizing Committee Chair
National Cheng Kung University
Taiwan

“有朋自遠方來，不亦樂乎!” This is a famous quote by Confucius, which means it is such a delight to have friends come from afar! Thank you for being here, and welcome to Taiwan!

It is our great privilege to host the 27th ICCE here at Kenting, Taiwan. At the south tip of Taiwan, Kenting National Park offers picturesque views and rich natural resources of both mountains and oceans. It reminds us that, as intelligent computing leads our way to the era of smart city and smart life, global sustainability goals are to be reached through convergent education wherein digital data, social networks, human activities and natures should be interconnected in a meaningful way. The conference theme this year, “Intelligent Computing and Convergent Education”, is designed to capture the harmonious spirit and goals of a balanced life.



Ju-Ling SHIH

Local Organizing Committee Chair
National Central University
Taiwan

This is in line with ICCE’s 30-year tradition of well-structured programs, high-quality technical papers, versatile technical and social activities, and most important of all, amazing people! Please join us in embracing this fascinating, 5-day event full of the fruitful exchanging of academic research work, making of new friends, refreshing memories with old colleagues, building strong research communities, and envisioning inspirational and innovative research.

We thank the APSCE Executive Committee for entrusting us with this important event. Our heartfelt thanks goes out to our sponsors, mainly, Taiwan’s Ministry of Science and Technology and National Cheng Kung University, and Conference Chair, Prof. Lung-Hsiang WONG, who has been with us through every step of the planning and executing stages. Also, we thank the

Committee, the International Program Committee, the 280 registered conference participants, colleagues and student volunteers from different affiliations and universities who have joined our Local Organizing Committee to help ensure a smooth execution and sound presentation of ICCE to our honorable guests.

We trust all of you will enjoy the conference, and take home with you fond memories from this trip to Taiwan.

謝謝！Thank you!

MESSAGE FROM THE INTERNATIONAL PROGRAM COORDINATION CHAIRS



Maiga CHANG

International Program Coordination
Chair
Athabasca University
Canada



Hyo-Jeong SO

International Program Coordination
Co-Chair
Eawha Womans University
South Korea

The International Conference on Computers in Education (ICCE) is an annual conference series encompassing a broad range of issues related to using Information and Communication Technology (ICT) for education, organized by the Asia-Pacific Society for Computers in Education (APSCE). ICCE 2019 takes place in Kenting, Taiwan, from December 2nd to 6th, 2019. It aims to bring together researchers from all over the world to share and exchange research and to develop and deploy new ideas that span the field of Computers in Education. Following the tradition of previous conferences in this series, ICCE 2019 is organized as a meta-conference, where there are seven Sub-Conferences, each of which focuses on specialized themes. Each Sub-Conference is organized by a program committee appointed by the respective Special Interest Group (SIG – see <https://apsce.net/sig.php>). These Sub-Conferences are:

- C1: ICCE Sub-Conference on Artificial Intelligence in Education/Intelligent Tutoring System (AIED/ITS)
- C2: ICCE Sub-Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences (LS)
- C3: ICCE Sub-Conference on Advanced Learning Technologies (ALT), Learning Analytics, Platforms and Infrastructure
- C4: ICCE Sub-Conference on Classroom, Ubiquitous, and Mobile Technologies Enhanced Learning (CUMTEL)
- C5: ICCE Sub-Conference on Educational Gamification and Game-based Learning (EGG)
- C6: ICCE Sub-Conference on Technology Enhanced Language Learning (TELL)
- C7: ICCE Sub-Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

The International Program Committee is led by a strong and dedicated team, which includes the Conference Chair, the Program Coordination Chair and Co-Chair, seven executive Sub-Conference Chairs and 288 experts in the field of Computers in Education from 35 different countries or economies. Former ICCE local organizing and program coordination chairs have played as consultants in overseeing the organization process of this conference. The conference received a total of 200 papers (142 full, 46 short, and 12 posters) from 24 different countries or economies. Table 1 provides the submissions by the country of the first

author in each paper.

Table 1. Submission Statistics (based on first author's country)

Countries or Economies			
Australia	4	New Zealand	1
Canada	2	Philippines	15
China	45	Singapore	10
Croatia	1	South Korea	1
Germany	6	Spain	1
Hong Kong	9	Sweden	3
India	11	Taiwan	22
Indonesia	7	Thailand	9
Japan	41	Turkey	3
Malaysia	6	United States	3

All papers were subjected to a rigorous review process by 3 to 5 reviewers from the respective Sub-Conference program committees. After the reviews were completed, a meta-review was provided for each paper. In total, 810 reviews and meta-reviews were received. After a discussion period within the individual program committees led by the Sub-Conference Executive Chairs and Co-Chairs, recommendations were made to the Program Coordination Committee Chair and Co-Chair, who oversaw the review process and quality for all Sub-Conferences. This resulted in 32 full, 68 short, and 49 poster acceptances across all of the seven Sub-Conferences. The overall acceptance rate for full papers is 22.54%. The acceptance rate for the full papers in the individual Sub-Conference closely mirrored the overall acceptance rate. This is a testimony to the continued maintenance of the quality of presentations in our conference. The complete statistics of paper acceptance is shown in Table 2.

Table 2. Paper Acceptance Statistics

	submission	submit as Full	Full	Full %	Short	Poster	Overall %
C1 - AIED/ITS	27	22	4	18.18%	7	8	70.37%
C2 - CSCL/LS	31	19	4	21.05%	9	10	74.19%
C3 - ALT/LA/PI	38	26	7	26.92%	11	10	73.68%
C4 - CUMTEL	26	21	5	23.81%	10	5	76.92%
C5 - EGG	25	19	4	21.05%	12	3	76.00%
C6 - TELL	25	16	4	25.00%	11	4	76.00%
C7 - PTP	28	19	4	21.05%	8	9	75.00%
ICCE 2019	200	142	32	22.54%	68	49	74.50%

In addition to full papers, short papers and posters, ICCE 2019 includes various program components, such as Keynote Speeches, Theme-based Invited Speeches, Workshops, Tutorials, Panels, Work-in-Progress Posters (WIPP), Extended Summary (ES), Doctoral Student

Consortia (DSC), and Early Career Workshop (ECW). All the papers in these program components are published in separate proceedings with their own ISBN numbers. Pre-conference events are held on the first two days of the conference, including 16 workshops, 4 tutorials, DSC, ECW, and APSCE Student Wing Workshop.

We would like to thank all who have contributed to making ICCE 2019 a successful conference. First of all, we would like to thank all paper authors for your contributions and for choosing ICCE 2019 as an outlet to present your research. We would also like to thank the IPC Executive Chairs/Co-Chairs and members, who undertook the responsibility of reviewing and selecting papers that represent research of high quality. Specially thanks to our Keynote and Invited Speakers for accepting our invitations and bring inspiring research to ICCE 2019 participants. The Local Organizing Committee deserves a big thank you for their hard work under the tremendous time pressure.

We hope that all participants will find the activities in ICCE 2019 interesting and inspiring, and have opportunities to meet old friends and establish new professional collaborations. Furthermore, we hope that participants will enjoy not only the academic activities, but also the vibrant and exciting culture experience in Kenting.

ORGANIZATION

Organized by: Asia Pacific Society for Computers in Education

Hosted by: National Cheng Kung University, Taiwan
National Central University, Taiwan

Standing Committee

- *Conference Chair*
Lung Hsiang WONG, Nanyang Technological University, Singapore
- *International Program Coordination Chair*
Maiga CHANG, Athabasca University, Canada
- *International Program Coordination Co-Chair*
Hyo-Jeong SO, Eawha Womans University, South Korea
- *Local Organizing Committee Chairs*
Fu-Yun YU, National Cheng Kung University, Taiwan
Juling SHIH, National Central University, Taiwan
- *Consultants*
Tanja MITROVIC, University of Canterbury, New Zealand
Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines
Jie Chi YANG, National Central University, Taiwan

Sub Conferences

C1: Sub-Conference on Artificial Intelligence in Education/Intelligent Tutoring Systems (AIED/ITS)

- *PC Executive Chair:*
Tomoko KOJIRI, Kansai University, Japan
- *PC Executive Co-Chairs:*
Roger NKAMBOU, Universite du Quebec a Montreal, Canada
Chih-Yueh CHOU, Yuan Ze University, Taiwan
Nguyen-Thanh LE, Humboldt-Universitat zu Berlin, Germany

C2: Sub-Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences

- *PC Executive Chair:*
Elizabeth KOH, Nanyang Technological University, Singapore
- *PC Executive Co-Chairs:*
Daniel BODEMER, University of Duisburg-Essen, Germany
Kate THOMPSON, Griffith University, Australia
Camillia MATUK, New York University, USA
Sahana MURTHY, IIT Bombay, India

C3: Sub-Conference on Advanced Learning Technologies, Learning Analytics, Platforms and Infrastructure

- *PC Executive Chair:*
Stian HAKLEV, École Polytechnique Fédérale de Lausanne, Switzerland
- *PC Executive Co-Chairs:*
Brendan FLANAGAN, Kyoto University, Japan
Elle WANG, Arizona State University, USA
Huang-Yao HONG, National Chengchi University, Taiwan

C4: Sub-Conference on Classroom, Ubiquitous and Mobile Technologies Enhanced Learning (CUMTEL)

- *PC Executive Chair:*
Daner SUN, The Education University of Hong Kong, Hong Kong
- *PC Executive Co-Chairs:*
Ying ZHAN, The Education University of Hong Kong, Hong Kong
Alice CHIK, Macquarie University, Australia
Fathi ESSALMI, University of Kairouan, Tunisia
Longkai WU, Nanyang Technological University, Singapore

C5: Sub-Conference on Educational Gamification and Game-based Learning (EGG)

- *PC Executive Chair:*
Ming-Puu CHEN, National Taiwan Normal University, Taiwan
- *PC Executive Co-Chairs:*
George GHINEA, Brunel University, UK
Morris S. Y. JONG, The Chinese University of Hong Kong, Hong Kong
Rita KUO, New Mexico Tech, USA

C6: Sub-Conference on Technology Enhanced Language Learning (TELL)

- *PC Executive Chair:*
Yun WEN, Nanyang Technological University, Singapore
- *PC Executive Co-Chairs:*
Kousuke MOURI, Tokyo university of Agriculture and technology, Japan
Qing MA, The Education University of Hong Kong, Hong Kong
Weichao CHEN, University of Virginia, USA
Chunping ZHENG, Beijing University of Posts and Telecommunications, China

C7: Sub-Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

- *PC Executive Chair:*
Ivica BOTIČKI, University of Zagreb, Croatia
- *PC Executive Co-Chairs:*
Marc JANSEN, Hochschule Ruhr West, Germany
Peter SEOW, Nanyang Technological University, Singapore
Rwitajit MAJUMDAR, Kyoto University, Japan
Gokhan AKCAPINAR, Hacettepe University, Turkey

Workshop & Interactive Events

- *Chair:*
Niwat SRISAWASDI, Khon Kaen University, Thailand
- *Co-Chairs:*
Patcharin PANJABUREE, Mahidol University, Thailand
Michelle P. BANAWAN, Ateneo de Davao University, Philippines

Tutorials

- *Chair:*
Ben CHANG, National Central University, Taiwan
- *Co-Chairs:*
Ahmed TLILI, Smart Learning Institute of Beijing Normal University, China
Calvin LIAO, National Taipei University of Nursing and Health Sciences, Taiwan

Posters

- *Chair:*
Andrei D. CORONEL, Ateneo de Manila University, Philippines
- *Co-Chairs:*
Swapna GOTTIPATI, Singapore Management University, Singapore
Fan OUYANG, Zhejiang University, China
Michelle P. BANAWAN, Ateneo de Davao University, Philippines

Doctoral Student Consortium (DSC)

- *Chair:*
Weiqin CHEN, Oslo Metropolitan University, Norway
- *Co-Chair:*
H. Ulrich HOPPE, University of Duisburg-Essen, Germany

Early Career Workshop (ECW)

- *Chair:*
Morris S.Y. JONG, The Chinese University of Hong Kong, Hong Kong
- *Co-Chairs:*
Jon MASON, Charles Darwin University, Australia
Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines

Panels

- *Chair:*
Kuo-Chen LI, Chung-Yuan Christian University, Taiwan
- *Co-Chair:*
Ali DEWAN, Athabasca University, Canada

Extended Summaries (ES)

- *Chair:*
Yanjie SONG, The Education University of Hong Kong, Hong Kong
- *Co-Chair:*
Chengjiu YIN, Kobe University, Japan

Merit Scholarships

- *Chair:*
Bo JIANG, Zhejiang University of Technology, China
- *Co-Chair:*
Madathil Warriem JAYAKRISHNAN, Indian Institute of Technology Madras, India

Special Interest Groups (SIGs)

- *S1: Artificial Intelligence in Education/Intelligent Tutoring Systems/Adaptive Learning (AIED/ITS/AL)*
Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines
- *S2: Computer-supported Collaborative Learning (CSCL) and Learning Sciences*
Chew Lee TEO, Nanyang Technological University, Singapore
- *S3: Advanced Learning Technologies (ALT), Open Contents, and Standards*
Jon MASON, Charles Darwin University, Australia
- *S4: Classroom, Ubiquitous and Mobile Technologies Enhanced Learning (CUMTEL)*
Chengjiu YIN, Kobe University, Japan
- *S5: Educational Gamification and Game-based Learning (EGG)*
Morris S.Y. JONG, The Chinese University of Hong Kong, Hong Kong
- *S6: Technology Enhanced Language Learning (TELL)*
Ting-Chia HSU, National Taiwan Normal University, Taiwan
- *S7: Practice-driven Research, Teacher Professional Development, and Policy of ICT in Education (PTP)*
Sridhar IYER, Indian Institute of Technology Bombay, India
- *S8: Development of Information and Communication Technology in the Asia-Pacific Neighborhood (DICTAP)*
Niwat SRISAWASDI, Khon Kaen University, Thailand
- *S9: Educational Use of Problems/Questions in Technology-Enhanced Learning*
Tomoko KOJIRI, Kansai University, Japan
- *S10: Learning Analytics and Educational Data Mining*
H. Ulrich HOPPE, University of Duisburg-Essen, Germany

Local Organizing Committee

- *Chairs*
Fu-Yun YU, National Cheng Kung University, Taiwan
Juling SHIH, National Central University, Taiwan
- *Coordinator:* Chung-I LIN, National Cheng Kung University, Taiwan
- *Finance:* Hsin Chih LIN, National University of Tainan, Taiwan
- *Publicity:*
Tsung-Yen CHUANG, National University of Tainan, Taiwan
Chun-Ming SU, National University of Tainan, Taiwan
- *Proceedings:* Wan-Lin YANG, National Cheng Kung University, Taiwan
- *Webmaster:*
Hao-Chiang Koong LIN, National University of Tainan, Taiwan
Wen-lin KUO, National University of Tainan, Taiwan
- *Social Media Coordinator:*
Chun-Ping WU, National University of Tainan, Taiwan
- *Technical:* I-Wen HUANG, National University of Tainan, Taiwan
- *Registration:*
Chi-Syan LIN, National University of Tainan, Taiwan
Yen-Tzu CHEN, National University of Tainan, Taiwan
- *Social Events:* Chien-I LEE, National University of Tainan, Taiwan
- *Sponsorship:* Wen-Chi WU, Asia University, Taiwan
- *Volunteers:* Yi-Chun LIU, Chia Nan University of Pharmacy & Science, Taiwan
- *Local Arrangement:*

Chih-Kai CHANG, National University of Tainan, Taiwan

Shu-Hsien HUANG, National Chin-Yi University of Technology

C1 PC Members

- Sagaya Amalathas, Taylors University, Malaysia
- Ryan Baker, University of Pennsylvania, USA
- Michelle Banawan, ateneo de davao university, Phillippine
- Kritya Bunchongchit, Mahidol University International College, Thailand
- Mario Chacón Rivas, Tecnológico de Costa Rica, Costa Rica
- Ben Daniel, University of Otago, New Zealand
- Benedict Du Boulay, Informatics Department, University of Sussex, UK
- Claude Frasson, University of Montreal, Canada
- Shinobu Hasegawa, JAIST, Japan
- Yusuke Hayashi, Graduate School of Engineering, Hiroshima university, Japan
- Bastiaan Heeren, Open University, The Netherlands, Netherlands
- Tomoya Horiguchi, Kobe University, Japan
- Sharon Hsiao, Arizona State University, USA
- Akihiro Kashiara, The University of Electro-Communications, Japan
- Tatsuhiro Konishi, Shizuoka University, Japan
- Noboru Matsuda, North Carolina State University, USA
- Tatsunori Matsui, Waseda University, Japan
- Tanja Mitrovic, Intelligent Computer Tutoring Group, University of Canterbury, Christchurch, New Zealand
- Kazuhisa Miwa, Nagoya University, Japan
- Riichiro Mizoguchi, Japan Advanced Institute of Science and Technology, Japan
- Helen Pain, The University of Edinburgh, UK
- Jose-Luis Perez De La Cruz, Universidad de Malaga, Spain
- Elvira Popescu, University of Craiova, Romania
- Ma. Mercedes T. Rodrigo, Department of Information Systems and Computer Science, Ateneo de Manila University, Phillippine
- Olga C. Santos, aDeNu Research Group (UNED), Spain
- Kazuhisa Seta, Osaka Prefecture University, Japan
- Thepchai Supnithi, NECTEC, Thailand

C2 PC Members

- Sakinah Alhadad, Griffith University, Australia
- Maaik Bouwmeester, New York University, USA
- Jürgen Buder, Leibniz-Institut für Wissensmedien, Germany
- Sanjay Chandrasekharan, Homi Bhabha Centre for Science Education, TIFR, India
- Bodong Chen, University of Minnesota, USA
- Gaowei Chen, The University of Hong Kong, Hong Kong
- Andrew Clayphan, The University of Sydney, Australia
- Mutlu Cukurova, University College London, UK
- Chandan Dasgupta, Indian Institute of Technology Bombay, India
- Pryce Davis, University of Nottingham, UK
- Amanda Dickes, Harvard University, USA
- Julia Eberle, Ruhr-Universität Bochum, Germany
- Heisawn Jeong, Hallym University, South Korea
- Beaumie Kim, University of Calgary, Canada

- Simon Knight, University of Technology Sydney, Australia
- Rose Liang, National University of Singapore, Singapore
- Pei-Yi Lin, National Cheng Chi University, Taiwan
- Chen-Chung Liu, National Central University, Taiwan
- Chee Kit Looi, Nanyang Technological University, Singapore
- Michelle Lui, University of Toronto, Canada
- Roberto Martinez-Maldonado, University of Technology Sydney, Australia
- Johanna Pöysä-Tarhonen, University of Jyväskylä, Finland
- Peter Reimann, Univ. of Sydney, Australia
- Lay Hoon Seah, Nanyang Technological University, Singapore
- Kristen Searle, Utah State University, USA
- Pratim Sengupta, University of Calgary, Canada
- Hillary Swanson, Northwestern University, USA
- Esther Tan, Delft University of Technology, Netherlands
- Seng Chee Tan, Nanyang Technological University, Singapore
- Chewlee Teo, MOE, Singapore
- Mike Tissenbaum, University of Illinois, USA
- Ralph Vacca, Fordham University, USA
- Qiyun Wang, Nanyang Technological University, Singapore
- Hongzhi Yang, The University of Sydney, Australia

C3 PC Members

- Nurbiha A Shukor, Universiti Teknologi Malaysia, Malaysia
- Luis Anido Rifon, Universidade de Vigo. Depto. of Telematics, Spain
- Judith Azcarraga, De La Salle University Manila, Philippines
- Manuel Caeiro Rodríguez, University of Vigo, Spain
- Yang-Hsueh Chen, National Chengchi University, Taiwan, Taiwan
- Yong-Sang Cho, Korea Education & Research Information Service, Korea
- Jean-Noel Colin, Faculty of Computer Science/FUNDP, Belgium
- Jie Geng, The Chinese University of Hong Kong, Hong Kong
- Swapna Gottipati, Singapore Management University, Singapore
- Tore Hoel, Høgskolen i Oslo og Akershus, Norway
- Wafa Johal, CHILI/LSRO Lab - EPFL, Switzerland
- Khalid Khan, Charles Darwin University, Australia
- Dan Kohen, Linnaeus University, Sweden
- Lam-For Kwok, City University of Hong Kong, Hong Kong
- Yongwu Miao, University Duisburg-Essen, Germany
- Hiroshi Nakano, Center for Management of IT, Kumamoto University, Japan
- Jennifer Olsen, Ecole Polytechnique Fédérale de Lausanne, Switzerland
- Nigel Stanger, University of Otago, New Zealand
- Jerry Chih-Yuan Sun, National Chiao Tung University, Taiwan
- Chew Lee Teo, Nanyang Technological University, Singapore
- Riina Vuorikari, Institute for Prospective Technological Studies (IPTs), Spain
- Minhong Wang, The University of Hong Kong, Hong Kong
- Tsuneo Yamada, The Open University of Japan, Japan

C4 PC Members

- Ivana Bosnic, University of Zagreb, Croatia
- Ivica Boticki, University of Zagreb, Croatia
- Huiying Cai, Jiangnan University, China
- Su Cai, Beijing Normal University, China
- Ben Chang, National Central University, Taiwan
- Chi-Cheng Chang, National Taiwan Normal University, Taiwan
- Chih-Kai Chang, National University of Tainan, Taiwan
- Chih-Hung Chen, National Taichung University of Education, Taiwan
- Guang Chen, Beijing Normal University, China
- Jun-Ming Chen, National Taiwan University, Taiwan
- Tzung-Shi Chen, National University of Tainan, Taiwan
- Gary Cheng, The Education University of Hong Kong, Hong Kong
- Feng-Kuang Chiang, Shanghai Normal University, China
- Tosti H. C. Chiang, National Taiwan Normal University, Taiwan
- Chih-Ming Chu, National Ilan University, Taiwan
- Hui Chun Chu, Soochow University, Taiwan
- Daniel Churchill, The University of Hong Kong, Hong Kong
- Haiguang Fang, Capital Normal University, China
- Xiaoqing Gu, East China Normal University, China
- Ping He, Tianjin University, China
- Natasa Hoic-Bozic, University of Rijeka, Croatia
- Martina Holenko Dlab, University of Rijeka, Croatia
- Iwen Huang, National Tainan University, Taiwan
- Gwo-Jen Hwang, National Taiwan University of Science and Technology, Taiwan
- M. Carmen Juan, Universitat Politècnica de València, Spain
- Yih-Ruey Juang, Jinwen University of Science and Technology, Taiwan
- Tai-Chien Kao, National Dong Hwa University, Taiwan
- Mohamed Ali Khenissi, University of Tunis, Tunisia
- Mi Song Kim, Nanyang Technological University, Singapore
- Chiu-Lin Lai, National Taiwan University of Science and Technology, Taiwan
- Nguyen-Thanh Le, Humboldt Universität zu Berlin, Germany
- Chen-Yu Lee, Ling Tung University, Taiwan
- Jing Leng, East China Normal University, China
- Chih-Lung Lin, National Pingtung University, Taiwan
- Chiupin Lin, National Tsing Hua University, Taiwan
- Fuhua Lin, Athabasca University, Canada
- Peilin Liu, National Chiayi University, Taiwan
- Chee-Kit Looi, Nanyang Technological University, Singapore
- Ma Luo, East China Normal University, China
- Qing Ma, The Education University of Hong Kong, Hong Kong
- Igor Mekterović, FER, Croatia
- Pey Tee Oon, The University of Macau, Macau
- Kuo-Liang Ou, National Tsing Hua University, Taiwan
- Peter Seow, Nanyang Technological University, Singapore
- Yanjie Song, The Education University of Hong Kong, Hong Kong
- Jun-Ming Su, National University of Tainan, Taiwan
- Ahmed Tlili, Beijing Normal University, China

- Yuen-Hsien Tseng, National Taiwan Normal University, Taiwan
- Noriko Uosaki, Osaka University, Japan
- Peter Wan, The Education University of Hong Kong, Hong Kong
- Sally Wan, ,
- Jijun Wang, Faculty of education, China
- Jingyun Wang, Kyushu University, Japan
- Sheng-Yuan Wang, National Taiwan University, Taiwan
- Silin Wei, Hangzhou Normal University, China
- Xuefeng Wei, Ludong University, China
- Yun Wen, Nanyang Technological University, Singapore
- Gary Wong, The University of Hong Kong, Hong Kong
- Bian Wu, East China Normal University, China
- Di Wu, Central China Normal University, China
- Fati Wu, Beijing Normal University, China
- Yuqin Yang, Central China Normal University, China
- Jane Yau, University of Mannheim, Germany
- Roger Yen, National Taipei University of Education, Taiwan
- Chengjiu Yin, Kobe University, Japan
- Shengquan Yu, Beijing Normal University, China
- Xinglong Zhao, NCET, China
- Gustavo Zurita, Universidad de Chile, Chile

C5 PC Members

- Liz Bacon, Abertay University, UK
- Tak-Wai Chan, National Central University, Taiwan
- Ben Chang, National Central University, Taiwan
- Chih-Kai Chang, National University of Tainan, Taiwan
- Chih-Tsan Chang, National Chung Cheng University, Taiwan
- Chiung-Sui Chang, Tamkang University, Taiwan
- Nian-Shing Chen, National Yunlin University of Science and Technology, Taiwan
- Zhi-Hong Chen, National Taiwan Normal University, Taiwan
- Hercy Cheng, Central China Normal University, China
- Anurag Deep, IIT BOMBAY, India
- Baltasar Fernández-Manjón, Complutense University of Madrid, Spain
- Sara De Freitas, Birkbeck, University of London, UK
- Susan Gwee, English Language Institute of Singapore, Singapore
- Toshihiro Hayashi, Kagawa University, Japan
- Christopher Holden, The University of New Mexico, USA
- Chung-Yuan Hsu, National Pingtung University of Science and Technology, Taiwan
- Mingfong Jan, National Central University, Taiwan, Taiwan
- Kristian Kiili, Tampere University of Technology, Finland
- David Ku, Tamkang University, Taiwan
- Chang-Yen Liao, National Taipei University of Nursing and Health Sciences, Taiwan
- Koong Lin, National University of Tainan, Taiwan
- Saurabh Mehta, vidyalankar Institute of Technology, India
- Hiroyuki Mitsuohara, Tokushima University, Japan
- Wolfgang Mueller, University of Education Weingarten, Germany
- Kuo-Liang Ou, National Tsing Hua University, Taiwan

- Michal Ptaszynski, Kitami Institute of Technology, Japan
- Ma. Mercedes T. Rodrigo, Ateneo de Manila University, Philippines
- Demetrios Sampson, Curtin University, Australia
- Junjie Shang, Peking University, China
- Chun-Yi Shen, Tamkang University, Taiwan
- Ru-Chu Shih, National Pingtung University of Science and Technology, Taiwan
- Masanori Sugimoto, Hokkaido University, Japan
- Kaoru Sumi, Future University Hakodate, Japan
- Han-Yu Sung, National Taipei University of Nursing and Health Sciences, Taiwan
- Ashwin T S, NITK, India
- Shu-Yuan Tao, Takming University of Science and Technology, Taiwan
- Meng-Jung Tsai, National Taiwan Normal University, Taiwan
- Ming Hsin Tsai, Asia University, Taiwan
- Harris Wang, Athabasca University, Canada
- Yi Hsuan Wang, Tamkang university, Taiwan
- Wing-Kwong Wong, National Yunlin University of Science & Technology, Taiwan
- Jianhua Wu, Central China Normal University, China
- Po-Han Wu, National Taipei University of Education, Taiwan
- Haoran Xie, The Education University of Hong Kong, Hong Kong
- Hsi-Hsun Yang, National Yunlin University of Science and Technology, Taiwan
- Kai-Hsiang Yang, National Taipei University of Education, Taiwan
- Sheng-Kai Yin, Cheng Shiu University, Taiwan
- Di Zou, The Education University of Hong Kong, Hong Kong

C6 PC Members

- Alex Boulton, University of Lorraine, France
- Mei-Mei Chang, NPUST, Taiwan
- Ying-Hsueh Cheng, National Pingtung University of Science and Technology, Taiwan
- Yoshiko Goda, Kumamoto University, Japan
- Yanhui Han, The Open University of China, China
- Chia-Ling Hsieh, National Taiwan Normal University, Taiwan
- Phil Hubbard, Stanford University, US
- Jiyou Jia, Peking University, China
- Ho Cheong Lam, The Education University of Hong Kong, Hong Kong
- Jiahang Li, Michigan State University, US
- Meei-Ling Liaw, National Taichung University, Taiwan
- Vera Lúcia Menezes de Oliveira E Paiva, Universidade Federal de Minas Gerais, Brazil
- Misato Oi, Kyushu University, Japan
- Yuichi Ono, University of Tsukuba, Japan
- Zahra Shahsavari, Shiraz University of Medical Sciences, Iran
- Jerry Chih-Yuan Sun, National Chiao Tung University, Taiwan
- Siew Ming Thang, The National University of Malaysia, Malaysia
- Takafumi Utashiro, Hokkai-Gakuen University, Japan
- Jane Vinther, University of Southern Denmark, Denmark
- Wanwisa Wannapipat, Khon Kaen University, Thailand
- Limei Zhang, Singapore Centre for Chinese Language Limited, Singapore
- Di Zou, The Education University of Hong Kong, Hong Kong

C7 PC Members

- Arif Altun, Hacettepe unv, Turkey
- Gargi Banerjee, IIT Bombay, India
- Fatma Bayrak, Hacettepe Üniversitesi, Turkey
- Omer Delialioğlu, Middle East Technical University, Turkey
- Khe Foon Hew, The University of Hong Kong, China
- Sui Lin Goei, Windesheim University of Applied Sciences, Germany
- Tore Hoel, Høgskolen i Oslo og Akershus, Norway
- Gonca Kızılkaya Cumaoğlu, Yeditepe University, Turkey
- Dan Kohen-Vacs, HIT, Israel
- Siu Cheung Kong, The Education University of Hong Kong, Hong Kong
- Shu-Shing Lee, Nanyang Technological University, Singapore
- Marcelo Milrad, Linnaeus University, Sweden
- Vildan Özeke, Gaziosmanpaşa University, Turkey
- Jan Pawlowski, Ruhr West University of Applied Sciences, Germany
- Su Luan Wong, Universiti Putra Malaysia, Malaysia
- Daner Sun, The Education University of Hong Kong, Hong Kong
- Yasemin Usluel, Hacettepe, Turkey
- Joke Voogt, Windesheim University of Applied Sciences, Germany
- Jayakrishnan Warriem, IIT Bombay, India
- Lee Yong Tay, Nanyang Technological University, Singapore
- Halil Yurdugul, Hacettepe University, Turkey

APSCE FELLOWS PROGRAM

Founded in 2019, the APSCE Fellowship recognizes outstanding members of the Asia-Pacific Society for Computers in Education (APSCE) in the field of computers in education. The title of APSCE fellow indicates, (1) Sustained and distinguished academic contributions to the advancement of research in the field of computers in education at the international level; (2) A strong track record in academic networking and services within the Asia-Pacific region.

The fellowship is for life, whose names shall be indicated on the APSCE website permanently. Furthermore, the APSCE fellows are entitled to complimentary lifetime voting APSCE memberships.

The number of new fellows named each year shall be capped at five (5). An APSCE Fellow must be an existing APSCE member in the year he or she is inducted.

The inaugural cohort of the APSCE Fellowship consists of the three existing APSCE Honorary Executive Committee (EC) members. Subsequently, the APSCE President, the APSCE Award Subcommittee Chair and the Honorary EC members formed the APSCE Fellow Committee to select additional fellows. After the first year (2019), the existing APSCE Fellows, the APSCE President and the Award Subcommittee Chair shall form the APSCE Fellow Committee each year to select new fellows. The APSCE President and the Award Subcommittee Chair are not eligible for APSCE Fellow inductions in the year which they are serving as APSCE Fellow Committee members.

The full APSCE Fellowship guidelines is available on <https://tinyurl.com/y3xmo7n6>

The inaugural cohort of APSCE Fellows are (in alphabetical order):

- Tak-Wai CHAN (Taiwan)
- H. Ulrich HOPPE (Germany)
- Chee-Kit LOOI (Singapore)
- Riichiro MIZOGUCHI (Japan)



Tak-Wai CHAN

Chair Professor, Graduate Institute of Networked Learning,
National Central University, Taiwan

Website: <http://chan.lst.ncu.edu.tw/>

Tak-Wai Chan is Chair Professor of the Graduate Institute of Network Learning Technology at National Central University in Taiwan. In the late Eighties, he pioneered research on virtual learning companions. In 1990, his team developed possibly the earliest networked learning system dedicated to support collaborative learning and competitive learning games. He continued to work on various networked learning models, intelligent future classrooms, and digital game-based learning in the Nineties. In 2000, his team built an online learning society called EduCity. By 2003, EduCity had 2,500 schools and 1.5 million online participants before it was transferred to a telecom company. Throughout the 2000's, he worked on mobile and ubiquitous learning, one-to-one technology enhanced learning, seamless learning while he continued to search ways of nurturing student interest in reading, writing and mathematics. At the beginning of this decade, realizing the need of a theory that can guide the design of future education in Asia, he worked with a small group Asian researchers, which grew bigger and an IDC Initiative was formed later, to build a learning design theory called Interest-Driven Creator (IDC) Theory. This theory may exert far-reaching impact in future Asian education because the considerably examination-driven Asian education needs to change.

Chan bears a personal mission to facilitate the building of regional research communities since the early nineties. He co-founded two conference series, ICCE and GCCCE, and, respectively, two corresponding international academic societies, APSCE and GCSCE, one for the Asia-Pacific regional community and the other for the global Chinese community.



H. Ulrich HOPPE

Full Professor, Department of Computer Science and Applied Cognitive Science (Engineering Faculty), University of Duisburg-Essen, Germany

Website: <https://www.collide.info/en>

Dr. H. Ulrich Hoppe holds a full professorship in the area of “Collaborative and Learning Support Systems” at the University of Duisburg-Essen (Germany). After his PhD on interactive programming in mathematics education in 1984, Ulrich Hoppe has worked for about ten years in the area of intelligent user interfaces and cognitive models in HCI, before he re-focused his research on intelligent support in educational systems and distributed collaborative environments in 1995. With his COLLIDE Research Group he has participated in more than ten European projects on Technology-enhanced learning. He was one of the initiators of the European Network of Excellence Kaleidoscope (2004-07). Since 2015 he is engaged as a PI in a Research Training Group on “User Centred Social Media” funded by the German National Science Foundation (DFG). His current research is focused on computational techniques for learning and knowledge building in various context, including higher education as well as vocational education and training. He is an active member of the Learning Analytics community where he particularly pursues the integration and adaptation of computational methods, such as combinations of (social) network analysis with other methods of data mining and artefact analysis.



Chee-Kit LOOI

Professor, National Institute of Education, Nanyang Technological University, Singapore

Website: <https://www.nie.edu.sg/profile/looi-chee-kit>

Chee-Kit Looi is Professor of Education at the National Institute of Education, Nanyang Technological University (NTU) in Singapore. He was the Founding Head of the Learning Sciences Lab, the first research centre devoted to the study of the sciences of learning in the Asia-Pacific region. He is also co-Director of the Centre of Research and Development in Learning, NTU. He organized ICCE 1995 and ICCE 2005 in Singapore, and served as President of APSCE from 1997 to 2011. He is a founding member of the Global Chinese Society of Computers in Education and served as its past president.

Chee-Kit's research in education is characterized by producing outcomes, processes or artifacts that impact practice. An early completed project involved the design of digital mathematics manipulatives which have been made available to all secondary schools in Singapore. He is the PI or co-PI of several research projects funded by the National Research Foundation, Singapore. His research work on creating routine practices of rapid collaborative learning using GroupScribbles has made significant inroads into transforming school practices in several primary and secondary schools. His research on seamless and mobile learning has made good progress toward creating a model of 1:1 computing in schools, remarkable in terms of achieving sustainability and scalability in ten over schools.

Chee-Kit was an associate editor for the JLS, and an editorial member of JCAL, ijCSCL, and IJAIED. He was a member of the Core Expert Group that developed the framework for assessing Collaborative Problem Solving in OECD PISA 2015.



Riichiro MIZOGUCHI

Fellow, Japan Advanced Institute of Science and Technology (JAIST), Japan

Website: <http://www.jaist.ac.jp/ks/english/portfolio/mizoguchi/>

Riichiro Mizoguchi received Ph.D. degree from Osaka University in 1977. He had been a full professor of the Institute of Scientific and Industrial Research, Osaka University from 1990 to 2012 and a research

professor of Research Center for Service Science, Japan Advanced Institute of Science and Technology (JAIST) from October, 2012 to March, 2019. He is currently Fellow of JAIST and Associate researcher, ISTC-CNR Laboratory for Applied Ontology, Trento, Italy. His research interests include Non-parametric data analyses, Knowledge-based systems, Ontology engineering and Intelligent learning support systems. Dr. Mizoguchi was President of International AI in Education Society and Asia-Pacific

Society for Computers in Education from 2001 to 2003 and President of Japanese Society for Artificial Intelligence(JSAI) from 2005-2007. He received honorable mention for the Pattern Recognition Society Award in 1985, Best paper award of the Institute of Electronics, Information and Communication Engineers in 1988, 10th Anniversary Memorial Paper Award of JSAI in 1996, Best paper awards of ICCE99 and ICCE2006, Best paper award of JSAI in 2006 and 2012, and Best paper award of Japan Society for Information and Systems in Education in 2010 and 2019. He was Vice-President of SWSA (Semantic Web Science Association) and Co-Editor-in-Chief of J. of Web Semantics from 2005 to 2009 and from 2008 to 2011, respectively. He is currently an associate editor of ACM TiiS and an editorial board member of Applied Ontology.

EARLY CAREER RESEARCHER AWARD WINNER (2019)



Cheng-Jiu YIN

Associate Professor at the Information Science and Technology Center of Kobe University, Japan

Website: <http://yin.istc.kobe-u.ac.jp/>

Dr. Chengjiu Yin is currently an Associate Professor at the Information Science and Technology Center of Kobe University, Japan. He received his PhD from the University of Tokushima, Japan in March 2008. He was a Visiting Scholar at the Department of Computer Science and Engineering of the University of Minnesota, United States in 2018. He is a member of IEEE, IPSJ, JSET, JSiSE and APSCE.

Dr. Yin is currently the Chair of CUMTEL SIG in APSCE, and the Co-Chair of ES (Extended Summary) of the International Conference on Computers in Education (ICCE 2019). He was the Co-Chair of CUMTEL SIG in APSCE during 2016-2017. He served as Executive Chair of ICCE C4 (CUMTEL) sub-conference twice, and Poster Coordination Chair of ICCE2017. In addition, he has served as the Guest Editor of Journal of Distance Education Technologies (EI, ESCI) thrice, and Interactive Learning Environments (SSCI) for four times. He has also successfully organized many other sub-conferences (e.g. IIAI-AAI-LTLE2012, IIAI-AAI-LTLE2013, IIAI-AAI-LTLE2014) and workshops (e.g. LEET2017, LEET2018, LEET2019).

Dr. Yin has more than 10 years of experience in the field of Educational Technology such as Language Learning, Mobile Learning and Learning Analysis. His research works have been published in many prestigious international journals such as Journal of Educational Technology & Society, Interactive Learning Environments, and others. He has published more than 150 academic papers — 45 of which are journal papers (including 13 SSCI/SCI-indexed and 7 EI-indexed journal papers). His current research foci are in Educational Data Mining and Learning Analytics. He was invited to deliver speeches on these topics by several universities, such as Wuhan University, Northeast Normal University and Asia University Educational Data. He was also an invited speaker at the International Joint Conference on Information, Media and Engineering (ICIME2018).

Dr. Yin received the Best Paper Award (Short Paper) at the 15th IEEE International Conference on Advanced Learning Technologies (ICALT 2015) and Best Poster Paper Award at the IIAI International Congress on Advanced Applied Informatics 2019. He has received many nominations for Best Paper Awards at various conferences including ICCE (Best Students Paper Candidate in ICCE2018, Best Overall Paper Candidate in ICCE2017, Best Technical Paper Candidate in ICCE2015, Best Overall Paper Candidate in ICCE2015).

LAST TEN YEARS'
DISTINGUISHED RESEARCHER AWARD WINNERS

2015 - APSCE Distinguished Researcher Award

Lung-Hsiang WONG, Nanyang Technological University, Singapore

2014 - APSCE Distinguished Researcher Award

Hiroaki OGATA, Kyushu University, Japan

2011 - APSCE Distinguished Researcher Award

Antonija MITROVIC, University of Canterbury, New Zealand

Chen-Chung LIU, National Central University, Taiwan

2009 - APSCE Distinguished Researcher Award

Fu-Yun YU, National Cheng Kung University, Taiwan

Tsukasa HIRASHIMA, Hiroshima University, Japan

LAST FIVE YEARS'
EARLY CAREER RESEARCHER AWARD WINNERS

2018 - APSCE Early Career Researcher Award

Ting-Chia HSU, National Taiwan Normal University, Taiwan

2017 - APSCE Early Career Researcher Award

Jon MASON, Charles Darwin University, Australia

2015 - APSCE Early Career Researcher Award

Morris Siu-yung JONG, The Chinese University of Hong Kong, Hong Kong

KEYNOTE SPEAKERS



Stephen J.-H. Yang

National Central University, Taiwan

Title:

Precision Education: New Challenges for AI in Education

Abstract: After decades of research on AI in education, we have solved some problems and are facing new challenges. Precision education is one of the new challenges of applying artificial intelligence, machine learning, and learning analytics for adaptive and individualized learning in a context-aware ubiquitous learning environment. The goal of Precision education is to identify at-risk students as early as possible and provide timely intervention based on student learning experiences. In this research, at-risk students are confined to students who were diagnosed could get low GPA (low academic performance), could drop/withdraw a course, or students who were low engagement in terms of learning behavior, emotion, and cognition. The idea of Precision education came from The Precision Medicine Initiative (<https://obamawhitehouse.archives.gov/precision-medicine>), which was proposed by President Obama in his 2015 State of the Union address, the Initiative is a new research effort to revolutionize the medical treatment of disease. As addressed in this Initiative, most treatments were designed for the average patients as a result of one-size-fits-all-approach without taking into account individual differences in people genes, environments, and lifestyles; similarly, most pedagogy is designed not fully consider students genes IQ, learning styles, learning environments, and learning strategies. Therefore, by carrying the research focus of Precision medicine, which is to improve the diagnosis, prediction, treatment, and prevention of disease, we define the new challenges of Precision education on emphasizing the improvement of diagnosis, prediction, treatment, and prevention of at-risk students.

Biography: Dr. Stephen J.-H. Yang is the Chair Professor of Computer Science & Information Engineering, National Central University, Taiwan. Dr. Yang received the Outstanding Research Award from Ministry of Science; Technology (2010, 2017) and Distinguished Service Medal from Ministry of Education, Taiwan. (2015). Dr. Yang was the Director of Department of Information and Technology Education, Ministry of Education, Taiwan (2013~2014), during his two years of service in Taiwan government, Dr. Yang was responsible for the national information & technology education, he launched Taiwan national digital learning initiative which includes the construction of 100G Taiwan Academic Network for national network infrastructure, the construction of Education Cloud for national data infrastructure, and innovation programs such as Taiwan MOOCs, flipped classroom, mobile learning. Dr. Yang also served as the Convener of Information Education Discipline, Ministry of Science & Technology. Dr. Yang received his PhD degree in Electrical Engineering & Computer Science from the University of Illinois at Chicago in 1995. Dr. Yang has published over 70 SSCI/SCI journal papers. His research interests include Artificial Intelligence, Big Data, learning analytics, educational data mining, and MOOCs. As shown on Google Scholar Citation, Dr. Yang publication citation indices has been over 11,000,

especially on the main research themes, Artificial Intelligence in education is ranked #6, Learning analytics is ranked #8, Educational data mining is ranked #3, MOOCs is ranked #4. Dr. Yang is currently the Co-Editors-in-Chief of the International Journal of Knowledge Management & e-Learning.



Hiroaki Ogata

Kyoto University, Japan

Title:

Toward Evidence-based Education through Learning Analytics

Abstract: The multi-disciplinary research approach of Learning Analytics (LA) has provided methods to understand learning and teaching process by analyzing logs collected during diverse teaching-learning activities and potentially enrich such experiences. This talk will propose the Learning Evidence Analytics Framework (LEAF) and draw a research road-map of an educational big data-informed evidence-based education system. It focuses on the approach of developing novel techniques by applying the knowledge base of LA that can help to extract evidence of effective teaching-learning practices. Finally, it shows that teachers can refine their instructional practices, learners can enhance learning experiences, and researchers can study on the dynamics of the teaching-learning cases with LEAF.

Biography: Hiroaki Ogata is a Professor at the Academic Center for Computing and Media Studies, and the Graduate School of Informatics at Kyoto University, Japan. His research includes Computer Supported Ubiquitous and Mobile Learning, CSCL, CALL, and Learning Analytics. He has published more than 300 peer-reviewed papers including SSCI Journals and international conferences. He received the APSCE Distinguished Researcher Award in 2014 and several Best Paper Awards. Also he has given keynote lectures in several countries. He is an associate editor of the IEEE Transactions on Learning Technologies, RPTL, and IJMLO, and also an editorial board member of IJCSCL, IJAIED, JLA and SLE. He is an Executive Committee member of SOLAR and APSCE.



Mina. C. Johnson

Arizona State University, USA

Title:

VR, Augmented, and Mixed Realities (XR) for STEM and Game-based Learning: Focus on Design

Abstract: The Embodied Games Lab specializes in both creating and efficacy testing educational XR content. The opportunities for researching STEM education with immersive Virtual Reality (VR) have grown tremendously over the past three years. High resolution, stand-alone headsets are now affordable for many classrooms. Soon, teachers will ask, "Where is the quality educational content?" We need to start creating those experiences now. Many instructors and technologists want to move past the gaze-only style of mobile, non-interactive VR. The new generation of interactive hand controls allows learners to be very active and embodied as they manipulate content - in VR especially. The talk will describe and give examples of a set of 17 best practices in VR. Additionally, not all phenomena should be re-envisioned in 3D. It is expensive, and can perhaps be overwhelming; the sense of "presence" created by IVR is profound and immediate. This talk will also focus on which phenomena should be chosen. It will highlight how to use the body and gestures to interact with mediated content, and methods for embedding assessment into content. These sorts of embedded and game-like assessments can help teachers understand if the learner is really comprehending, or simply 'going through the motions'.

Biography: Dr. Johnson-Glenberg is a Research Professor at Arizona State University in the Psychology department. She is also an entrepreneur who is the Founder/President of the award-winning learning technology company called Embodied Games. Her ASU spin-out company creates, researches, and distributes innovative virtual and mixed reality (XR) educational content for 4th graders through life-long learners. As a graduate student at the University of Colorado at Boulder, she cut her teeth on designing software to remediate dyslexia in primary school children. For the past decade, she has led both a research lab and a small company in designing and producing STEM education games and simulations. She has a varied background as a screenwriter, experimental psychologist, and psychometrician; these disparate skills serve her well in leading a creative game design team dedicated to creating efficacious content. She presents nationally and internationally, and she has published in multiple peer-reviewed journals. Her most recent publication on VR and STEM education can be found in *Frontiers in Robotics and AI* <https://doi.org/10.3389/frobt.2018.00081>.

The Embodied Games lab has been at the forefront of creating "embodied" and kinesthetically-active games using sensors and joint-tracking technologies. Because the majority of Dr. Johnson-Glenberg's games have been federally funded, they can be downloaded and played for free at www.embodied-games.com.



Joke Voogt

University of Amsterdam, Netherlands

Title:

Preparing Learners for Living and Working in the Digital Age

Abstract: Due to technological developments young learners have to cope with many challenges related to living and working in the digital age. Often the term of 21st century competencies is introduced to point to the competencies students need to acquire to be able to cope with these challenges. However it is not that easy to incorporate 21st century competencies in the school curriculum. It requires a focus on meaningful learning through authentic tasks. Curriculum and assessment practices need to allow for such an approach to learning and schools and teachers need to be prepared to teach 21st century competencies. The focus of the presentation will be on research-based examples of meaningful learning through authentic tasks using technology and the conditions that need to be in place to realize them in educational practice.

Biography: Dr. Joke Voogt is professor of ICT and Curriculum at the Faculty of Behavioural and Social Sciences of the University of Amsterdam and professor of Educational Innovation and ICT at Windesheim University of Applied Sciences. Since the 1980s she has been investigating the integration of Information and Communication Technology in educational practice in national and international settings. She was part of the coordination team of Second Information Technology in Education Studies (1998-2008) of the International Association for the Evaluation of Educational Achievement and is currently involved as a curriculum specialist in the OECD Education 2030 study: The future of Education and Skills. Dr. Voogt is co-editor of the first and second edition of the International Handbook of Information Technology in Primary and Secondary Education (2008, 2018). She is co-founder and chair of the EDUsummIT, the international Summit for research, policy and practice that emerged from the first edition of the Handbook. Dr. Voogt has more than 100 scientific publications. Her areas of expertise are curriculum innovation, teacher learning and development, 21st century skills, in particular related to the integration of technology in education. Dr. Voogt has a master degree in Educational Science and Technology (with distinction) and a doctoral degree (PhD) in Educational Sciences from the University of Twente, The Netherlands.

THEME-BASED INVITED SPEAKERS



Bodong CHEN

University of Minnesota, USA

Title:

Creating Conditions for Knowledge Building for the Public Good

Research fields: Computer-supported collaborative learning; Learning analytics; Knowledge building; Higher order competencies; Social media

Abstract: Knowledge creation and innovation are becoming pervasive and essential in meeting today's societal demands. This has led to increasing pressure on education to explore new pedagogical models, to move from knowledge transmission to educational paradigms that are more creative, collaborative, and connected. Knowledge Building represents a longstanding effort to redesign education as a knowledge-creating enterprise. Knowledge Building immerses learners in creative work with ideas from the earliest grades; it engages learners in continually improving knowledge of value to their class community; it challenges learners to take greater collective responsibility for their work. Students are asked to identify promising ideas, participate in meta-cognitive dialogues, play epistemic games, assess themselves, and grapple with complex, emergent knowledge structures. These competencies and high-level learner agency espoused by Knowledge Building are needed more than ever in the era of automation and A.I.

In this talk, I will introduce ongoing design-research projects that attempt to further stretch students' knowledge building in classrooms to connect more deeply with the public. Overall, I posit that with proper pedagogical and socio-technical conditions learners are capable of meaningfully participating in public discourse while advancing curricular goals. In particular, I will introduce computer-supported collaborative learning (CSCL) environments we designed to support ninth graders to study "Energy and Matter" in connection with the "Green New Deal" in the United States; I will also showcase computational research conducted by sixth graders in relation to the United Nations Sustainability Development Goals. Alongside the explication of the design principles and designed environments, I will share episodes of students' knowledge building and evidence of their knowledge advancement, expressiveness, and higher-order competencies. In doing so, I hope to challenge our entrenched beliefs about learner capabilities and highlight the potential of nurturing conditions in K-12 classrooms for knowledge building that connects with and contributes to the public good.



Xiaoqing GU

East China Normal University, China

Title:

**Embrace Data Intelligence in Designing for the e-Schoolbag:
Addressing the 1:1 Challenges**

Research fields: ICT in Education CSCL ICT-integrated Pedagogy Learning behaviour and student model in Technology systems learning analytics

Abstract: The popularity of personal mobile devices has fuelled interest in deploying this technology to transform the classroom. In general, 1:1 computing refers to the projects where technology is available to all students and their teachers. Such developments are now widespread in China, where the personal digital device in class has also been termed the 'e-Schoolbag' initiative.

One finding echoed in 1:1 projects across the world is that when these technologies are implemented, the effect is much less transformative than expected. Our studies intended to indicate how best to innovate 1:1 classroom practice from two perspectives: first, the innovative learning design in support of teaching and learning in the 1:1 classroom; second, learning analytics in an ICT-rich environment to support learning diagnosis and learning improvement.

I am going to unfold the innovative learning design for 1:1 classroom with a project centered on the development of e-Textbook. In a tradition of instruction that relies heavily on the textbook, the emerging e-Textbooks may become implicated in how teachers could innovate practice within the 1:1 class. I will first introduce a large scale e-Textbook project, where the textbook is still at the central place of the 1:1 classroom; follow that I will introduce a case study, in which we invited a classroom teacher to take up the role of designer in shaping the content of the class text. Both the design and the access to its outcome by students were made possible by the tablet technology at the heart of the e-Schoolbag idea.

The challenge of 1:1 classroom lies not only on the innovation of learning design, but also the learning intervention on the active learning activities that follows. In this sense, I am going to start with a large scale project with a heavy feature on harnessing data intelligence to innovate learning; follow that I will introduce a case study to illustrate how the learning analytic could help deal with the challenges on 1:1 classroom.



Yuichi ONO

University of Tsukuba, Japan

Title:

Flipped-Classroom Course Model with ICT Support to Activate Discussion in Foreign Language Classrooms

Research fields: Foreign language education

Abstract: There is a general agreement among foreign-language researchers and instructors that course models should meet the requirements of "vigorously changing societies" in the twenty-first century. On the basis of this general agreement, there have been various types of learning paradigms proposed and assumed in order to construct appropriate foreign language course models. In this thematic-based invited talk, I discuss the flipped-classroom course model to enhance discussion course model.

My talk starts with some theoretical backgrounds related to the current research such as Interest-Based Learning, Seamless Learning, Project-Based Learning Skills, Information Literacy Skills, and Foreign Language Anxiety in constructing the proposed course model. I introduce the recent policies set out by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Japan; the relationship of national curriculum with the so-called one-to-one (one tablet PC to one student) policy.

Secondly, I discuss some merits of utilizing online educational resource (OER); such as YouTube, TED Talks, Kahn Academy, cMOOC/xMOOC, etc, and suggest the potentials to draw learner's willingness to cultivate and deepen their interest through choosing these authentic educational online materials. The learning support system, or quiz generator, to encourage learners nicely to familiarize these authentic materials in terms of vocabulary learning and listening practice is also introduced.

Thirdly, I report some findings of my research concerning Flipped-Classroom Course Model that was implemented to enhance EFL learners' information skills. Specifically, the change of learner's factors such as proficiency, foreign language anxiety, PBL skill awareness are analyzed through the project and how they proceed their learning under the ICT environment within and outside the classroom.

Lastly, my talk concludes that the course model employing ICT-supported Flipped Classroom is successful so far in enhancing interactive activities in the classroom in the sense of active learning and information literacy acquisition, followed by some future implications.

PANELS

ASSESSMENTS FOR COMPUTATIONAL THINKING IN PRIMARY AND SECONDARY SCHOOLS2

CHAIR: SIU CHEUNG KONG, THE EDUCATION UNIVERSITY OF HONG KONG, HONG KONG

CO-CHAIR: CHEE KIT LOOI, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

PANELISTS:

TING-CHIA HSU, NATIONAL TAIWAN NORMAL UNIVERSITY, TAIWAN

JU-LING SHIH, NATIONAL CENTRAL UNIVERSITY, TAIWAN

A LEARNING THEORY DESIGN FOR ASIA IN THE 21ST CENTURY: INTEREST-DRIVEN CREATOR THEORY (IDC)5

CHAIR: BEN CHANG, NATIONAL CENTRAL UNIVERSITY, TAIWAN

PANELISTS:

TAK-WAI CHAN, NATIONAL CENTRAL UNIVERSITY, TAIWAN

WENLI CHEN, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

SIU CHEUNG KONG, THE EDUCATION UNIVERSITY OF HONG KONG, HONG KONG

CHEE KIT LOOI, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

LUNG HSIANG WONG, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

QUESTIONS, DESIGN, INDICATORS FOR CSCL6

CHAIR: GAOWEI CHEN, HONG KONG UNIVERSITY, HONG KONG

PANELISTS:

CHEW LEE TEO, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

BODONG CHEN, UNIVERSITY OF MINNESOTA - TWIN CITIES, USA

LUNG HSIANG WONG, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

DISCUSSANT: ELIZABETH KOH, NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

Assessments for Computational Thinking in Primary and Secondary Schools

Siu Cheung KONG ^{a*}, Chee Kit LOOI ^b, Ting-Chia HSU ^c, & Ju-Ling SHIH ^d

^a*The Education University of Hong Kong, Hong Kong*

^b*Nanyang Technological University, Singapore*

^c*National Taiwan Normal University, Taiwan*

^d*National University of Tainan, Taiwan*

*sckong@eduhk.hk

Abstract: Computational thinking (CT) is essential for the young generation in the 21st century to succeed in their learning and daily-life in the digitalized society. It is important for learners in primary and secondary schools to have opportunities to not only meaningfully develop the concepts and practices fundamental to CT competencies, but also for the progress and outcome of their CT development to be assessed in daily learning contexts. The existing literature lacks a universally agreed-upon method of assessment for CT development; and many existing methods for assessment of CT development focus on university learners. This panel aims to address the gap of research discussion about assessments for CT development among young learners in primary and secondary schools. This panel will consist of panel presentations and open-floor discussion about the related experiences, existing challenges, expected dimensions, and possible approaches of CT assessments in primary and secondary schools. This panel can inform the evidence-based evaluation of CT education in the 21st century.

Keywords: Assessments, computational thinking, primary schools, secondary schools

1. Introduction

In recent years, computational thinking (CT) is advocated to be an essential competency for learners to not only take advantages but also make contributions in the digitalized society in the 21st century (Barr, Harrison, & Conery, 2011; Shute, Sun, & Asbell-Clarke, 2017; Wing, 2006). This drives a growing trend in the recent years that primary and secondary education sectors over the world arrange school curriculum delivery to integrate elements of CT education (Barr & Stephenson, 2011; Grover & Pea, 2013; Lye & Koh, 2014; Shute et al., 2017). The successful implementation of CT education in school learning contexts have two main concerns: to suitably plan what and how schools can teach in the curricular activities for CT education; and to suitably measure what and how learners have learned from the curricular activities for CT education (Lye & Koh, 2014; Chen et al., 2017; Voogt, Fisser, Good, Mishra, & Yadav, 2015). In recent years, the research community has many discussions about the first concern and builds the foundation of curriculum design and pedagogical implementation for CT education in school curriculum. It is now the time to extend the research discussion to address the second concern—the assessments for the progress and outcome of CT development among learners, especially the ones for young learners in primary and secondary schools, as the existing literature on assessment of CT development focuses on the assessment methods for university learners.

2. Abstracts of Individual Panelists' Presentation

2.1 *Evaluation of Computational Thinking among Senior Primary School Students: The Components and Methods (Siu Cheung KONG)*

Teaching practitioners in school education generally face challenges to comprehensively evaluate the progress and outcome among young students in computational thinking (CT) development. This

presentation will introduce the essential components and suitable methods for evaluating CT concepts, CT practices and CT perspectives among senior primary school students. The review of literature from 2010 onwards identifies that it is essential to address nine components of CT concepts, seven components of CT practices, and three components of CT perspectives when teaching practitioners evaluate students' CT development. Thus, it is appropriate for teaching practitioners to evaluate students' CT concepts by test designs with multiple choice questions; their CT practices by rubric-rating of programming projects and test designs with task-based questions; and their CT perspectives by student questionnaire surveys. This presentation can inform the direction of teaching practitioners to foster creative problem-solvers through the meaningful assessments in CT curriculum in school education.

2.2 Using Task-Based and Situated Test Items to Evaluate Computational Thinking of Primary and Secondary School Students (Ting-Chia HSU)

Computational thinking of students could be evaluated with computer and without computer. For example, Bebras provides the students with the multiple choice tests although the students use a computer with the Internet to fill out the answers. In Taiwan, some senior high school students participate in the test named APCS (Advanced Placement Computer Science) containing both conceptual tests of programs and writing programs with computer. They can use the scores of APCS to apply to some universities in Taiwan. Consequently, most of the time, the students have to solve task-based and situated test items so that their concepts and practices of computational thinking can be assessed. Finally, their perspectives, such as relating to self-efficacy, attitude, motivation and so on, are evaluated with proper questionnaires. More tools and examples will be introduced in the panel.

2.3 Potential of Multimodal Assessments of Computational Thinking in the Interdisciplinary Robotic Game (Ju-Ling SHIH)

In my educational project, STEAMing the ships for the Great Voyage, computational thinking is integrated in an interdisciplinary robotic game. Student groups role-play the five main countries in the Age of Discovery using computers to write block programming to control the robotic ships to voyage on the world map. In the game, computational thinking is the key to deploying diplomatic strategies, cooperating or competing with other countries, solving historical crisis, and completing tasks by trading or battling for spices.

Multimodal assessments take multiple sources of data for both quantitative and qualitative data analysis. Other than questionnaires and tests, it includes video/audio recordings of all the participants which can be textualized for content analysis, with the addition to the coding logs from the system, as well as the observations of their behaviors. The purpose is to diagnose the participants' gaming strategies, the dynamic interpersonal and inter-group interactions, and critical/creative/computational thinking skills.

2.4 Assessment of the Effects of the Informal Curriculum of Computing Education in Singapore (Chee Kit LOOI)

In Singapore, the IMDA government agency has been promoting CT education through running or supporting programmes for schools such as Code for Fun (in which students learn Scratch programming) and Digital Making (in which students do physical computing on the microBit). Many schools (but not all schools) opt to offer these computing-related programmes, and hence we allude to them as the informal curricula which also include community-based or private tuition programmes. The objective is to interest students in computing/coding, with a view towards nurturing students to take up computing or related subjects, thereby contributing to the future pipeline of ICT manpower in Singapore. The Ministry of Education promotes CT education through offering the Computing subject in Secondary 3 and 4, and the Computer Science subject in Junior Colleges (grades 11-12). With a rich ecology of CT education provision involving informal learning and formal learning, we discuss research efforts to understand the outcomes of these current efforts to promote and teach computing and coding in Singapore. The relevant assessment questions are: What are the pathways of students learning

CT through the informal curricula in schools? What and how are the students learning CT in the different parts of the ecosystem? With respect to what and how the students are learning Computational Thinking in the different parts of the ecosystem, we need to collect data from students, teachers, principals and vendors participating in these computing-related applied learning programmes, and analyse students' perceptions and CT artifacts to assess CT readiness, progression and growth. In this panel, I provide a discussion of the methods to explore these assessment questions.

3. Discussion and Conclusion

The panel will first consist of panelist presentations which share the related experiences of assessing the progress and outcome of CT development among learners in primary and secondary schools across different countries / regions around the world. The panelists will share their insights into the focuses and approaches desirable for CT evaluation; the design of test items for CT evaluation; the models of CT evaluation in game-based learning; and the methods of CT evaluation in informal curriculum. The panel will then discuss what challenges that primary and secondary schools are facing in the deployment of CT assessments; what dimensions that CT assessments are expected to be covered according to different grades of young learners; and what approaches that are possible for teaching practitioners to accurately assess CT development among learners in primary and secondary schools. The panel is expected to provide participants with insights related to the universally agreed-upon methods of assessment for CT development.

References

- Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology*, 38(6), 20-23.
- Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads*, 2(1), 48-54.
- Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education*, 109, 162-175.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.
- Lye, S.-Y., & Koh, J. H.-L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, 41, 51-61.
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158.
- Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies*, 20(4), 715-728.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.

A Learning Theory Design for Asia in the 21st Century: Interest-Driven Creator Theory (IDC)

Ben Chang^{a*}, Tak-Wai Chan^a, Wenli Chen^b, Siu Cheung Kong^c, Chee Kit Looi^b, Lung Hsiang Wong^b, Ju-Ling Shih^d

^a *Graduate Institute of Learning and Instruction, National Central University, Taiwan*

^b *National Institute of Education, Nanyang Technological University, Singapore*

^c *The Education University of Hong Kong, Hong Kong*

^d *Department of Information and Learning Technology, National University of Tainan, Taiwan*

* bchang.tw@gmail.com

Abstract: If we researchers in Asia want to bring about any profound transformation in education, we have to face a great barrier to resistance: a significant part of formal education in Asia remains ‘examination-led education’, that is, educational practice is largely governed by a short-term goal: getting higher scores in public examinations. This leads to over-emphasizing cognitive outcomes, resulting some serious drawbacks: learning and teaching are distorted; many students do not enjoy learning; it is hard for students to develop 21st century competencies.

The 21st century, however, marks an era of exponential change and emergence. Our world demands its citizens a lifetime of creative and critical thinking, endlessly delivering innovations, new values and productivity to thrust social and economic development. Have our societies, schools and families found the right ways to prepare the young generation for the 21st century? Will our children thrive in the 21st century?

In contrast to the enormous educational movements described above, researchers have long been undertaking small-scale experiments, intending to improve students’ learning experience, test their learning outcomes, and develop knowledge on learning and teaching as well as some micro-level theories. However, to transform Asian education with a sustainable impact on a large scale, we need a macro theory to guide the design and direct research towards the creation of a form of quality education for Asia. A group of Asian researchers has created a learning theory named IDC Theory (Interest-Driven Creator Theory), and they will elaborate on this theory in this panel.

Questions, Design, Indicators for CSCL

ChewLee TEO, Lung Hsiang WONG, Bodong CHEN, Gaowei CHEN, Elizabeth KOH

^a *Chew Lee Teo, Lung Hsiang Wong, Elizabeth Koh, National Institute of Education, Singapore*

^b *Bodong Chen, University of Minnesota, USA*

^c *Gaowei Chen, Hong Kong University, Hong Kong*

*Chew Lee, Teo @corresponding.author

Abstract: Computer-supported-collaborative-learning (CSCL) has been one of the major thrusts in educational initiatives across the globe. For example, "collaborative learning" has been the buzzword since the second ICT master-plan in Singapore. The language of CSCL has since taken root in many teachers' professional development effort and repertoire of practice. However, though the processes of research, development, and implementation of CSCL in schools in Singapore have increased extensively in recent years, the mode of assessment and the deepening in understanding of collaborative learning and its environment continued to lag behind the adoption rate. Many teachers still found difficulties to trust the process of collaborative learning and to confidently say that all their students have learned in a collaborative setting. Of course, mindset and belief of teachers are of enormous significance, but we would also like to ascribe this phenomenon among teachers to the complexity of latent attributes such as learners' cognitions and emotions in the learning process, which differ significantly from traditional practice. This sets the motion for a 'call for multimodality' in CSCL to define a more holistic way through comprehensive sets of data and analyses to understand the process and outcome of CSCL (Schneider & Blikstein, 2015). This 'multimodality' has tremendous potential but yet to reach an easy-to-implement status, as it is now still at the early research phase. Given the myriads of development happening around the world, this panel aims to bring together existing research on various dimension and collectively derive a framework from providing teachers a systemic understanding of CSCL indicators to inform the key challenges or the critical activities in their CSCL practice.

Keywords: Computer-supported collaborative learning; assessment; indicators; analytics

1. Introduction

Computer-supported-collaborative-learning (CSCL) has been one of the major thrusts in educational initiatives across the globe. For example, "collaborative learning" has been the buzzword since the second ICT master-plan in Singapore. The language of CSCL has since taken root in many teachers' professional development effort and repertoire of practice. However, though the processes of research, development, and implementation of CSCL in schools in Singapore have increased extensively in recent years, the mode of assessment and the deepening in understanding of collaborative learning and its environment continued to lag behind the adoption rate. Many teachers still found difficulties to trust the process of collaborative learning and to confidently say that all their students have learned in a collaborative setting. Of course, mindset and belief of teachers are of enormous significance, but we would also like to ascribe this phenomenon among teachers to the complexity of latent attributes such as learners' cognitions and emotions in the learning process, which differ significantly from traditional practice. This sets the motion for a 'call for multimodality' in CSCL to define a more holistic way through comprehensive sets of data and analyses to understand the process and outcome of CSCL (Schneider & Blikstein, 2015). This 'multimodality' has tremendous potential but yet to reach an easy-to-implement status, as it is now still at the early research phase. Given the myriads of development happening around the world, this panel aims to bring together existing research on various dimension and collectively derive a framework from providing teachers a systemic understanding of CSCL indicators to inform the key challenges or the critical activities in their CSCL practice.

2. The Design-Indicators-Question frame.

Designing and implementing CSCL environment requires teachers to consider various dimensions in the collaborative learning process such as social (skills, awareness), technological (social presence and content support, online resource), pedagogical situation (definition of activities, artefacts, content and group work), and motivation (engagement). There are many pedagogical scaffold tools designed for teachers, for example, Persico et al. (2010) derive at "design pattern (DP) instances, that are based on a set of indicators, to support monitoring and evaluation of CSCL interactions, as well as on the methods to gauge these indicators. The question-design-indicator used in this panel discussion attempt to explore the connections between the different aspects of CSCL practice in various CSCL classrooms activity to understand the design-implementation-evaluation process.

Question-Design-indicator frame (Table 1): We begin the unpacking with teachers' questions about what they hope to see in the CSCL activities which will lead them to think about the design and enactment of the activities. We then continue to map different analytics and visualisation as indicators that could inform teachers what they set out to do. Afterwhich we returned to the prioritisation questions that would challenge the teachers to think about the two areas of prioritisation: (i) how they are adopting the outcomes, diagnostic or emergent; (ii) are they more concern with students taking agency of learning or are their control over the instructional progression.

We will be discussing three cscl scenarios based on the question-design-indicator frame. We hope the discussion can allow teachers to think about their intent of CSCL so as to (i) scaffold deep and sustained collaborative teaching and learning beyond superficial group-work; (ii) supports leaders and educators in analysing, monitoring, and improving CSCL activities; (iii) facilitates design of curriculum assessment of students' engagement, 21CC and deep learning.

3. CSCL Scenario #1: Students presented with the individualised “report” that explained their collaborative activities.

This scenario presents an on-going research effort of individual "reports" presented to students in a knowledge-building classroom. The individualised cards include evidence of students' participation and contribution to the collective inquiry in class from an online discussion forum, Knowledge Forum. The cards provide visualisation of students' soft-skills development such as contribution patterns, leadership in ideas, and graphs of scaffold support (sentence starter) used by students in their notes. The content learning is presented to students with the proxy of the keywords used by the students in their collaborative discussion as compared to the keywords used by their class. The list of keywords adopted for the analytics are selected by teacher and researchers referencing to both curriculum document and students' discussion. This CSCL scenario is unpacked in table 1, according to the question-design-indicator framework (Table 1). The panel will present preliminary results based on students and teachers' reflection about the two areas of prioritisation.

4. CSCL Scenario #2: Seamless environment in and out of class, online and offline environment for CSCL practice

Students' collaborating seamlessly in and out of classroom: Authentic collaboration happens in and out of classroom naturally (Looi et al, 2010). A student engaged in the class inquiry on light would be observing shadows and rainbow on their way home in a different manner, they might be triggered more questions to ask their friends walking home with them (not their collaborating partners), they might challenge some of the explanations they hear in class. They might actively discuss about the phenomenon with their families. How do teachers make sense of these children's thoughts as these children bring them back to the classroom? How do teachers capture these rich learning happening outside classroom that might change the dynamic interaction in the CSCL lesson? Panelist will build on to the CSCL scenario by providing insights of seamless learning and its importance and realisation in

daily classroom practice through the dimensions and characterisation of mobile-assisted seamless learning (MSL) (Lung & Looi, 2011) using the question-design-indicator frame.

5. CSCL scenario #3: Collaborative and progressive discourse (Bodong Chen)

Supporting collaborative discourse in CSCL activities: Facilitating whole-class and small group collaborative discussions is a core practice of CSCL. Such discourse-centric pedagogy demands teachers to instill confidence in individual students, establish nurturing classroom culture, and push students to improve reasoning (Michaels & O'Connor, 2012). Individual students should, through this collaborative discourse, achieve better understanding and take ownership in their learning. The whole issue about supporting collaborative discourse is that the teacher might have to deviate from their planned procedure and seek new ways of responding to students' interest and learning. It might involve reflective noticing, questioning, and revoicing, embraces opportunistic knowledge processes and working emerging problems (Zhang et al., 2018). Panelist will share his extensive research on technology and analytic development, e.g. Promising Idea Tool ideas, to discuss the scaffolds given to teachers and students in a collaborative and progressive discourse. The panel will discuss on the basic design idea of these tools in a classroom environment and the implication of such design on pedagogy to realise the potential of young children in evaluating and reflection on ideas generated by their peers (Chen et al, 2015).

6. Discussions

The integration of different cognitive, socio-cognitive, and technological aspects of learning through CSCL has been such a break-through concept in the past two decades. It has indeed impacted the way school leaders, teachers and students think about learning. However, the traditional mode of learning has been so deeply entrenched in the practice that many in the mainstream schools continue to superimpose the traditional transactional learning onto CSCL. CSCL community needs to address the challenges. One of the challenges that especially resonated with us is that that "the interdependence between different aspects needs to be better represented and discussed, rather than considering them cause-effect relationships" (Dillenbourg et al. 2009). The learning happening in CSCL is much more productive than merely constrained by a cause-effect relationship, which is typically a transactional approach to learning. Unless we can present a principled and system-based approach to CSCL and translate this interpretation to teachers (Scardamalia et al., 2012), we hope this panel can contribute to the discussion of a principled and system view of CSCL practice in classrooms.

Reference

- Persico, D., Pozzi, F., & Sarti, L. (2009). Design patterns for monitoring and evaluating CSCL processes. *Computers in Human Behavior*, 25(5), 1020-1027.
- Barros-Castro, R. A., Córdoba-Pachón, J. R., & Pinzón-Salcedo, L. A. (2014). A systemic framework for evaluating computer-supported collaborative learning—Mathematical problem-solving (CSCL-MPS) initiatives: Insights from a Colombian case. *Systemic Practice and Action Research*, 27(3), 265-285.
- Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, 7(3), 89-116.
- Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. *Socializing intelligence through talk and dialogue*, 347-362.
- Looi, C. K., Seow, P., Zhang, B., So, H. J., Chen, W., & Wong, L. H. (2010). Leveraging mobile technology for sustainable seamless learning: a research agenda. *British journal of educational technology*, 41(2), 154-169.
- Zhang, J., & Yuan, G. (2018). Cross-Classroom Interaction for Knowledge Building: A Design Experiment in Four Grade 5 Science Classrooms. In *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS)*. London, UK: International Society of the Learning Sciences.
- Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In *Technology-enhanced learning* (pp. 3-19). Springer, Dordrecht.

Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In *Assessment and teaching of 21st century skills*(pp. 231-300). Springer, Dordrecht.

Wong, L. H., & Looi, C. K. (2011). What seams do we remove in mobile-assisted seamless learning? A critical review of the literature. *Computers & Education*, 57(4), 2364-2381.

Annex A. Table 1 Question-design-indicator table to guide design and analysis

The following table unpack a sample CSCL Activity: students collaborating in groups within the class to pursue an inquiry task, facilitated by teachers.

Learning outcomes	Teachers' questions	Design	Indicators: Types of visualisations and information to present to teachers	Questions: Using the design-indicators to prioritise students' agency over teachers instructional control; prioritise	Questions: Using design-indicator to prioritise emerging quality and teaching moves.
E.g. cognitive and metacognitive development	<p>Content coverage/knowledge & 21CC: I need to understand the learning outcomes of my students in this lesson:</p> <p>What is my objective (knowledge and skills) in this lesson?</p> <p>What activities can I design to achieve these objectives in a collaborative manner?</p> <p>How do I know if they have achieved if they are collaborating?</p> <p>What is the students' prior knowledge? Is that prior knowledge enough to the learning purpose?</p>	<p>Hard content skills versus soft skills: The collaboration process is as, if not more important, than the artefacts produced by the group while collaborating: Translated to practice, it is more important for the teacher and students to have a strong sense of what is happening in their interaction and collaboration is more important than making sure they get to the final product. Individual versus group: individual learners contributions and learning are representative of the collaborative process and artefacts. Individual reporting can be designed within CSCL but it should be compromise the collaborative work.</p> <p>Knowledge versus 21CC: Is critical thinking taking place? Are participants negotiating and sharing knowledge, solving problems together? To what extent are they aware and in control of their learning process?).</p>	<p>The analysis of individual and collective contributions; the relationships among contributions; and the analyses of students' reflection.</p> <p>Archival of process, e.g. qualitative analysis of students' postings</p> <p>Visualization of cues/clues of the cognitive dimension in students' writing (to provide a quantitative assessment of the qualitative 21CC). E.g. frequency of use of certain thinking scaffold such as "I need to understand"; frequency of use of keywords.</p> <p>Visualisations: of how students are working with one another, not just as a social activity but as a knowledge creating activity: of students' reflection.</p>	<p>Students are provided a 'report' of their collaborative activities</p> <p>How have I been sharing personal question/ideas in relation to the brainstorming and generative conversation phase.</p> <p>What is the consolidation of my own point of view in relation to class' problem; finding or proposing an agreement;</p> <p>What is the summary of the class at this point of time</p> <p>What are some of the disagreements and can I make a suggestions.</p> <p>reflection: how have I been learning?; planning further steps; constructive review of one's contribution in relation to group/class progress?</p>	<p>Teacher's reflecting on students' learning and planning the next activity ahead.</p> <p>What are the promising ideas or discussion reflected by the students?</p> <p>How are the hard skills developing along side the soft skills?</p>

MAIN CONFERENCE C1
ARTIFICIAL INTELLIGENCE IN EDUCATION/INTELLIGENT TUTORING SYSTEM
(AIED/ITS) AND ADAPTIVE LEARNING

- FULL PAPER -

LEVERAGING LSTM IN THE FINE-GRAINED ANALYSIS OF THE INCUBATION EFFECT IN PHYSICS PLAYGROUND..... 12
MAY MARIE P. TALANDRON-FELIPE & MA. MERCEDES T. RODRIGO

**NOVEL WRITING SUPPORT SYSTEM BY TARGET READERS’ STORY ARCS AND CHARACTERS’ EMOTIONAL CHANGES
..... 22**
ATSUSHI ASHIDA, MASATAKA TOKUMARU, TOMOKO KOJIRI

**CLASSIFICATION OF EMOTIONS IN PROGRAMMING FROM FACE AND LOG FEATURES USING REPRESENTATIVE
INTERVALS 32**
THOMAS JAMES TIAM-LEE & KAORU SUMI

USING DATA MINING TECHNIQUES TO ASSESS STUDENTS’ ANSWER PREDICTIONS..... 42
ALISA LINCKE, MARC JANSEN , MARCELO MILRAD, ELIAS BERGE

- SHORT PAPER -

**EFFECTS ON FOSTERING COMPUTATIONAL THINKING BY EXTERNALIZING A SOLUTION WITH CONSTRUCTION OF
A PROBLEM-SOLVING MODEL..... 51**
KAZUAKI KOJIMA & KAZUHISA MIWA

TOWARDS ADAPTIVE PROVISION OF EXAMPLES DURING PROBLEM SOLVING 57
FAIZA TAHIR, ANTONIJA MITROVIC & VALERIE SOTARDI

CAN EEG SIGNAL PREDICT LEARNERS’ PERCEIVED DIFFICULTY? 63
AMAN KUMAR, PANKAJ CHAVAN & RITAYAN MITRA

**IDENTIFYING SIGNIFICANT INDICATORS OF EYE-MOVEMENT AND EEG-BASED ATTENTION TO PREDICT READING
PERFORMANCE 69**
SONG LAI, BINGBING NIU, JIAQI LIU[^] & FATI WU

PROMOTING REFLECTION ON QUESTION DECOMPOSITION IN WEB-BASED INVESTIGATIVE LEARNING 75
YOSHIKI SATO, AKIHIRO KASHIHARA, SHINOBU HASEGAWA, KOICHI OTA, RYO TAKAOKA

MODELING VIDEO VIEWING STYLES WITH PROBABILISTIC MODE SWITCHING 81
HIROAKI KAWASHIMA, KOUSUKE UEKI & KEI SHIMONISHI

- POSTER PAPER -

EXTERNALIZATION SUPPORT FOR HYPOTHESES CREATION PROCESS OF DISCOVERY LEARNING IN BIOLOGY 87
TAKUYA WADA, TOYA OTAGAKI & TOMOKO KOJIRI

**CORRELATING WORKING MEMORY CAPACITY WITH LEARNERS’ STUDY BEHAVIOR IN A WEB-BASED LEARNING
PLATFORM 90**
ALISA LINCKE, DANIEL FELLMAN , MARC JANSEN , MARCELO MILRAD, ELIAS BERGE, BERT JONSSON

**SIMULATABLE OPEN LEARNER MODELS OF CORE COMPETENCIES FOR SETTING GOALS FOR COURSE
PERFORMANCE 93**
CHIH-YUEH CHOU, WEN-CHIEH CHIH, SHU-FEN TSENG & ZHI-HONG CHEN

INVESTIGATING FUNCTIONAL FIXEDNESS AMONG NOVICE STUDENT PROGRAMMERS..... 96
MAY MARIE P. TALANDRON-FELIPE & KENT LEVI A. BONIFACIO

Leveraging LSTM in the fine-grained analysis of the Incubation Effect in Physics Playground

May Marie P. TALANDRON-FELIPE^{ab*} & Ma. Mercedes T. RODRIGO^{a*}

^a*Ateneo De Manila University, Philippines*

^b*Central Mindanao University, Philippines*

*maymarie.talandron@upou.edu.ph, mrodrigo@ateneo.edu

Abstract: Incubation Effect (IE) refers to the phenomenon where one gets stuck in a problem-solving activity, decides to take a break, and afterwards revisits the unsolved problem and eventually solves it. While studies on IE were all limited to traditional classroom activities, this research aimed to continue the study of IE in the context of a computer-based learning environment and find features that would predict the incidence of revisiting an unsolved problem and its positive outcome. A prior IE model was developed using a logistic regression but the hand-crafted features used were from aggregated data and do not reflect specific characteristics of students' actions. Further analysis was conducted in this study and used a deep learning technique which significantly improved the performance of the IE model. In order to interpret the learned features of the neural network, a combination of dimension reduction, visualization technique, and clustering were used. It was found that the coarse-grained features are consistent with the fine-grained features but action level features were also discovered which provided more evidence that there was an improvement on how students tried to solve the problem after incubation.

Keywords: Incubation Effect, Physics Playground, LSTM

1. Introduction

Taking a break from a series of failed attempts to solve a problem may facilitate the solution process as shown in prior work (Fulgosi & Guilford, 1968; Gilhooly, Georgiou, & Devery, 2013; Penalzoza & Calvillo, 2012; Sio & Ormerod, 2015). This momentary break is known by the name incubation (Sio & Ormerod, 2015). During some incubation periods, an internal mental process associates new information with past information to generate solution ideas (Medd & Houtz, 2002). In the context of education, students who get stuck in a problem-solving activity may temporarily engage in another task, after which, they return to the original problem and find a solution. When the student solves the problem after incubation, the phenomenon and its positive result is called the Incubation Effect (IE).

IE is divided into three phases (Gilhooly et al., 2013): (a) pre-incubation phase, (b) incubation phase, and (c) post-incubation phase. The pre-incubation phase is when a student attempts to solve a problem and gets stuck. When the student decides to take a break from the problem-solving task and engages in another task, it signals the beginning of the incubation phase. The post-incubation phase happens when the student goes back to the original problem and tries to solve it again. The benefits of incubation prompted researchers to incorporate breaks into educational activities which were shown to have positive results (Lynch & Swink, 1967; Medd & Houtz, 2002; Rae, 1997; Webster, Campbell, & Jane, 2006). Prior studies (Ellwood, Pallier, Snyder, & Gallate, 2009; Fulgosi & Guilford, 1968; Gilhooly et al., 2013; Penalzoza & Calvillo, 2012; Sio & Ormerod, 2015) investigated specific factors that could lead to successful incubation in the context of classroom tasks and suggested that engaging in a different activity may produce a better outcome. On the other hand, (Penney, Godsell, Scott, & Balsom, 2004) claimed that engaging in a task with similar nature would promote priming which allows students to realize the correct solution to the problem but (Segal, 2004) said that the task during incubation has no effect on its outcome.

The incidence of incubation effect has also been investigated (Martinez et al., 2016; Talandron, Rodrigo, & Beck, 2017; (Talandron & Rodrigo, 2018) in the context of a computer-based environment called Physics Playground (PP) which is a two-dimensional game that is designed for high school

students to better understand concepts in Physics. (Martinez et al., 2016) found evidence that taking a break helped some students to solve a problem in which they were previously stuck. For example, if the student cannot solve level 3 in playground 3, the student can leave it and try to solve other levels then just go back to it again later. To further explore IE on PP, Talandron, Rodrigo, and Beck (2017) attempted to model IE and examined possible factors that predict the successful outcome of incubation. However, the study was limited to hand-crafted features from aggregated data which means that individual attempts comprising the 3 phases of IE were not analyzed in a fine-grained level and the actual activity during incubation were not taken into consideration. Thus, further analysis is needed to look at the large amount of unused data from the previous studies and come up with better features and model.

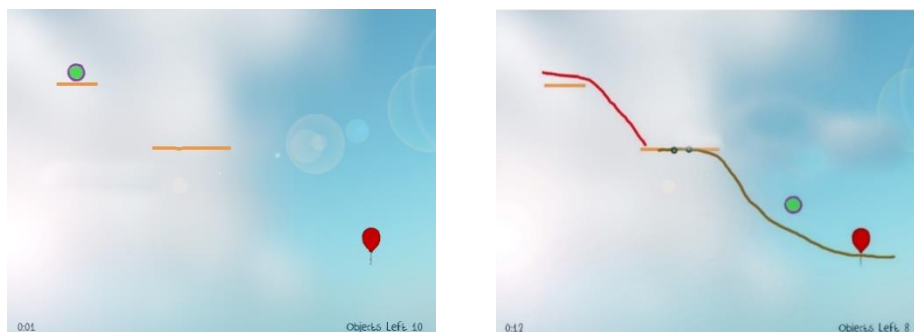
The main objective of the study is to conduct a fine-grained level modeling of the incubation effect among students playing Physics Playground and further understand the factors that predict IE. Student interaction logs contain a rich amount of data that can be utilized to extract information on student actions and behaviors in each attempt. Specifically, this study would like to answer the following:

(RQ1) What fine-grained features predict incubation effect? What are the features that more likely give a positive result during the post-incubation phase? What's the difference between the actions of the students to solve level X during pre-incubation and during post-incubation?

(RQ2) How will the extracted features perform against the hand-crafted features (Talandron et al., 2017) in predicting incubation effect? How does the model using hand-crafted coarse-grained features perform versus the model with fine-grained learned features?

2. Physics Playground

Physics Playground (PP), formerly known as Newton's Playground, is a two-dimensional computer-based game designed for students in the secondary level to better understand the concepts of qualitative Physics. The game simulates how the physical objects operate in relation to Newton's laws of motion: balance, mass, conservation and transfer of momentum, gravity, and potential and kinetic energy (Shute & Ventura, 2013). The game has different problems with varying levels of difficulty and solutions. The main objective of each problem is to guide a green ball to a red balloon. To solve each level, the players must draw objects (i.e., ramp, lever, pendulum, springboard) using the computer mouse and these objects become part of the game environment. Figure 1 (a) shows an example level of PP which requires a ramp to lead the ball to the balloon. All objects drawn obey the basic rules of physics relating to gravity and Newton's three laws of motion (Shute & Ventura, 2013). Once the player draws a ramp, the ball will then follow its path until it reaches the red balloon as shown in Figure 1 (b).



(a) A level that requires a ramp

(b) Sample solution to a ramp problem

Figure 1. A sample level in Physics Playground.

When the student solves a level, he/she receives either a gold or silver badge. A badge is awarded if the student solves the level – a gold badge if the problem was solved using at or below a par number of objects determined by the game designers; otherwise, a silver badge is given.

3. Methods

3.1 Data

The analyses for this study used a dataset collected from a total of 176 high school students in the Philippines: 29 from a public junior high school in Baguio City (School 1); 31 from a private university also in Baguio City (School 2); 56 from a private university in Cebu (School 3); and 60 from a private university in Davao City (School 4). These students were considered average in terms of their academic performance.

The students were given an orientation to introduce them to Physics Playground and to explain the game mechanics. Before playing, they were asked to answer a pre-test which was comprised of 16 multiple-choice type questions about simple machines and laws of Physics in relation to the learning objectives of PP. Then the students were given about 2 hours to play Physics Playground where their interactions with the game were automatically recorded into a log file. They were allowed to choose the problem they would like to solve, they could leave the problem, and return to it at a later time. After the session, the students answered a post-test which was also based on the topics covered in PP.

While playing PP, student's interactions with the game were automatically recorded along with each action's time stamp. The actions recorded were divided into 4 categories: Menu Events, Level Events, Play Events, and Agent Events. Menu Events refer to interactions when the player is in the main menu of the game while Level Events are actions related to each individual level within a playground. Play Events are the player's interactions within the PP environment once the player started to play a specific level. Agent Events refer to the interactions of and with the objects or simple machines drawn by the player to solve the level. These include the level, start time, end time, objects drawn, badge, etc. where we can derive other information as in prior work (Martinez et al., 2016; Palaoag, Rodrigo, & Andres, 2015; Palaoag, Rodrigo, Andres, Andres, & Beck, 2016; Talandron et al., 2017) such as attempt duration, number of restarts and revisits, sequence of levels, number of badges earned.

3.2 Fine-Grained vs. Coarse-Grained Features

To distinguish coarse-grained (Talandron et al., 2017) and fine-grained analysis, operationalizing IE in the PP interaction logs were done in a hierarchical manner. Figure 2 shows the levels of analysis as well as the relationship of the entities. The coarse-grained level analysis involved features from levels 1 to 3 of the diagram and the fine-grained analysis will include features on levels 4 and 5.

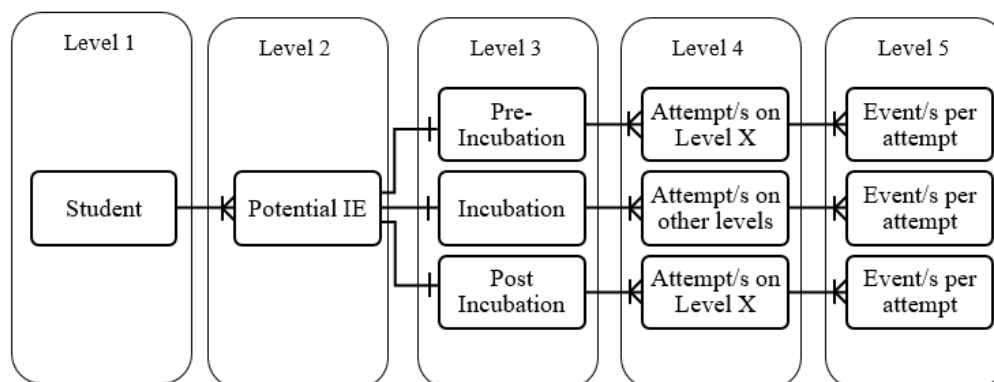


Figure 2. The levels of analysis.

The data was filtered to only include the interaction logs of students who exhibited potential IEs. Unnecessary columns were then removed and only included the problem ID, series of actions taken to solve the problem, and the result which indicates whether the student solved the problem or not. The canonical solution to each problem which was the basis of the problem type was also integrated into the logs.

3.3 Modeling

To prepare the data for modeling using LSTM, it is essential to structure the data for the specification of the timestep and batch size such that the timestep corresponds to the number of actions to solve a problem and the batch size as the number of problems or attempts per student. Another step was the transformation of the data type from string to numeric. String data must be encoded as numbers to be used as input or output for machine learning and deep learning models and the *Scikit-learn* library has provided the tool to do this. *Sklearn's LabelEncoder* module finds all classes and assigns each a numeric id starting from 0. For the output labels, *np.utils.to_categorical* was used to convert array of labeled data (from 0 to nb_classes-1) to one-hot vector.

The model using LSTM was developed using Keras, a high-level neural networks API on top of TensorFlow with Python as the underlying programming language. To realize the objectives, this study focused on the following training tasks:

- Given a sequence of attempts for a series of problems, classify each attempt based on the following:
 - son – solved new (successfully solved the problem at first attempt)
 - unn – unsolved new (failed to solve the problem at first attempt)
 - sops – solved a problem that has already been previously solved
 - unps – failed to solve a problem that has already been previously solved
 - sopu-nb – no incubation, replayed and solved a previously unsolved problem
 - unpu-nb – no incubation, replayed and failed to solve a previously unsolved problem
 - **sopu – revisited and solved a previously unsolved problem (IE-True)**
 - **unpu – revisited and failed to solve a previously unsolved problem (IE-False)**

The “sopu” corresponds to IE-True while “unpu” corresponds to IE-False. The analyses were focused on these 2 classes. The term “replay” means the player re-played the same level consecutively without interval while “revisit” means a “break” or interval occurred before the same level is played again.

The input features include the time of the action, the problem ID, and the series of actions in the attempts to solve the problem. The output label was the result of each attempt which was coded based on the 8 classes described in this section. For this task, a one-layer deep LSTM was used. To predict the output label, the hidden state at the last timestep was passed through a fully connected layer and a subsequent *softmax* layer. The batch size used was the number of attempts per student which was 91 and timestep was set to 100 so that the network would consider all the actions for each attempt as it backpropagates when calculating gradients for weight updates. The final number of epochs was set to 200. The values of other hyperparameters followed the configuration described in section 3.3. A student level cross-validation was used which was to ensure that a student’s data was only either on the training set or the testing set.

Another issue that had to be addressed was class imbalance. Majority (71%) were first attempts on a new problem and the remaining 29% were divided into 3 types: 13% replay previously solved problems, 9% revisit previously unsolved problem (potential IE), 7% replay unsolved problems (no incubation). This was addressed using the *sklearn.utils.class_weight.compute_class_weight* from the *Scikit-learn* library which computes for the appropriate weight based on the given training data. The computed values were stored in a dictionary which was then implemented during training. A student-level cross validation was done to ensure that each student’s data was either on the training set or the testing set.

3.4 Analysis of Features

The application of t-SNE (Maaten & Hinton, 2008) and X-means clustering algorithm as described in Wang et al. (2017) was done to map the input samples with the learned features. To answer the research questions previously stated, the data were analyzed as follows:

RQ1: What fine-grained features predict incubation effect? The learned representations are expected to be features that predict performance. The features derived from the input data was visualized using t-SNE as well as the prediction results. Then, X-means algorithm was used to cluster the data points in order to identify distinct groups. Quantitative analysis was conducted on each cluster

based on the features that could be derived from the LSTM input data to find significant differences or effect of these features on each cluster.

RQ2: How will the extracted features perform against the hand-crafted features in predicting incubation effect? The performance of the model and the coarse-grained features presented in Talandron, Rodrigo, and Beck (2017) were compared to the fine-grained model and the extracted fine-grained features.

4. Results and Discussion

4.1 The Fine-Grained Model

LSTM was used to model IE in a fine-grained level as described in section 3.3. A zero-padding technique was applied in order to achieve a uniform number of attempts per student which was used as the batch size during training and the same technique for the number of actions per attempt which was used the number of timesteps. The input vector has a total of 773,500 rows (85 students with 91 attempts each and each attempt with 100 actions). Model performance was measured based on the confusion matrix which is shown in Table 2 (recall = 91.62%, precision = 82.55%, f-score = 86.84%, kappa = 0.821).

Table 1

Fine-Grained IE model confusion matrix

Actual	Predicted		
	IE-True (<i>sopu</i>)	IE-False (<i>unpu</i>)	Others
IE-True (<i>sopu</i>)	175	2	14
IE-False (<i>unpu</i>)	15	95	22
Others	22	36	3251

Using t-SNE, both the actual and the predicted IE-True and IE-False instances were visualized on a two-dimensional plot. Several runs of t-SNE were conducted with varying values for perplexity and in order to decide on the most appropriate value, the one with the highest t-SNE nearest neighbor accuracy was selected. In this result, the perplexity of 20 yield the highest t-SNE nearest neighbor accuracy of 81%. The input data used in t-SNE includes all the features that could be derived from the LSTM input data and that could have been learned by the neural net such as time, problem, actions taken to solve the problem, duration of incubation, duration for each attempt, problem difficulty, productivity, and problem type. The t-SNE output reduced these features into two dimensions which were then the basis for the t-SNE plot. Figure 3 shows the t-SNE plot of IE-True (*sopu*) and IE-False (*unpu*) based on the actual (a) and prediction results (b).

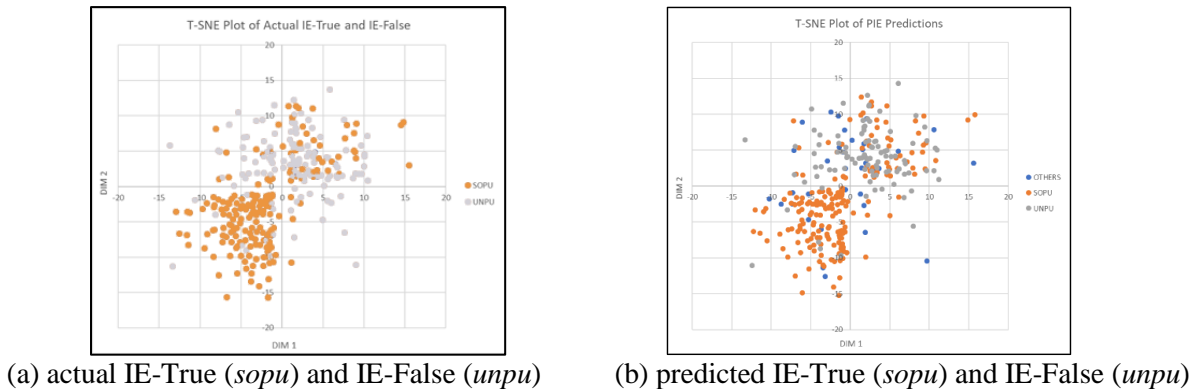


Figure 3. The t-SNE plot of IE-True (*sopu*) and IE-False (*unpu*).

Since the model’s recall was at 91.62%, it was expected that their t-SNE plot should look similar. More importantly, both plots show two apparent groups, one on the third quadrant of the plot which was composed of almost all IE-True instances, and one on the first quadrant which includes majority of IE-False but was mixed with some IE-True. In order to get the distinct clusters, we applied x-means clustering on the t-SNE results and then plot the clustering results on the t-SNE data. This was done on both the actual and predicted IE-True and IE-False as shown in Figure 4(a) and Figure 4(b), respectively.

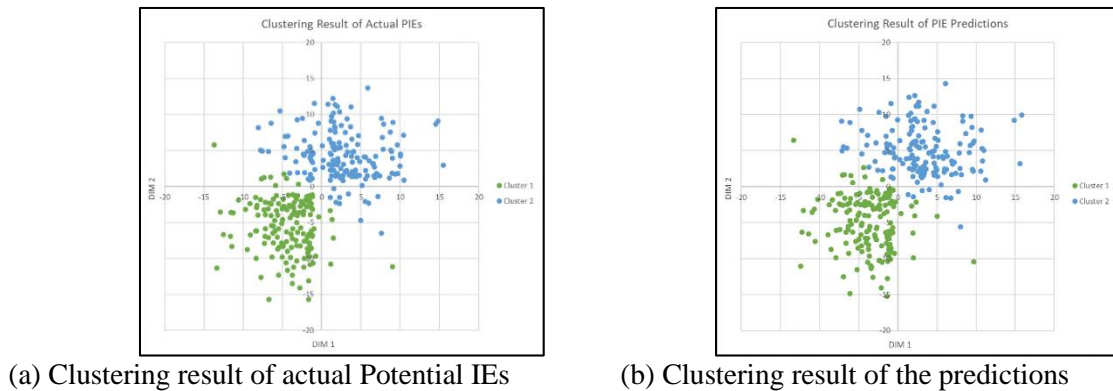


Figure 4. The clustering results of both actual IEs and predicted IEs

Based on the clusters of prediction results and referencing Figure 3(b) on Figure 4(b), Cluster 1 is predominantly composed of SOPU predictions (IE-True). A quantitative analysis was done for all the features derived from the LSTM input data in order to extract distinct features for the clusters. As previously mentioned, these features include the time of the revisit, incubation duration, problem difficulty, student’s productivity, the problem type, and the different actions done to solve the problem.

4.1.1 Time of Revisit and Problem Difficulty

It was found that more potential IEs (revisit) occurred during the last 30 minutes of the session. However, when investigated in terms of how productive revisits were at each 30-min interval, it was found that more revisits at the later time resulted to IE-False. A t-test was conducted to compare the difference between the time of revisit in cluster 1 vs those in cluster 2. There was a significant effect of the time of revisit on the clusters at the $p < 0.05$ level [$F(1, 321) = 7.01, p = 0.008$]. The IE incidence for each cluster in 30-minute bins is also shown in Figure 5. Since instances in cluster 1 were mostly successful IEs, it can be inferred that potential IEs in the early part of a time-limited session are more likely to be beneficial.

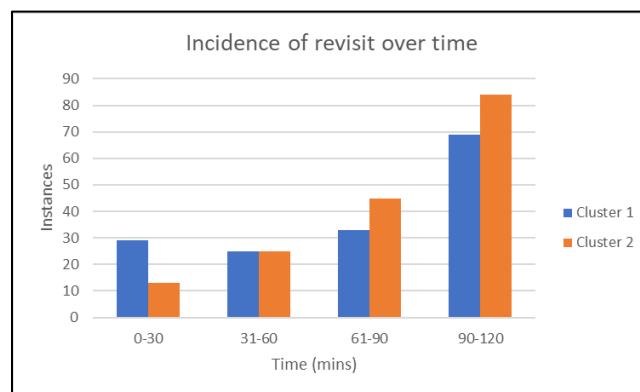


Figure 5. Predicted revisit instances at each time interval from clusters 1 and 2

Also, from the coarse-grained IE model (Talandron et al., 2017), problem difficulty was a significant feature such that revisiting a problem with lower difficulty rate more likely results to IE-True. This was also analyzed in a fine-grained level and the difference was significant on cluster 1

(mean = 39.36%, sd = 16.64%) and cluster 2 (mean = 45.39%, sd = 17.33%) at the $p < 0.05$ level [$F(1, 321) = 10.06, p = 0.002$].

4.1.2 Duration of Incubation

The incubation duration between clusters were compared and a significant difference was found between cluster 1 (mean = 10.95, sd = 17.99) and cluster 2 (mean = 16.06, sd = 20.81) at the $p < 0.05$ level [$F(1, 321) = 5.52, p = 0.02$]. This finding is also consistent with the coarse-grained analysis where it was found that a lengthy incubation could lead to IE-False (Talandron et al., 2017) and specifically, incubation duration that was more than 40 minutes resulted to IE-False (Talandron and Rodrigo, 2018).

4.1.3 Productivity Rate

Another notable feature in predicting IE-True in the coarse-grained analysis (Talandron et al., 2017) was the student's productivity which means how well the student was performing at the time of revisit. Similarly, a comparison was made in the fine-grained level analysis to see how this feature affected the clusters. This was computed as the number of problems solved over all attempts made at the time of revisit. The difference was significant between cluster 1 (mean = 64.48%, sd = 17.89%) and cluster 2 (mean = 56.97%, sd = 19.75%) at the $p < 0.05$ level [$F(1, 321) = 12.30, p < 0.001$]. This means that having been productive by the time of revisit would more likely lead to a positive revisit outcome. Figure 6 shows the potential IE instances in clusters 1 and 2 and the student's productivity rates in 25% bins.

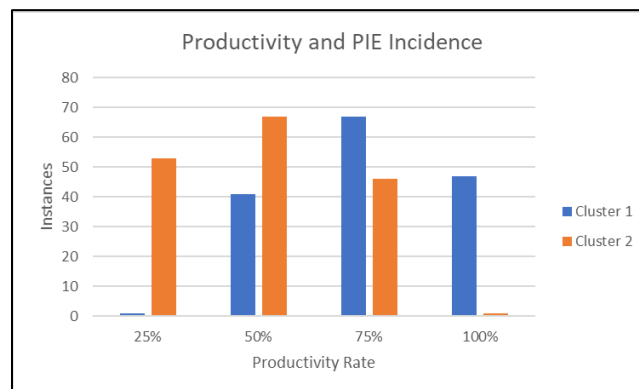


Figure 6. Productivity of IE instances from clusters 1 and 2

4.1.4 Problem Type

The problem type is a binary feature which indicates whether the attempt preceding the revisit had a similar solution to the problem during the revisit. There were 208 potential IEs preceded by an attempt on a problem with a similar solution and 125 (60.10%) resulted to IE-True. In terms of its effect on clusters 1 and 2, the Chi-Square test of independence was conducted and found a significant effect, $\chi^2(1, N=323) = 39.55, p < 0.001$, between the clusters and the problem type. This means, being preceded by a problem with a similar solution was considered as a significant factor in predicting IE-True.

4.1.5 Actions

There's a total of 22 events or actions from the logs and some of these actions can be found on other games such as 'click', 'pause', 'hover/preview tutorial', 'watch tutorial', 'erase'. Other actions are very specific to Physics Playground such as 'collision', 'diver', 'draw pin', 'lever', 'pendulum strike', 'springboard', 'ramp', 'stacking', 'stacking warning', 'draw freeform'.

The analysis for these actions was divided into three: 1) an analysis on the common game actions between cluster 1 and cluster 2; 2) an analysis on the common game actions between the pre-incubation period and post-incubation period; and 3) an analysis on solution specific actions during the pre-incubation period and post-incubation period to explore whether there has been a change on the

approach to solve the problem. For the common game actions, a comparison on the frequency of these actions between the clusters were analyzed and a significant effect, at $p < 0.05$ level, was found on the incidence of pause, hover tutorial, and erase between cluster 1 and cluster 2 as shown in Table 3.

Table 2

Comparison of common game actions between clusters 1 and 2

Actions	Cluster 1 (mean)	Cluster 2 (mean)	F	p
Erase	10.89	6.55	7.77	<0.01
Hover Tutorial	0.41	0.11	4.68	<0.05
Pause	0.39	1.21	143.42	<0.001

Based on this, we can infer that higher incidence of erase in cluster 1 means better awareness when incorrect actions were made, lower incidence of pause can be an indication that they were more confident of what they're doing and higher incidence of hover tutorial means they are making sure that they are drawing the object correctly. Further, a comparison of these actions between the pre-incubation and post-incubation period was conducted to see if there was a difference before and after incubation. The difference was significant for pause between cluster 1 pre-incubation (mean = 0.08) and cluster 1 post-incubation (mean = 0.03) at the $p < 0.05$ level [F (1, 381) = 109.19, $p < 0.001$]. It was also significant for erase, cluster 1 pre-incubation (mean = 0.03) and cluster 1 post-incubation (mean=0.05) at the $p < 0.05$ level [F (1, 381) = 6.72, $p = 0.009$]. Third, solution specific actions were compared between pre-incubation and post-incubation and a significant difference, at the $p < 0.05$ level, was found for ramp and springboard as shown in Table 4.

Table 3

Comparison of solution-specific actions between clusters 1 and 2

Actions	Pre-incubation (mean)	Post-incubation (mean)	F	p
Ramp	0.02	0.07	14.72	<0.001
Lever	0.015	0.017	0.085	0.77
Springboard	0.007	0.025	7.87	0.005
Pendulum	0.010	0.014	0.88	0.35

With the significant difference on the frequency of these actions during the pre-incubation period and post-incubation period, it can be inferred that incubation has made an impact on how the students try to solve the problem during the revisit.

4.2 Coarse-Grained Model vs Fine-Grained Model

The findings were compared to establish consistency of results as well as the necessity of fine-grained analysis to improve model performance. All the significant features from the coarse-grained analysis (Talandron et al., 2017) are consistent with the extracted features in the fine-grained analysis which are the time of revisit, incubation duration, productivity before revisit, and problem difficulty. Aside from these, new fine-grained features were discovered during the analysis which are the problem type, common game actions and solution-specific actions which provided evidence that incubation improved students' actions in their attempt to solve a previously unsolved problem after incubation.

The fine-grained model performed better in all metrics used (precision, recall, f-score, kappa) compared to the coarse-grained model as shown in table 5. There was a notable increase of 31.82% in precision and 5.97 in kappa.

Table 4

Comparison of model performance

IE Model	precision	recall	f-score	kappa
Coarse-grained model (Talandron et al., 2017)	50.73%	89.61%	64.78%	0.224
Fine-grained model	82.55%	91.62%	86.84%	0.821

This improvement could be attributed to the problem-level or attempt-level data that the fine-grained level analysis was able to use. The difference between the student's actions before and after incubation was a significant factor on the model's prediction performance.

5. Conclusion, Limitations, and Future Work

This study investigated and modeled the incubation effect phenomenon among students playing an educational game in a fine-grained level. The initial investigation found that students' IE success rates matched their non-IE success rates, implying that IEs may indeed benefit students who are stuck (Martinez et al., 2016). This study continued the work and answered the following:

RQ 1: What fine-grained features predict incubation effect?

The significant features found were time of revisit (low), duration of incubation (low), problem difficulty (low), student's productivity at the time of revisit (high), similarity with the preceding problem. In terms of game actions, the following were discovered: erase (high), pause (low), hover tutorial (high). For problem specific actions, improvement in the student's drawing of ramp and springboard were observed.

RQ 2: How will the extracted features perform against the hand-crafted features in predicting incubation effect?

The coarse-grained features (Talandron et al., 2017) that were manually engineered were consistent with the extracted features but action-level features were also discovered providing more evidence on the improvement of students' actions to solve the problem after incubation. Moreover, a notable increase of 31.82% in precision (82.55% vs 50.73% of the coarse-grained model) and 5.97 in kappa (0.821 vs 0.224 of the coarse-grained model) were achieved when compared to the previous model which used aggregated data.

This study contributes to researches using computer-based learning environments in studying phenomenon of a similar construct with IE since interaction logs of test subjects can be recorded automatically and hence more accurately. Second, findings quantified the pedagogical practice where teachers instruct students who are stuck at a problem to skip it and go back to it at a later time. It showed that incubation can be an effective technique in solving problems where activities performed during the break are similar or related tasks and the features extracted from this study could be translated to design features that could be used in other educational games to improve students' performance. Third, this study contributed to the growing applications of deep learning on educational data specifically by analyzing interaction logs from a computer-based learning environment to improve predictive models of behaviors and phenomena in the context of education. It has also shown that combining deep learning with other machine learning techniques such as dimensionality reduction, visualization, and clustering, can pave way to understand the learned features of a neural network.

Based on the limitations of the data collection method which used stealth assessment, another experiment can be recommended which would involve a control group and an experimental group where one group is instructed and allowed to incubate and other is not to further study the benefits of incubation versus no incubation in an experimental setup. Second, the features that predict IE could be translated into a game design such as rules or game mechanics of an educational application or software to further study the integration of IE as a pedagogical strategy.

Acknowledgements

The authors would like to thank Mr. Joshua Martinez and Mr. Jun Rangie Obispo for their valuable contribution in the initial investigation of incubation effect in Physics Playground which paved way for this study and Prof. Valerie Shute for her assistance on questions about Physics Playground specifically about the canonical solutions. We also would like to thank the Ateneo Laboratory for the Learning Sciences and the Ateneo de Manila University Department of Information Systems and Computer Science. This study was made possible through a grant from the Philippines' Department of Science and Technology Philippine Council for Industry, Energy and Emerging Technology Research and Development entitled "Stealth assessment of student conscientiousness, cognitive-affective states, and learning using an educational game for Physics."

References

- Ellwood, S., Pallier, G., Snyder, A., & Gallate, J. (2009). The Incubation Effect: Hatching a Solution? *Creativity Research Journal*, 29(1), 6–14. <https://doi.org/10.1080/10400410802633368>
- Fulgosi, A., & Guilford, J. P. (1968). Short-term Incubation in Divergent Production. *American Journal of Psychology*, 81(2), 241–246.
- Gilhooly, Georgiou, G., & Devery, U. (2013). Incubation and creativity: Do something different. *Thinking & Reasoning*, 19(2), 137–149.
- Lynch, M. D., & Swink, E. (1967). Some effects of priming, incubation and creative aptitude on journalism performance. *Journal of Communication*, 17(4), 372–382.
- Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.
- Martinez, J. C., Obispo, J. R. C., Talandron, M. M. P., & Rodrigo, M. M. T. (2016). Investigating the Incubation Effect among Students playing Physics Playground. *24th International Conference on Computers in Education*, 1–8. Mumbai, India: Asia-Pacific Society for Computers in Education.
- Medd, E., & Houtz, J. C. (2002). The effects of facilitated incubation on fourth graders' creative writing. *Educational Research Quarterly*, 26(2), 13.
- Palaoag, T. D., Rodrigo, M. M. T., & Andres, J. M. L. (2015). *An Exploratory Study of Persistence Markers within a Game-based Learning Environment*. Presented at the 23rd International Conference for Computers in Education, China.
- Palaoag, T. D., Rodrigo, M. M. T., Andres, J. M. L., Andres, J. M. A. L., & Beck, J. E. (2016). Wheel-Spinning in a Game-Based Learning Environment for Physics. *International Conference on Intelligent Tutoring Systems*, 234–239. Springer.
- Penaloza, A. A., & Calvillo, D. P. (2012). Incubation provides relief from artificial fixation in problem solving. *Creativity Research Journal*, 24(4), 338–344.
- Penney, C. G., Godsell, A., Scott, A., & Balsom, R. (2004). Problem variables that promote incubation effects. *The Journal of Creative Behavior*, 38(1), 35–55.
- Rae, C. M. (1997). The creative power of doing nothing. *Writer*, 110(7), 13–15.
- Segal, E. (2004). Incubation in Insight Problem Solving. *Creativity Research Journal*, 16(1), 141–148.
- Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.
- Sio, U. N., & Ormerod, T. (2015). Incubation and Cueing Effects in Problem-Solving: Set Aside the Difficult Problems but Focus on the Easy Ones. *Thinking and Reasoning*, 21(1), 113–129.
- Talandron, M. M. P., & Rodrigo, M. M. T. (2018). Let's Take a Break: Analysis of the Incubation Effect Among Students Using a Learning Game for Physics. *Proceedings of the 26th International Conference on Computers in Education*. Presented at the 26th International Conference on Computers in Education, Philippines.
- Talandron, M. M. P., Rodrigo, M. M. T., & Beck, J. E. (2017). Modeling the Incubation Effect Among Students Playing an Educational Game for Physics. *Artificial Intelligence in Education: 18th International Conference on Artificial Intelligence in Education Proceedings*, 371–380. <https://doi.org/10.1007/978-3-319-61425-0>
- Wang, L., Sy, A., Liu, L., & Piech, C. (2017). Learning to Represent Student Knowledge on Programming Exercises Using Deep Learning. *10th International Conference on Educational Data Mining*. Presented at the International Conference on Educational Data Mining, Wuhan, China.
- Webster, A., Campbell, C., & Jane, B. (2006). Enhancing the creative process for learning in primary technology education. *International Journal of Technology and Design Education*, 16(3), 221–235.

Novel Writing Support System by Target Readers' Story Arcs and Characters' Emotional Changes

Atsushi ASHIDA^{a*}, Masataka TOKUMARU^b, Tomoko KOJIRI^b

^aGraduate School of Science and Engineering, Kansai University, Japan

^bFaculty of Engineering Science, Kansai University, Japan

*k088944@kansai-u.ac.jp

Abstract: Writing includes such elements of communication skills as expressing authorial intention in an understandable way that is acceptable to readers. Our research introduces a writing activity that cultivates communication skills by introducing a method for writing novels that are liked by the estimated target readers. An aspect of novels that may reflect individual preferences is the emotional experience produced by them. Therefore, this research proposes a method for creating novels that are liked by target readers from an emotional viewpoint. We first adopt story arcs as a representation of a reader model that expresses the emotional changes that readers are expected to experience from a novel. Second, we define the candidates of emotions for designing scenes from the viewpoint of the emotions of characters. Last, we develop a support system that not only suggests candidates for story arcs and characters' emotions but also provides feedback when the emotions selected for characters fail to support the selected story arc.

Keywords: Communication skills, novel writing, story arcs, characters' emotions

1. Introduction

Students need "21st Century Skills" to succeed in the information age. Their importance as well as methods for teaching and cultivating them have been proposed (Griffin and Care, 2014). One category of the 21st Century Skills is the ability to communicate with others. This skill consists of various sub-skills; precisely representing what a student wants to convey, changing how one talks based on the communication targets, and understanding the intentions of others; such sub-skills must be cultivated to improve communication. Writing is an activity that needs two of these skills, such as precisely representing what a student wants to convey, and changing how one talks based on the communication targets (expected readers). We believe that if authors can obtain a method for designing writings based on their theme and their target readers, they can improve a part of their communication skills. Therefore, we introduce a writing activity as a training method for cultivating a part of communication skills. The objective of our research is to develop a support system with which authors can create the outline their writings based on themes and target readers. We focus on the writing of fictional novels, where authors need to design everything in a novel such as the settings, and the scenes.

When writing a novel, authors often create a skeletal outline of a novel, which is called plot (Hunter and Begoray, 1990). In writing training, students need to create a plot that can express their theme and that are liked by their target readers. Several methods have been studied to support plot creation. Nishihara and Miura (2015) proposed a system that simplifies the creation of plots by defining the candidates of character behaviors and showing the results of changes in the novel's world affected by them. Akimoto and Ogata (2011) established a method that automatically generates a novel from characters and their roles based on Propp's Character Theory (Propp, 2010). These studies enable authors to construct novels with consistent stories based on Propp's Character Theory. Also, Watanabe and Arasawa (2014) analyzed components of novels and developed a novel-composition support system by having authors fill in the components. However, these studies ignore themes and target readers.

Our study develops a system that encourages students to write novels based on their themes and target readers. We previously proposed a support method for deriving a story that can express intended

themes (Ashida and Kojiri, 2018) and effectively derived the settings and scenes that can convey themes. The current research seeks to create the novels to be liked by the expected target readers. Cheong and Young (2006) proposed a system that automatically generates a story that readers might like using planning techniques. It assumes that the suspense that readers do not know what will happen next is one element that satisfies readers. It defines suspense as the number of possible scenes that readers can assume and developed a system that generates a story that evokes an enjoyable level of suspense by arranging the given scenes using inference techniques. This system requires authors to create scenes beforehand without considering their target readers. However, because the liked scenes are different based on the readers, created story are not always liked by the target readers. In addition, it automatically creates a story and does not support authors to create stories.

This research supports authors who want to create a sequence of scenes that will be enjoyed by their target readers. One factor that may satisfy them is the emotions they experience while they are reading. From this viewpoint, our research introduces a story arc that expresses the expected transition of readers along with the story and proposes a method that utilizes the story arc as a reader model. In addition, since readers usually empathize with specific characters, the emotions of characters should reflect the story arc. This paper proposes a method that supports authors who are establishing the emotions of their characters based on the reader model and provides feedback if gaps are found between the characters' emotional changes and the selected story arc.

2. Approach

The steps for creating scenes to compose a story that follows the reader model and approaches for supporting each step are shown in Figure 1. Zillmann's Mood Management Theory (Zillmann, 1988) reported that "people acquire various emotions from the media and actively select media in consideration of the emotions that they want to acquire." Choosing a novel is similar; readers select novels based on the emotions that they want to experience, including happiness, sadness, excitement, and anticipation. The feeling after reading is often derived both from a novel's last scene and the aggregation of scenes; authors first need to select the emotional changes that their target readers are expected to experience as they read.

In this step since many authors lack experience consciously imagining their expected readers, explicitly defining a reader model is difficult. We define it as the emotional changes experienced by readers while they are reading a novel. To support authors who want to develop reader models, we establish several possible patterns of emotional changes as candidates for reader models from which authors can select.

Emotional changes in readers are produced by the flow of specific scenes. Readers often empathize with characters while they are reading (Oatley, 2002). In this paper, as one of the empathizing, we assume that readers experience the same emotions of characters; thus the characters' emotional changes throughout the scenes should be defined carefully to derive the expected emotional changes in target readers. However, in this step, some authors fail to understand that they must consider the characters' emotional changes or know which emotions should be defined to follow a reader model. We propose a support method that presents candidates for the character's emotions and provides feedback if the defined characters' emotions do not elicit the intended reader emotions in the reader model.

When the emotions of the characters are determined, authors are expected to establish concrete scenes that can express them. Our research currently does not support this step.

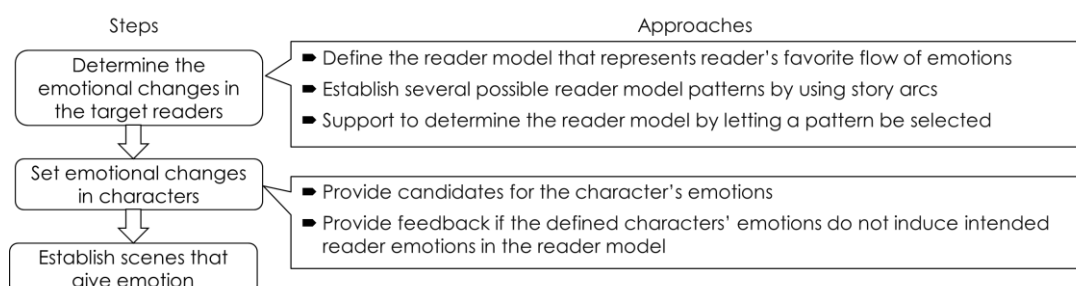


Figure 1. Steps for designing story and approaches of support system

3. Determination of Reader Model Using Story Arcs

Since authors often experience difficulty scrutinizing the emotional changes of readers, a simple representation of a reader model is desirable. Although many emotions exist, Cohn et al. (2009) argued that “happiness” is representative of positive emotions and “unhappiness” represents negative emotions, and emotional changes can be represented by the levels of happiness. Therefore, this research adopts the levels of happiness to represent the simple reader model.

A story arc expresses a story’s outline from an emotional aspect (Fig. 2). The horizontal axis corresponds to the story’s progress, and the vertical axis indicates the happiness level expressed by the story at that point. Reagan et al. (2016) reported that the story arcs of stories written in English can be classified into six types by clustering a huge number of story arcs using machine learning techniques. We adopt the six story arcs defined by Reagan et al. as candidates for reader models.

Table 1 lists the types of story arcs defined as candidates of reader models along the basic story structure, which consists of introduction, development, turn, and conclusion. This structure is often used in Japanese creative writing. The model represents the types of story arcs by their transitions, not by the absolute values of the happiness levels, because we believe that how readers feel reflects the differences in the story’s emotions. We provide authors with candidate reader models that allow them to choose the one that is most likely to be preferred by their target readers.

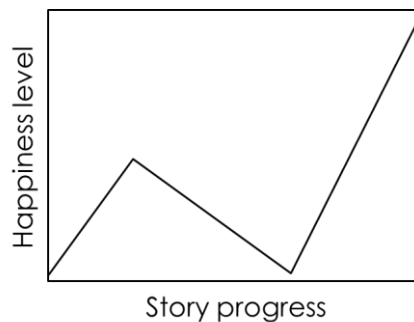


Figure 2. Example of story arc

Table 1. Types of Story Arcs as Reader Models

Types	Changes in happiness levels in story arcs		
	Introduction to development	Development to turn	Turn to conclusion
Type 1	Increase	Increase	Increase
Type 2	Decrease	Decrease	Decrease
Type 3	Increase	No change	Decrease
Type 4	Decrease	No change	Increase
Type 5	Increase	Decrease	Increase
Type 6	Decrease	Increase	Decrease

4. Determination Support of Characters' Emotions Following Reader Models

4.1 Candidate Emotions

To create preferred stories for target readers, authors need to design the emotions of characters based on selected reader models. For supporting authors who want to set the character's emotions, we provide candidates of emotions and encourage authors to select from them.

Plutchik (2001) proposed a model of human emotions called Plutchik's wheel of emotions. This model argues that human emotions can be expressed by both basic and complex emotions that occur as a combination of basic emotions. Figure 3 shows the eight basic emotions placed on a wheel. Those at the opposite pole are unlikely to occur simultaneously. The combination of two basic emotions other than the one at the opposite end creates a complex emotion called a *dyad*. In this research, eight basic emotions and 24 dyads are provided as emotion candidates.

Readers often feel emotions from empathized characters (Oatley, 2002). However, they do not always empathize with the protagonist. Some readers may prefer friend of the protagonist, or the cool villain. Our research allows authors to create characters that give the emotional changes of a reader model for each section of the story, including introduction, development, turn, and conclusion. Then it provides candidate emotions to authors and helps them consider the emotions of the characters for each section of the story to fit the defined reader model.



Figure 3. Basic emotions of Plutchik's wheel of emotions (Plutchik, 2001)

4.2 Feedback Based on Story Arcs and Setting of Characters' Emotions

4.2.1 Changes in Happiness Level by Emotional Changes

The selected characters' emotional changes should adhere to the reader model expressed by the story arc. However, since there are eight basic emotions and 24 dyads, authors do not always select the appropriate ones. Therefore, we propose a support method that compares the story arcs of the selected reader model and the happiness changes produced by the selected emotional changes and provides feedback to the authors when these two elements do not correspond. We conducted two surveys for determining the changes in happiness levels based on the selected emotional changes.

In the first survey, ten university students were given 32 words that represent emotions and ranked them based on happiness levels, where 1 is the highest and 32 is the lowest. Figure 4 shows the results. In Figure 4, The emotional words were arranged by their mean ranks from all the participants. The blue circles correspond to the mean ranks and the error bars show the standard deviation. Based on these results, we divided these emotional words into two groups: the happiness group, which consists of *joy* to *surprise* and the unhappiness group, which is formed by *sentimentality* to *despair*. However, ranking the emotional words within each group is difficult.

The second survey analyzed the ranks for each pair of emotional words within the happiness and unhappiness groups. Ten university students and an associate professor participated. We asked them in the second survey to group the emotional words into two or three classes based on the happiness levels.

Then for every pair, if the relations of the differences in the happiness levels are identical for more than six participants, these relations were set as their relative ranks. We set 35 relative relations in the happiness group and 102 in the unhappiness group.

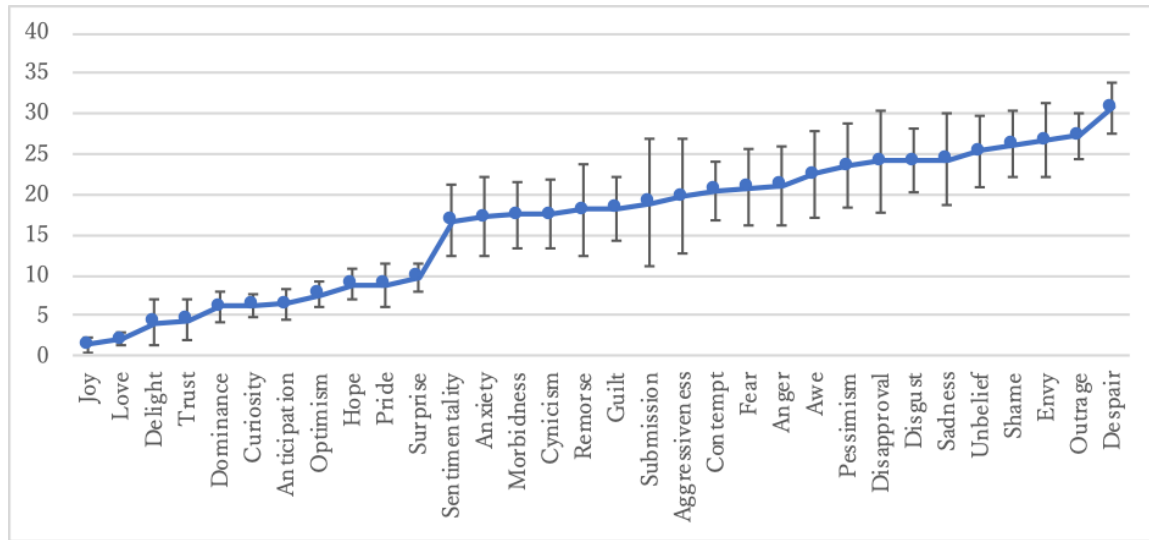


Figure 4. Ranks of emotional words by happiness levels

4.2.2 Feedback Generation

Based on the differences in the happiness levels between two emotional words defined by the surveys, the system provides feedback when the selected characters' emotional changes are inappropriate for the reader model. That is, for two sections of the story, such as introduction, development, turn, and conclusion, the system compares the transition of the happiness levels between the empathized characters' and those in the selected story arc and comments if their transitions are not the same.

Next we explain an example of feedback generation. Assume that the author selects the Type 4 story arc in Table 1 as a reader model and defines the emotions of character A and B, as shown in Figure 5. The transitions of the happiness levels derived from the characters' emotions and the reader model are shown in Table 2. If an author believes that the reader has empathy for character B, character B's transition of happiness levels between the turn and the conclusion does not correspond to the reader model, which indicates *increase* and character B's emotional change shows *decrease*. Therefore, the system provides feedback: "Note the happiness level contradiction between the reader model and character B's emotions in the turn to the conclusion."

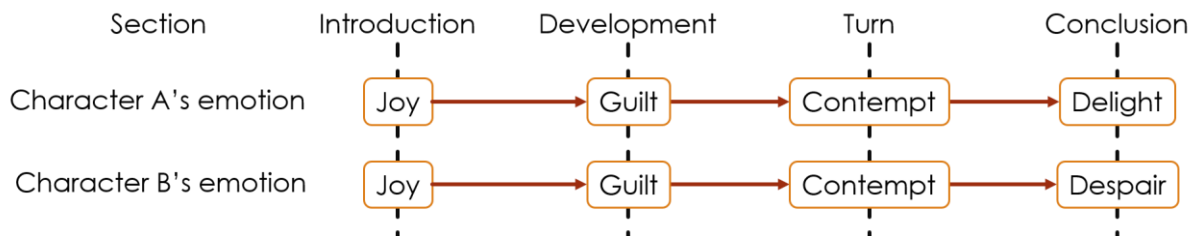


Figure 5. Example of characters' emotions

Table 2. Transitions of Happiness Levels by Characters' Emotions and Reader Model

		Changes in happiness level		
		Introduction to development	Development to turn	Turn to conclusion
Character's emotion	Character A	Decrease	No change	Increase
	Character B	Decrease	No change	Decrease
Reader model (story arc)		Decrease	No change	Increase

5. Prototype System

We developed a prototype system with which authors can select story arcs as reader models and define the emotional changes of their characters. We implemented our developed system as a new module of a plot creation support system that we developed in previous research (Ashida and Kojiri, 2019). Our system defines the plot's format and provides an interface for organizing it based on the format. The system is implemented using HTML and JavaScript.

The system's interface of the system is shown in Figure 6. The upper half is *the story arc selection part* and the lower half is *the character's emotion setting part*. In the former, six types of reader models can be selected from *the list of story arcs*. When one reader model is selected from the list, its corresponding graph is depicted below.

In the character's emotion setting part, the characters are created by *the add character button* and the characters' emotions can be inputted for each section. 32 emotional words are provided by *the list of emotions*. When the basic emotions are selected, the selected emotion on the Plutchik's wheel above is highlighted. If a complex emotion is selected, the two basic emotions that compose the selected emotion are highlighted. The text areas below the list of emotions are used for organizing the ideas of the scenes. A character with whom the readers might empathize is defined at *the relationship setting button*.

When the feedback button is clicked, the selected story arc and the selected emotional words are compared. When differences are detected in the changes of the happiness level, the feedback in Figure 7 is presented in *the feedback area* in Figure 6.

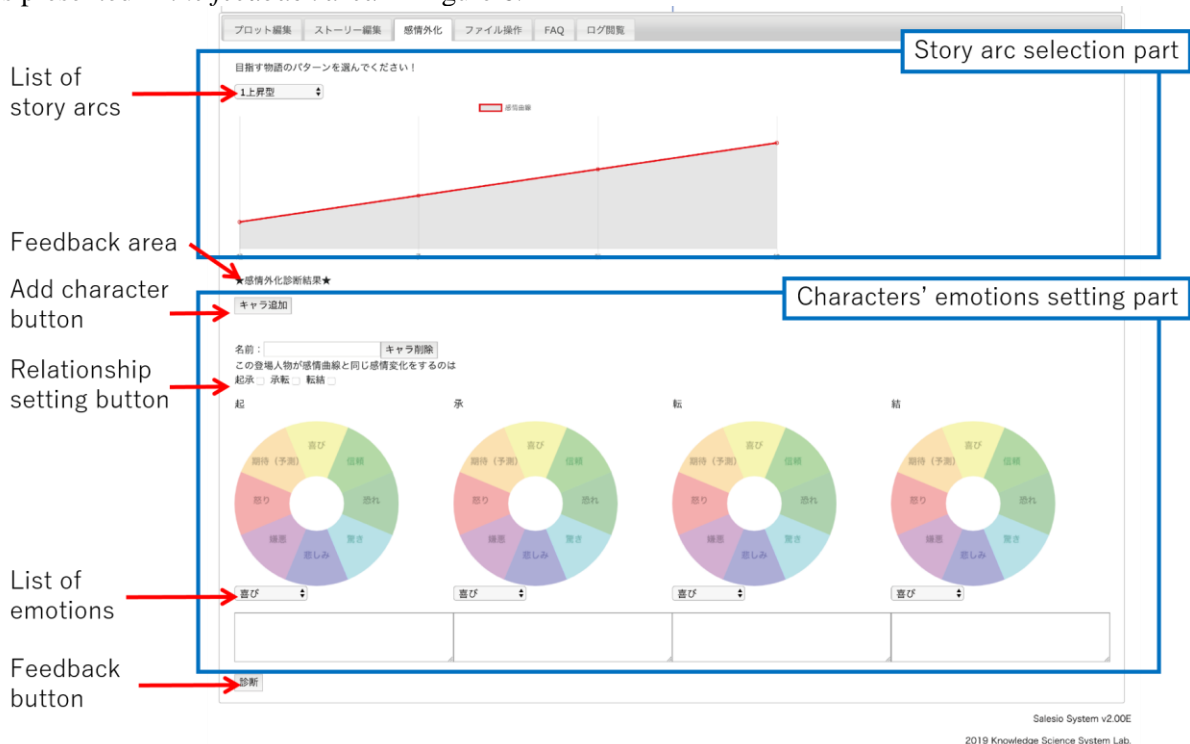


Figure 6. Interface of prototype system

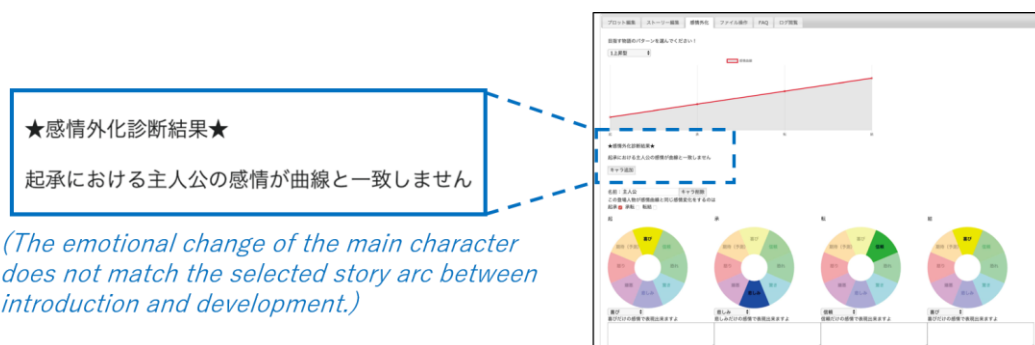


Figure 7. Example of feedback message

6. Experiment

6.1 Evaluation Experiment

6.1.1 Experimental Setting

To evaluate the effectiveness of our developed system for plot creation, we conducted an evaluation experiment with 11 university students. According to the pre-questionnaire, we found out that the participants' writing experience and motivation were different. They were divided into two groups: an experimental group of six participants and a control group of five participants.

The following is the experimental procedure:

1. Determine target readers;
2. Create a plot;
3. Answer questions about the created plot;
4. Answer a questionnaire.

In Step 1, the participants determined an age group and gender for their targeted readers from the following choices. The age group choices were children, junior- or senior-high school students, and adults. The gender choices were male and female. In Step 2, the control group participants created a plot using only the plot creation interface of the plot creation support system, which provides its format (Ashida and Kojiri 2019). The experimental group participants determined the story arc and their characters' emotions with the developed system before using the plot creation support system. In Step 3, they explained why they created the scenes in their plots. For Step 4, the questionnaire items shown in Table 3 were given to the experimental group and those in Table 4 were given to the control group. Regarding author confidence in the created plot, Item 1 is shared by both groups. We asked this question to check whether the proposed method helped create plots that are preferred by the target readers. Items 2 and 3 in Table 3 evaluate whether the story arc's setting and the characters' emotions contributed to the plot creation. Item 4 in Table 3 questions the appropriateness of the feedback provided by the system.

Table 3. *Questionnaire Items for Experimental Group*

Item	Question	Choices
1	Are you confident that your created plot will be preferred by your target readers?	1. Yes. 2. I have some confidence. 3. I have very little confidence. 4. No.
2	Did choosing a story arc help you create the plot?	1. Yes. 2. No. 3. Can't decide.
3	Did setting the characters' emotions based on the selected story arc help create the plot?	1. Yes. 2. No. 3. Can't decide.
4	Was the feedback proper?	1. Yes. 2. No 3. I received no feedback.

Table 4. *Questionnaire Items for Control Group*

Item	Question	Choices
1	Are you confident that your created plot will be preferred by your target readers?	1. Yes. 2. I have some confidence. 3. I have very little confidence. 4. No.

6.1.2 Results

First, we evaluated whether our proposed system contributed to the smooth creation of plots. If they can be created smoothly, Modification of scenes can be reduced. Table 5 shows the number of times that scenes were modified, the total number of scenes, and the average number of modifications per scene. Participant D is blank because we failed to obtain his log data. The results show that the average number of modifications per scene for the experimental group is significantly less than for the control group: t test ($t(5) = -2.698, p < 0.05$).

The questionnaire results are shown in Tables 6 and 7, indicating the number of participants who selected each choice. According to the answer for Item 1, all the participants in the experimental group were confident of their plots to some extent, while some in the control group were not. Items 2 and 3 in Table 6 show that most of the participants in the experimental group found that selecting the story arc and the characters' emotions supported their plot creation.

These results indicate that our system, especially its functions for selecting story arcs and characters' emotions, contributed to the smooth creation of plots in which participants felt confident. According to Item 4 in Table 6, only one participant in the experimental group answered "no." This suggests that the method of feedback generation helped some of the authors. Since the number of participants of our experiment was small, further experiments are needed to verify the effectiveness of the system's feedback.

Table 5. Results of Evaluation Experiment

	ID	Number of times scenes were modified	Number of scenes in created plot	Average number of modifications per scene
Experimental group	A	20	19	1.05
	B	11	13	0.85
	C	12	13	0.92
	D	-	14	-
	E	8	12	0.67
	F	5	10	0.50
Control group	G	20	15	1.33
	H	12	11	1.09
	I	18	13	1.38
	J	29	19	1.53
	K	38	14	2.71

Table 6. Questionnaire Results of Experimental Group

Item \ Choices	1	2	3	4
	1	1	5	0
2	4	2	0	
3	5	0	1	
4	2	1	2	

Table 7. Questionnaire Results of Control Group

Item \ Choices	1	2	3	4
	1	0	3	2

6.2 Verification Experiment

6.2.1 Experimental Setting

We experimentally verified whether the authors created plots that represent the selected story arc using our system. Five university students participated as *reader participants* to differentiate them from the participants in the first experiment.

The reader participants were asked to read the plots created by the participants in the experimental group of the former experiment and to select the happiness level transition that they felt by the plots from the given choices for each section, such as introduction to development, development to turn, and turn to conclusion. The given choices were increase, decrease, no change, and cannot answer. We evaluated whether the selected transitions of the happiness levels correspond to the story arcs selected by the authors of the plots as reader models.

Since the age-group and gender of each reader participant did not correspond to the target readers of the plots, we did not verify whether they preferred the created plots.

6.2.2 Results

Table 8 shows the number of reader participants whose selected happiness level transitions are the same as those of the story arc transitions that the authors selected. The accuracy rate indicates the ratio of correct answers among the total number of answers. Based on the accuracy rates, participants A, B, E, and F created plots that provided the expected transitions of emotions to the reader participants.

We examined the correlations between the accuracy rates and the answers for Items 2 and 3 in Table 6 by setting the following values to the answers: yes = 1, no = 0, and can't decide = 0.5. We identified a very strong correlation between the accuracy rate and the answers to Item 2 ($r = 0.94$) and a strong correlation between the accuracy rate and Item 3 ($r = 0.60$). These results indicate that those who utilized the selection of the story arc and the characters' emotions created plots that influenced the reader participants as the author intended. However, we failed to prove whether stories, which were created from the plots, are preferred by the target readers. Thus, we must conduct further experiments to investigate whether the stories created using our system satisfy target readers.

Table 8. *Results of Verification Experiment*

Plot	Number of correct answers			Accuracy rate
	Introduction to development	Development to turn	Turn to conclusion	
A	5	2	5	80%
B	1	5	5	73%
C	5	1	0	40%
D	1	0	5	40%
E	1	5	5	73%
F	4	5	5	93%

7. Conclusion

We proposed a method that creates plots for novels that are preferred by target readers. To represent reader models, we defined a story arc as an element that expresses the emotional changes that readers are expected to experience while reading the novel. In addition, for designing scenes from the viewpoint of characters' emotions, we defined the emotion candidates and the differences in their happiness levels.

We also developed a support system for creating plots based on the determination of reader models by selecting story arcs from multiple story arc patterns and setting characters' emotions by providing emotion candidates. Our experimental results suggest that the proposed system can support the smoothly creation of a plot that authors are confident about and that the created plots can express the selected story arc. However, our current experiments were conducted with too few participants. We need further experiments to verify our proposed system's effectiveness.

This system allows authors to freely set the emotions for each character. However, emotional dependencies are often found between characters. For instance, if a mother and child share a good relationship, the mother tends to feel happy when her child is happy. If such emotional dependency is ignored, deriving concrete scenes might be difficult. In our future work, we need to establish a method that allows authors to consider character relationships and possible emotions based on defined relations.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 18H03346.

References

- Akimoto, T., & Ogata, T. (2011). A consideration of the elements for narrative generation and a trial of integrated narrative generation system. *Proceedings of the 7th International Conference on Natural Language Processing and Knowledge Engineering*, 369-377.
- Ashida, A., & Kojiri, T. (2019). Plot-creation support with plot-construction model for writing novels. *Journal of Information and Telecommunication*, 3(1), 57-73.
- Ashida, A., & Kojiri, T. (2018). Question-based idea generation system for writing novels. *Proceedings of International Symposium on Educational Technology*, 165-167.
- Cheong, Y. G., & Young, R. M. (2006). A computational model of narrative generation for suspense. *Proceedings of the 21st National Conference on Artificial Intelligence*, 8-15.
- Cohn, M. A., Fredrickson, B. L., Brown, S. L., Mikels J. A., & Conway, A. M. (2009). Happiness Unpacked: Positive Emotions Increase Life Satisfaction by Building Resilience. *Emotion*, 9(3), 361-368.
- Griffin, P., & Care, E. (2014). *Assessment and teaching of 21st century skills: Methods and approach*. Springer.
- Nishihara, S., & Miura, M. (2015). A web-based application for writing novels. *Procedia Computer Science*, 60(1), 1014-1020.
- Hunter, W. J., & Begoray, J. (1990). A framework for the activities involved in the writing process. *Writing Notebook*, 7(3), 40-42.
- Oatley, K. (2002). Emotions and the story worlds of fiction. *Narrative Impact: Social and Cognitive Foundations*, 36-69.
- Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots. *American Scientist*, 89(4), 344-350.
- Propp, V. (2010). *Morphology of the Folktale: Second Edition*. University of Texas Press.
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1), 1-12.
- Watanabe, T., & Arasawa, R. (2014). Computer-supported novel composition based on externalization. *Procedia Computer Science*, 35(1), 1662-1671.
- Zillmann, D. (1988). Mood management through communication choices. *American Behavioral Scientist*, 31(3), 327-340.

Classification of Emotions in Programming from Face and Log Features Using Representative Intervals

Thomas James TIAM-LEE* & Kaoru SUMI

Future University Hakodate, Japan

*g3117002@fun.ac.jp

Abstract: This paper discusses a machine learning approach for classifying student emotions while doing programming exercises. Detection of academic emotions in programming from face features has previously been shown to be a difficult task because people don't tend to display as much expression as compared to more social activities. In our approach, we show that adding log features in addition to face features can improve the performance of classifiers. Furthermore, we show that identifying representative intervals of each emotion type based on human annotations can be used to build models to classify emotion over longer periods of time. We believe that our study can contribute in the development of better intelligent programming tutors that can respond to the affective state of students.

Keywords: Affective computing, Student modelling, Education

1. Introduction

Recent advances in artificial intelligence have allowed for intelligent tutoring systems to respond to the affective states of students. Many of these so-called "affective tutoring systems" have been motivated by several studies that correlate the emotional experience of students to their achievement (Rodrigo et al., 2009; Daniels et al., 2009) and self-regulated learning (Mega, Ronconi, & De Beni, 2014; Cho, & Heron, 2015). The ability to observe and recognize the student's affective states is part of why human tutors are very effective (Lehman et al., 2008). Likewise, intelligent tutoring systems that can respond to student emotions have been shown to yield a variety of benefits in the students' learning (Malekdazeh, Mustafa, & Lahsasna, 2015).

A key part of these types of systems is the automatic detection of student affect. A comprehensive meta-analysis of various approaches for affect detection can be found in the work of D'Mello and Kory (D'Mello, & Kory, 2015). A number of these studies used cameras are used to extract facial features to detect basic emotions such as joy, sadness, and anger (Zatarain-Cabada et al., 2015). However, in addition to these basic emotions, there has also been an interest in detecting what are referred to as "academic emotions" (Pekrun et al., 2002). These are emotions that are prevalent in academic settings, such as engagement and boredom. In AutoTutor (D'Mello, & Graesser, 2012), the student learns by interacting with a virtual conversational tutor avatar while an affect detector module uses a combination of posture, face, and log data to infer academic emotions. In VALERIE (Palarie et al., 2005), academic emotions are detected using a wide array of sensors including a video camera, a speech recognizer, and a psychological signal sensor.

However, there are some domains in which detection of affect has remained difficult. One of these is the domain of programming. Intelligent programming tutors (IPT) are a subclass of intelligent tutoring systems that teach programming. IPTs pose challenges that are not present in traditional ITSs (Crow, Luxton-Reilly, & Wuensche, et al., 2018). In these systems, students spend the majority of their time writing, testing, and debugging code instead of interacting with a virtual agent. Despite this, previous studies have shown that students frequently transition between different academic emotions while taking part in this kind of activity (Bosch, D'Mello, & Mills., 2013; Bosch, & D'Mello, 2013). However, the displays of affect are more naturalistic and difficult to detect.

Studies on detection of affect in programming are relatively limited. In a series of studies (Grafsgaard et al., 2011; Grafsgaard et al., 2013a; Grafsgaard et al., 2013b), displays of affect are detected using posture, gesture, and facial features, but are also induced by the nature of the experimental setup which involved a human tutor communicating with the student through an interface. An attempt to classify emotional states in a purely-coding setup was done by Bosch, Chen, and D'Mello by training classifier models using face features captured from a video camera (2014). They found that fine-grained and fixed judgment emotions was difficult, with the classifiers failing to perform above chance. They achieved better results for spontaneous judgments on confusion and frustration. More recently, negative affect was detected from students' keyboard and mouse actions in programming (Vea & Rodrigo, 2016).

Despite its difficulty, there is value in being able to infer emotions of students while doing programming tasks. In our previous studies, we have shown that a system that can provide adaptive feedback when confusion is detected in the context of programming can potentially help students solve more problems in programming practice (Tiam-Lee & Sumi, 2017, 2018a, 2018b). However, the model in this study was trained on a small dataset and was not rigorously evaluated for its accuracy. Expanding on this idea, we performed analysis on a larger dataset and found that both face features and log features can be meaningful in detecting emotions in programming (Tiam-Lee & Sumi, 2019). In this study, we continue to explore this idea by showing that combining log features with face features can improve the detection of emotions on fixed point judgment intervals. Furthermore, we show that identifying representative intervals of each emotion can potentially improve the classification of emotion over longer intervals, an insight that can be useful in building better emotion detectors in programming.

2. Methodology

2.1 Data Collection

We collected data from students by asking them to participate in a simulated programming session in which they must solve a series of coding exercises. In each exercise, they must write the body of a function according to a given specification. For example, in one of the exercises the function should return the area of a square given the length of its side as an argument. A custom application was used for this task which automatically logged all code changes and compilations. It also recorded a video of the student's face. Figure 1 shows the setup of the data collection.



Figure 1. Setup of the programming session.

Each student participated for 45 minutes or until all exercises were solved. After this, the student was shown replays of different time intervals of the session. These were selected based on key moments such as compilations, and when the student starts or stops typing. For each time interval, the student was shown the video of his face and the snapshot of his or her code at that time. Based on this, the student was asked to provide an emotion label which describes best what he or she felt during that time interval. The options were “engaged”, “confused”, “frustrated”, “bored”, and “neutral” (no apparent emotion). This is based on a previous study which identified the common emotions that are prevalent in programming tasks (Bosch, D’Mello, & Mills., 2013). To minimize subjectivity, we provided the definitions for each emotion, shown in Table 1. Our methodology produced self-report affect judgments that are fixed (intervals were chosen by the system, and not by the student).

A total of 73 students from Japan and the Philippines participated in the study. All students were enrolled in a freshman introductory programming course at the time of the study. All participants signed informed consent forms which contained an explanation about the study, the data to be

collected, and how it will be used. After removing bad videos (e.g., face not captured properly, face not detected well), we were left with session data from 39 students.

Table 1

Definition of Affective State Labels

Emotion	Definition
Engaged	You are immersed in the activity and enjoying it.
Confused	You have feelings of uncertainty on how to proceed.
Frustrated	You have strong feelings of anger or disappointment.
Bored	You feel a lack of interest in continuing with the activity.
Neutral	There's no apparent feeling.

2.2 Data Processing

We extract log features by taking the number of code insertions, code deletions, compilations, and submissions. We also consider if the compilation was successful or not (e.g., syntax or runtime error), and if the submission was passed all test cases or failed some of them.

To extract face features from the videos, we used OpenFace, a toolkit which can estimate facial landmarks and detect the presence of action units from images and videos using state-of-the-art machine learning techniques (Baltrusaitis, Zadeh, & Lim, 2018). Figure 2 shows an example of the estimation of facial landmarks. Action units (AUs) are a taxonomy of fundamental actions of facial muscles, some examples of which are opening the mouth or raising the eyebrows (Ekman and Friesen, 2003). We processed each video per frame and extracted the head pose features, which included the x, y, and z positions of the head in 3-D space and the x, y, and z rotation of the head. We also extracted the intensity of 17 AUs that OpenFace can detect. The intensity is a numerical value representing how strongly the AU was displayed in the frame.

We merged the consecutive intervals with the same reported emotion, resulting in intervals of widely varying lengths. An instance representing a time interval of the session is constructed by aggregating all observations within the time interval. First, the head pose and AU features were z-standardized on a student level to account for individual differences. We then take the sum of each log feature and the median, maximum, and standard deviation of all face features across the interval. We also perform some data cleaning by disregarding frames where OpenFace failed to detect the face properly using rule-based conditions.

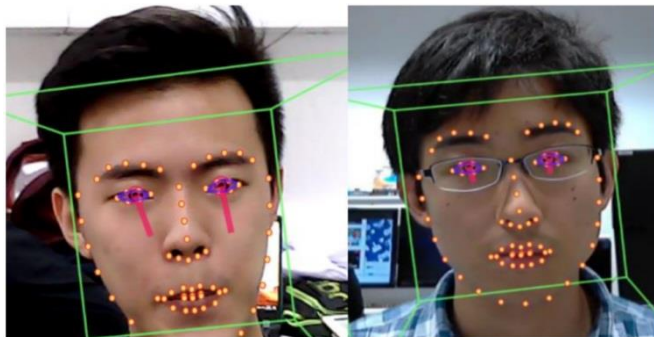


Figure 2. Estimation of facial landmarks by OpenFace.

2.3 Model Training

We trained machine learning models to classify affect on fixed-length, fine-grained sub-intervals. To do this, we used the caret package in R. We did not consider those labelled “neutral”, as we were interested in detecting the type of emotion, if it is present. We obtained the sub-intervals by splitting each interval into fixed-length sub-intervals, with a sliding window of $L/2$, where L is the length of the sub-interval. We trained models using 10, 20 and 60 second fixed-length sub-intervals to compare the performance of the classifiers across different L values. The L values were selected manually to test the ideal length of the interval for detection of affect.

To prevent overfitting and to ensure that the models can generalize to other unknown students, we used a student-level cross-fold validation. We randomly divided the students into 3 sets (each set contained 13 students), and used 2 sets for training and 1 set for evaluation in each fold. We also repeated this process 10 times to account for random sampling error. We built a separate classifier for each academic affective state, and tried different classifier models, namely: C4.5 Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, and Multilayer Perceptron. We selected the best fit performing model for each. Our baseline is using face features only, as was presented in previous studies on emotion detection in programming, so we first, we trained models using face features only, and then compared the performance against models that were trained with both face and log features to see if the addition of log features can improve performance.

2.4 Models for Classification of Longer Intervals

We also performed detection of affect over the original intervals prior to splitting. These intervals varied widely in terms of the length. To perform the classification, we trained hidden Markov models for each of the affective state labels based on the sequence of classifications produced by the localized classifiers, i.e. we combine the prediction output of each of the affective state classifiers for each sub-interval and used it to train a discrete state hidden Markov model for each affective state.

To classify the affective state of an unknown interval, first we divide it into several fixed-length sub-intervals of length L , with a sliding window of $L/2$. Then, we use the localized affect classifiers to classify each sub-interval for the presence of each of the affective states. We then use this as an input to each of the hidden Markov models for each affective state and classify it based on the model that yields the highest probability of generating that sequence of states. Figure 3 shows a diagram of this process. We found that this approach can yield good results when the localized affect classifiers are trained on representative intervals, which can be obtained from judgments from external human annotators. We discuss more of this on the next section.

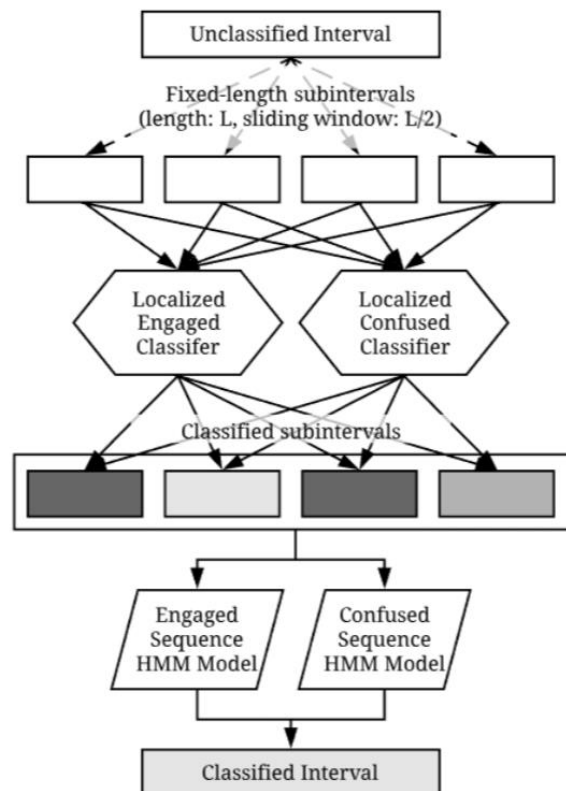


Figure 3. Classifying longer intervals from sequence of sub-interval classifications (engagement and confusion).

3. Results

3.1 Classification of Fine-Grained Intervals

Figure 4 shows the performance of the best classifiers for each affective state label on a student-level cross-fold validation. We treated each affective state as a binary classification problem (e.g., “engaged” or “not engaged”, where all intervals labelled “confused”, “frustrated” and “bored” were considered as the negative class). Because we are dealing with imbalanced datasets, we use Cohen's kappa κ as the metric for performance. $\kappa = 0$ means that the model is not any better than a model that predicts randomly by chance, while higher values are better.

By using face features alone, it is very difficult to predict the presence of affective states beyond random chance. This is not very surprising as it is consistent with previous findings in the literature. However, by combining log features and face features, the performance for classification on the 20 second and 60 second sub-intervals could be improved. Among the sub-interval lengths, $L = 60$ had the best performance. Among the affective states, engagement yielded the best performance, reaching $\kappa = 0.30$ on the 20 second interval and $\kappa = 0.45$ on the 60 second interval. Both of these used the multilayer perceptron classifier. However, detection for the other affective states remained poor despite the improvement.

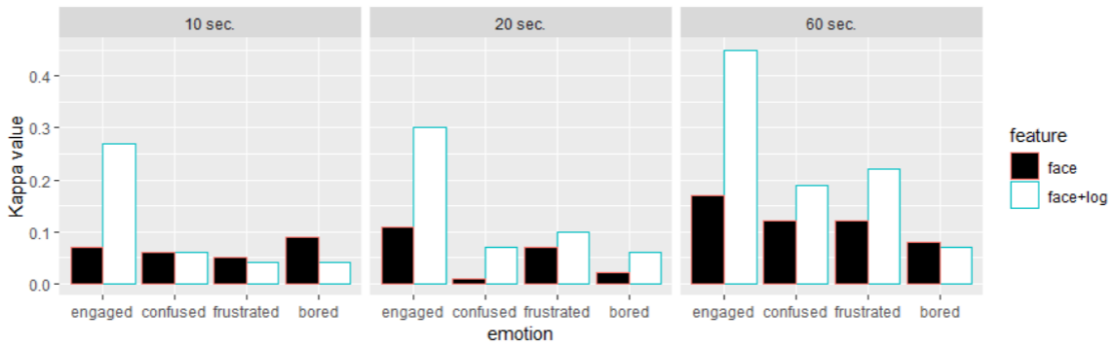


Figure 4. Performance of Best Classifiers on Fine-Grained Detection for Students' Self-Report.

We posit that one of the reasons for the difficulty of fine-grained affect classification is the sparseness of the displays of affect over an emotional episode. For example, if the student reports confusion for several consecutive intervals, he is reporting a persistent emotion that is not necessarily displayed to an observer in those intervals. Even though the student was asked to do the self-report for each individual interval locally, it is impossible to separate it from the emotions experienced in the surrounding intervals and from the context of the situation. Thus, we think that only a subset of the intervals are truly representative of the reported emotion in terms of the display of affect.

In order to determine these representative intervals, We tried to get a more localized judgment of affect by asking two human annotators to manually watch each sub-interval ($L = 10, 20, 60$), and provide an affect judgment of “engaged”, “confused”, “frustrated”, or “bored”. One of the annotators is a professor who has taught introductory programming courses, while the other one is a master course student who has served as a teaching assistant for the same course, so both of them have experiences on working with students who are learning how to code.

Because the annotators did not experience the programming session first-hand, they are free from any persistent emotion that might influence their judgments. We also shuffled the order the sub-intervals were presented to the annotators to further minimize the chance of associating the current judgment with the student or with the context of a particular episode, in an attempt to get localized affect judgments within each interval. We took all the sub-intervals where the two annotators and the student's self-report are all in agreement and considered them to be the representative sub-intervals of that affective state. The percentage of agreement between the two annotators is further discussed in Section 4.2.

We performed classification on the representative sub-intervals resulting into significantly better results. However, because the original data contained a low number of reports for frustration and boredom, there were very representative intervals identified to make a meaningful classification. The results for engagement and confusion for classifying representative intervals, both for face features only and both face and log features, are shown in Figure 5.

The performance of the models for representative intervals was better in both classifying engagement and confusion against all other emotions, reaching $\kappa = 0.65$ and $\kappa = 0.85$ for classifying engagement on 20 second and 60 second intervals, respectively and $\kappa = 0.49$ and $\kappa = 0.72$ for classifying confusion on 20 second and 60 second intervals, respectively. It should be noted that our intention for this is not to disregard all the other intervals that could not be classified easily, but to identify the important intervals which can be used later on to improve classification of longer sequences, which is discussed in the next section.

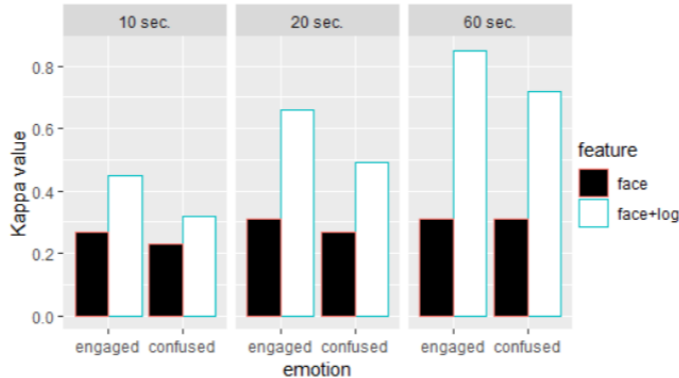


Figure 5. Performance of Best Classifiers on Fine-Grained Detection for Student Reports Combined with Annotator Judgment.

3.2 Classification of Longer Intervals

To classify longer intervals, we trained hidden Markov models from the sequence of predictions made by the localized classifiers discussed in the previous section. One hidden Markov model was trained for each affective state. Classification is performed by getting the model with the higher probability of generating the unknown sequence.

Because the localized interval classifiers trained on the actual student self-reports did not generalize very well on a student level apart from engagement, we used the localized classifiers trained on the representative intervals identified by the human annotators as described in the previous section. Unfortunately, since we could not build frustration and boredom models for representative intervals due to the lack of data, we could only perform classification for engagement and confusion. To ensure the models are generalizable, the models were still evaluated also on a student-level cross-fold validation, in which both the HMM models and the localized classifiers were trained only with instances that belonged to students that are in the training set.

Table 2 shows the results of the classification of the original intervals from the student reports using 10 second, 20 second, and 60 second sub-interval length for both face features only and face features combined with log features. We found that the best performing model was using both face and log features and splitting the interval into 20 second intervals with a sliding window of 10. This model was able to classify 76% of the intervals correctly, with $\kappa = 0.45$. It could also be observed that using face feature and log features improved the performance of the models when compared to using face features alone.

Table 2

HMM Classification Results

Face Features Only			Face and Log Features		
Interval	Acc.	κ	Interval	Acc.	κ
10 sec.	0.6	0.15	10 sec.	0.65	0.24
20 sec.	0.62	0.11	20 sec.	0.76	0.45
60 sec.	0.58	-0.14	60 sec.	0.58	0.22

4. Discussion

4.1 Contributions of this Study

In this study we performed two types of classification on affective states in programming: fine-grained classification of fixed-length intervals, and classification of longer intervals based on sequences of sub-intervals. Our study has two main contributions: (1) adding log features can improve the performance of classifier models for affect in activities like coding, and (2) using judgments of external observers to identify representative intervals can improve classification on longer intervals.

For the first contribution, on both types of classification, we have shown that adding log features in combination with face features can improve the prediction performance of classifier models. This suggests that the student activity is valuable in estimating emotion, especially in types of activities like programming where there is relatively more reserved and naturalistic displays of affect. For the second contribution, we have shown that the sequence of classifications from a model trained on representative sub-intervals can be used to classify longer sequences of affect. Although the models that we were able to train were far from excellent, it presents an improvement to previous attempts in the domain of affect detection in a programming setting.

4.2 Challenges in Fine-Grained Affect Detection

Fine-grained affect detection can generally be useful for intelligent programming tutors for systems to be constantly aware of the student's state, as opposed to being only aware in specific scenarios. In line with this, it is also interesting to determine the ideal length of the interval for discriminating different emotion types. In study, even though we were able to improve the performance of classifiers for fine-grained affect detection by adding log features, the resulting models for confusion, frustration, and boredom did not really achieve a significant performance above chance.

The difficulty of fine-grained affect detection may be caused by the fact that (1) emotions experienced by the students are not localized to short intervals, but rather influenced by the events that happen outside the interval, and (2) displays of affect do not necessarily occur constantly all throughout an emotional episode.

Table 3

Agreement Rate with Human Annotators (A/N: ratio of intervals where annotators made the same judgment over all intervals, A+S/N: ratio of all intervals where annotators and self-report are the same judgment over all intervals, A+S/S: percentage of all self-reports under the emotion that was correctly identified by both annotators)

Interval	Emotion	A/N	A+S/N	A+S/S
10 sec.	Engaged	52%	27.66%	65.1%
	Confused	17%	5.83%	25.55%
	Frustrated	0.67%	0.17%	0.57%
	Bored	0%	0%	0%
20 sec.	Engaged	52.67%	35.17%	66.57%
	Confused	17.17%	4.67%	25.69%
	Frustrated	1.17%	0.67%	2.96%
	Bored	0%	0%	0%
60 sec.	Engaged	53.27%	37.07%	72.12%
	Confused	11.53%	4.05%	21.31%
	Frustrated	1.25%	0.93%	3.95%
	Bored	0%	0%	0%

In this study, we asked external observers to identify which parts of an emotional episode are representative of that emotion by identifying the intervals where the observers' judgment and the student report agree with one another. Using these intervals, we were able to train fine-grained interval

classifiers that generalize significantly better on student-level cross-fold validation. This suggests that the representative sub-intervals are indeed much easier to discriminate from one another and can be considered as the sub-intervals that define each affective state. One of limitations that we encountered in this study, however, is that we did not find enough representative intervals for frustration and boredom to make a meaningful classification model. Table 3 shows the agreement of the two annotators and the student report for each of the affective states. Based on this data, around 65% to 70% of engaged sub-intervals could easily be recognized locally, but only 20% to 25% of the confused sub-intervals could easily be recognized locally.

The lack of agreement on frustration and boredom might have been caused by the fact that the student reports for these two affective states were relatively less than the other two, and thus there simply weren't enough examples for enough representative intervals to be identified. However, it is also a possibility that these two emotional states are not easy to detect for humans in short, localized intervals on face and log features alone, and must be supplemented with additional information.

4.3 Classification on Longer Sequences

In this study we also classified longer sequences based on prediction output on fixed-length sub-intervals. Figure 6 shows 10 randomly selected examples that were correctly classified by the HMM models from each class, as well as the corresponding classifier outputs for each sub-interval in the sequence. Each cell in the figure represents a 20 second sub-interval in the sequences, with a 10 second sliding window (i.e, half of the sub-interval overlaps with the previous one). There are four possible discrete symbols, since we considered two classes in this case: (1) both classifiers returned false, (2) only the “engaged” classifier returned true, (3) only the “confused” classifier returned true, and (4) both classifiers returns true.

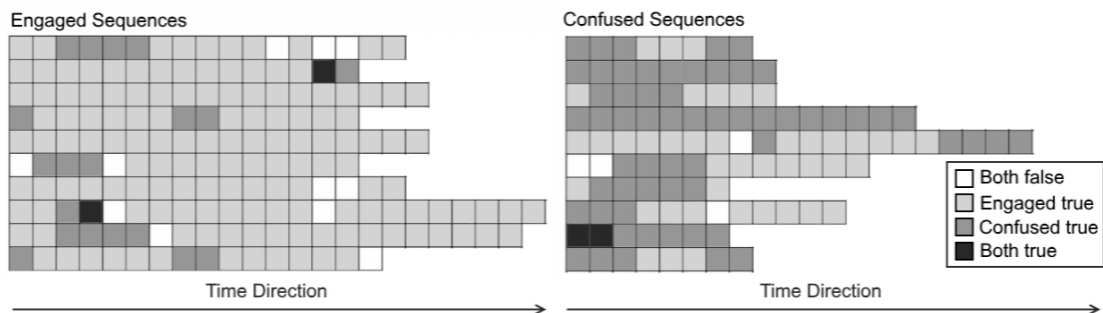


Figure 6. Examples of “engaged” and “confused” sequences correctly classified by the HMM models.

By visually inspecting the sequences, it could be seen that in the cases where the student reported engagement (i.e., the engaged sequences), most of the sub-intervals were classified as “engaged”. However, there were also some cases wherein some parts were classified as “confused”, but in these cases the confusion eventually turns into engagement. These sequences most likely represent confusions that were eventually resolved. On the other hand, for the “confused” sequences, there appeared to be more prevalent sub-intervals that were classified as “confused”.

It can be said that both “engaged” and “confused” sub-intervals could be observed from within both types of sequence, and this further supports the idea that affect judgments made locally within an interval may not always reflect the actual emotional experience experienced. Furthermore, it should also be noted that one limitation of our study was that we did not consider co-occurring affective states.

4.4 Useful Features on the Prediction of Affect

To investigate which features are useful in predicting each affect, we used the RELIEF-F feature ranking algorithm. RELIEF-F ranks the features based on how good they are at discriminating each instances from their nearest neighbors of a different class.

For engagement, document insertions was ranked highly across all sub-interval lengths, both by using all sub-intervals and by using representative sub-intervals only. Document insertions tend to

occur more in moments of engagement. For example, in the 20 second fixed-length sub-interval case, document insertions were significantly larger in “engaged” sub-intervals ($\mu = 16.4$) compared to the other sub-intervals ($\mu = 8.22$) with a $p < 2.2 \times 10^{-16}$.

Compilation errors was also ranked highly for classifying engagement and confusion. Upon further analysis, we found that compile errors occurred significantly less on “engaged” ($\mu = 0.01$) sub-intervals as compared to “not engaged” sub-intervals ($\mu = 0.03$, $p < 0.002$) and significantly more on “confused” ($\mu = 0.04$) sub-intervals compared to “not confused” intervals ($\mu = 0.02$, $p < 0.01$).

AU04 (brow lowerer) was ranked highly in the classification of confusion and frustration. AU04 was observed in the data when the eyebrow is furrowed, as well as when the subjects gaze downwards. This often happened during typing when the eyes quickly shift between looking at the screen and at the keyboard without turning the head down. Figure 7 shows some examples of AU04. Aside from this, the ranking of the other face features varied depending on the length of the sub-interval, even when only representative intervals are considered.



Figure 7. Displays of AU04 (brow lowerer) by a Filipino (left) and Japanese (right) student.

5. Conclusion

In this study, we have presented improvements to the detection of academic emotions in a programming setting by combining log information to the face features. We have also developed an approach for classifying longer sequences by training models on representative intervals which combine students' self-reports with the localized judgments of external annotators.

Acknowledgements

The authors would like to thank Mr. Fritz Kevin Flores, Mr. Manuel Carl Toleran, and Mr. Kayle Anjelo Tiu for assisting in the facilitation of the data collection process in the Philippines.

References

- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 59-66). IEEE.
- Bosch, N., & D’Mello, S. (2013, June). Sequential patterns of affective states of novice programmers. In The First Workshop on AI-supported Education for Computer Science (AIEDCS 2013) (pp. 1-10).
- Bosch, N., D’Mello, S., & Mills, C. (2013, July). What emotions do novices experience during their first computer programming learning session?. In International Conference on Artificial Intelligence in Education (pp. 11-20). Springer, Berlin, Heidelberg.
- Bosch, N., Chen, Y., & D’Mello, S. (2014, June). It’s written on your face: detecting affective states from facial expressions while learning computer programming. In International Conference on Intelligent Tutoring Systems (pp. 39-44). Springer, Cham.
- Cabada, R. Z., Estrada, M. L. B., Hernández, F. G., Bustillos, R. O., & Reyes-García, C. A. (2018). An affective and Web 3.0-based learning environment for a programming language. *Telematics and Informatics*, 35(3), 611-628.
- Cho, M. H., & Heron, M. L. (2015). Self-regulated learning: the role of motivation, emotion, and use of learning strategies in students’ learning experiences in a self-paced online mathematics course. *Distance Education*, 36(1), 80-99.

- Crow, T., Luxton-Reilly, A., & Wuensche, B. (2018, January). Intelligent tutoring systems for programming education: a systematic review. In *Proceedings of the 20th Australasian Computing Education Conference* (pp. 53-62). ACM.
- Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology*, 101(4), 948.
- D'Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4), 23.
- D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*, 47(3), 43.
- Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.
- Grafsgaard, J. F., Boyer, K. E., & Lester, J. C. (2011, October). Predicting facial indicators of confusion with hidden Markov models. In *International Conference on Affective computing and intelligent interaction* (pp. 97-106). Springer, Berlin, Heidelberg.
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2013, September). Automatically recognizing facial indicators of frustration: a learning-centric analysis. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 159-165). IEEE.
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2013, July). Embodied affect in tutorial dialogue: student gesture and posture. In *International Conference on Artificial Intelligence in Education* (pp. 1-10). Springer, Berlin, Heidelberg.
- Lehman, B., Matthews, M., D'Mello, S., & Person, N. (2008, June). What are you feeling? Investigating student affective states during expert human tutoring sessions. In *International conference on intelligent tutoring systems* (pp. 50-59). Springer, Berlin, Heidelberg.
- Malekzadeh, M., Mustafa, M. B., & Lahsasna, A. (2015). A review of emotion regulation in intelligent tutoring systems. *Journal of Educational Technology & Society*, 18(4), 435-445.
- Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of educational psychology*, 106(1), 121.
- Paleari, M., Lisetti, C., & Lethonen, M. (2005). *Virtual Agent for Learning Environment Reacting and Interacting Emotionally*. Eurecom. Fr, 3-5.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2), 91-105.
- Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lahoz, M. B. V., ... & Tabanao, E. S. (2009, July). Affective and behavioral predictors of novice programmer achievement. In *ACM SIGCSE Bulletin* (Vol. 41, No. 3, pp. 156-160). ACM.
- Tiam-Lee, T. J., & Sumi, K. (2017, September). A comparison of Filipino and Japanese facial expressions and hand gestures in relation to affective states in programming sessions. In *Workshop for Computation: Theory and Practice (WCTP)* (pp. 143-157). World Scientific, Singapore.
- Tiam-Lee, T. J., & Sumi, K. (2018a, June). Adaptive feedback based on student emotion in a system for programming practice. In *International Conference on Intelligent Tutoring Systems* (pp. 243-255). Springer, Cham.
- Tiam-Lee, T. J., & Sumi, K. (2018, October). Procedural Generation of Programming Exercises with Guides Based on the Student's Emotion. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1465-1470). IEEE.
- Tiam-Lee, T. J., & Sumi, K. (2019, June). Analysis and Prediction of Student Emotions While Doing Programming Exercises. In *International Conference on Intelligent Tutoring Systems* (pp. 24-33). Springer, Cham.
- Vea, L., & Rodrigo, M. M. (2016, August). Modeling negative affect detector of novice programming students using keyboard dynamics and mouse behavior. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 127-138). Springer, Cham.
- Zatarain-Cabada, R., Barrón-Estrada, M. L., Camacho, J. L. O., & Reyes-García, C. A. (2014, August). Affective Tutoring System for Android Mobiles. In *International Conference on Intelligent Computing* (pp. 1-10). Springer, Cham.

Using Data Mining Techniques to Assess Students' Answer Predictions

Alisa LINCKE^{a*}, Marc JANSEN^{a,b}, Marcelo MILRAD^a, Elias BERGE^c

^aComputer Science and Media Technology Department, Linnaeus University, Sweden

^bAffiliation B, University of Applied Sciences Ruhr West, Germany

^cHypocampus AB, Sweden

*alisa.lincke@lnu.se

Abstract: Estimating students' knowledge and performance, modeling their learning behaviors, and discovering and analyzing their different characteristics are some of the main tasks in the field of research called educational data mining (EDM). According to Chounta (2017), the predicted probabilities that a student will answer a question correctly can provide some insights into the student's knowledge. Based on this point of departure, the main objective of this paper is to apply different data mining techniques to predict the probabilities that students will answer questions correctly by using their interaction records with a web-based learning platform called Hypocampus. Five different machine learning algorithms and a rich context model were used on the Hypocampus dataset. The results of our evaluation indicate that the gradient-boosted tree and the XGboost algorithms are best in predicting the correctness of the student's answer.

Keywords: adaptive learning, machine learning, rich context model, answers probability prediction

1. Introduction

One of the biggest challenges for educators is to meet the individual needs of students while facing the constraints of time. One way to personalize education is by using adaptable learning systems (Papoušek, 2015). In order to efficiently provide students with personalized and adaptive digital content and a meaningful learning experience, it is crucial that the learning system gets over time an *understanding* not only of the students' current knowledge level but also his/her progression. One traditional way of assessing the knowledge level is letting students take a placement test (Hodara, 2015). However, to make a placement test adaptive, the system needs to be able to draw conclusions from every answered question. According to Chounta (2017), predicted probabilities that students will answer questions correctly can provide some insights into students' knowledge. By using the answers to predict the probability of answering correct on other questions a learning system might be able to recommend questions with a suitable level of difficulty. This would make the placement test more efficient, i.e., needing fewer questions to get an accurate picture of the students' knowledge level. Predicting the probabilities of students' answering correct may also be valuable in order to maximize students' engagement. If we know the probabilities of students answering questions correctly then we can optimize the studies with regard to engagement and knowledge level. Using probabilities, we can objectively measure a student's knowledge on a particular given subject. This measurement can be used as a valuable feedback to the students. Previous studies have suggested that an adaptive fail rate in a quiz increases student engagement (Papoušek, 2015). By choosing questions with a difficulty level that increases the chances of a student answering correct around 60% of the questions seems to hit a sweet spot where the average student experiences the quiz challenging without being too difficult. Therefore, this paper aims to estimate the probability that students will answer questions correctly by using different data mining techniques on data provided by the Hypocampus¹ system.

¹<https://www.hypocampus.se>

Hypocampus is an adaptive web-based learning platform used by medical students. It contains a library with many interactive reading materials (e.g., course literature) that students can use for self-studies in order to learn about a particular subject matter and to revise and review their current knowledge. The platform provides also quizzes for each reading material in order to help students to check and assess their current knowledge for each particular subject matter. In addition, it offers customized learning paths based on quantitative educational studies, visualizations of learning progress for students and teachers, and adaptive individual learning pathways. The learning platform optimizes the learning content according to the principles of retrieval practice (Karpicke, 2008).

Our research contributes to the student knowledge estimation research area with a particular focus on: (a) *providing a set of features that allows getting good prediction accuracy on students' answers (Table 2)*; (b) *an approach that works on the subject based level*; (c) *an approach that uses different data types besides the binary representation of students' knowledge state*. The remainder of this paper is structured as follows. Section 2 provides a short overview of the challenges and existing research related to modeling and predicting students' knowledge. Section 3 describes the proposed approach, our dataset, feature extraction, and the models used in this study. In Section 4, we present and discuss the evaluation results of our efforts while in Section 5 we provide our conclusions and outline and present possible lines of future work.

2. Related Work

Modeling and predicting the knowledge of students in online learning systems is a well-identified problem (Piech, 2015; Duong, 2013; Pelánek, 2017; Pardos, 2011; Tato et al., 2017). Unknown students' knowledge background, access to the learning resources such as reading material, quizzes, exams, courses at any time and order brings different learning behaviors (e.g., accessing the learning material in different sequences, some students just doing quizzes, exams without reading the material on the web platform, others first read the material and afterwards doing quizzes to check their knowledge about this material). Furthermore, students can use other reading resources besides those provided by the learning platform (such as books and notes from the classes). All of these are examples of challenges to model students' knowledge.

One approach for measuring academic achievements and student's knowledge is the Item Response Theory (IRT) models (Reise, 2014; Chen, 2005). IRT models allow measuring different students' abilities (intelligence, individual learning ability, attitude, academic achievements) by using answers on questions as test-based assessment. It predicts the probability that a student will answer the question correctly as a function with two parameters: student's knowledge level and the question difficulty (Chaundhry, 2018; Galvez, 2009). This modeling approach showed good practical use in estimating students' performance and making adaptive quizzes (dynamically decide which question to show based on student's answers). However, this approach does not model the evolution of students' knowledge over time (Chaundhry, 2018; Khajah, 2014).

Students generate a vast amount of interactional data in online learning platforms that allows to use data mining and machine learning techniques to better estimate students' knowledge and performance. A Recurrent Neural Network (RNN) is a type of artificial neural network for processing sequential data such as text, videos, sensors data, and stock markets (Hecht-Nielsen, 1992). RNN takes into account the current information and the previous/historical information to model student's knowledge. Student's study behavior data collected in web-based learning systems can be represented as sequences of study behaviors. For example, a sequence of performed exercises (Piech, 2015), a sequence of questions answered in a quiz, a sequence of reading and quiz sessions, a sequence of attempts and hints to solve a programming task (Wang, 2017; Duong, 2013). These study behavior sequences can be used as an input to RNN in order to predict whether the student will complete successfully a new exercise, or to predict the next type of learning activity (reading/quiz). Usually, the data consists of binary variables indicating whether the student will complete the task successfully or not and the label indicating the skill name or knowledge component (Wilson, 2016, Piech, 2015). The possible concern with RNN is the vanishing and exploding gradients (Xiong, 2016) that influence the accuracy and performance of the algorithm. Long short-term memory (LSTM) is introduced as a solution to the vanishing problem (Hochreiter, 1997). LSTM is an extension for RNN that allows to remember the inputs over a long period of time and decide whether to store, not store or delete the

information based on the importance of that information (Hochreiter, 1997). In a study predicting students' learning gains, Lin (2017) compared BKT, RNN and LSTM models. The LSTM model showed the highest accuracy in predicting students' learning gains.

Different classification models such as decision trees, random forest classifier, logistic regression classifier and vector support machine were used to predict whether a student would pass or fail a final exam (Bucos, 2018; Bunkar et al., 2012). In their study, the support vector machine and logistic regression obtained the best accuracy in predicting failing or passing the exam (Bucos, 2018). The features dataset that they have used contains information about students' gender, average grade, past examination grades, class attendance and others that are not always available in distance learning web based educational systems. Mueen (2016) has analyzed collected data from learning management systems (LMS) (such as forum data, assignments grades, learning material access duration, quizzes) in order to predict the students' performance by using different machine learning algorithms. In his study, the best accuracy score (86%) was achieved by a Naïve Bayes classification algorithm (Mueen, 2016). Tato et al. (2017) used probabilistic approach, a Bayesian neural network of logical reasoning skills to predict the learner's knowledge state with around 85% accuracy.

In summary, extensive research has been carried out on modeling students' knowledge and performance. However, to the best of our knowledge, it is still not clear how existing models handle different features collected by on-line learning platforms (for example, time answering the questions, time since last reading the material, time since last time doing quiz, repetitions) and how they model subject based knowledge and not skill based knowledge, without experts annotations or input, without information about grades, class attendances, and personal information (e.g., gender, age, study year).

3. Our proposed approach

The approach suggested in this paper is to predict the probability whether a question will be answered correctly based on a general machine-learning (ML) pipeline (Pentreath, 2015), see Figure 1 below. There are five main steps: data collection (Dataset), data preprocessing, feature extraction, selection and transformation, model application, and the last step is the model evaluation. This process is iterative and can be repeated many times until the model that performs best will be defined. In terms of technologies and tools, we have used Apache Spark MLlib and Scikit-learn libraries for performing data preprocessing, feature extraction, and transformation steps. In the model application step we use different machine learning models from the Apache Spark MLlib (Meng, 2016), the Keras deep learning library, and the Contextualization Service (Sotsenko, 2016).

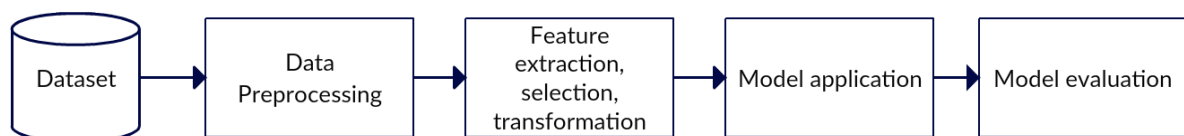


Figure 1. ML pipeline for predicting the probability if a question will be answered correctly.

3.1 Dataset

For this study, we have obtained the data from the Hypocampus web-based learning platform. The platform offers reading material for various subjects (e.g., Dermatology, Surgery, Gynecology, Internal Medicine and others). Every subject has a number of topics called chaptergroups. Every chaptergroup consists of chapters and quizzes. Most of the quizzes contain from 4 to 15 questions, some of them can have up to 28 questions. There are two types of questions: multiple-choice and text questions. The multiple-choice questions have several options to choose and only one is correct. After answering the multiple-choice questions, the system will show the direct feedback to students whether the answer is correct or incorrect. The text question contains a problem description and a student should provide a text answer on the problem. After answering a text question, the system does not check the text answer provided by student, but rather shows the answer and explanations to the problem. Once the student has seen the answer she/he should correct herself by selecting "I knew the answer" (correct) or "I need to

read more” (incorrect). Table 1 describes a summary of the collected data for 300 medical students used as part of our study that took place over a period of 10 months between 2017-2018.

Table 1

Summary of collected data.

	Number of records
Students	300
Quiz records	121 423
Multiple-choice questions (msq)	18 092
Text questions	103 331
Correct answers (msq)	14 580
Incorrect answers (msq)	3 420

The dataset (Table 1) contains information about quiz records from different subjects gathered over a period between one to three months. Quiz records include information about user identification number, question type (multiple-choice or text), question identifier, time answering a question, time reviewing the feedback from the system after answering the question, student’s answer (true – correct and false – incorrect), student’s text answer on the text questions, timestamp, course identifier, and question session. After collecting all these data, the preprocessing and feature extraction steps are performed in order to prepare the dataset to be used by data mining techniques.

3.2 Data Preprocessing

As part of the data preprocessing step we apply three filters: (a) selecting records that were collected in the system production mode; (b) selecting only multiple-choice question types, because this type of questions are more reliable for our model evaluation than text type questions (text type question correctness given to students and not to the system); (c) removing records that have missing values related to the question id information.

3.3 Feature Extraction and Transformation

We performed preprocessing on the collected data and extracted and selected specific features. After performing several iterations in the ML pipeline (Figure 1) the following 18 features were identified and described as presented in Table 2. There are five categorical features (F1, F2, F3, F4, F5), three time-related features (in seconds) (F6, F7, F8), and ten numerical features (F9, F10, F11, F12, F13, F14, F15, F16, F17, F18). Some of the features are directly used from the original dataset (e.g., user id, chapter id, question id, question session) and other features are calculated (e.g., number of correct/incorrect answers, attempt number) for each record in the dataset.

Table 2

Feature overview

	Feature Type	Name	Description
F1	Categorical	User ID	User identification
F2	Categorical	Chapter ID	Chapter identification
F3	Categorical	Question ID	Question identification
F4	Categorical	Question Session ID	Question session identification
F5	Categorical	Time of the day	Morning, lunch, afternoon, evening, night
F6	Time related features	Time since last doing quiz	Represents the time duration since last time the student was doing a quiz
F7	Time related features	Time since last reading a chapter	Represents the time duration since last time the student was reading a chapter
F8	Time related features	Reading time	Total reading time per chapter (in

			seconds)
F9	Numerical	Correct	Number of questions answered correctly per chapter, per attempt, per question session
F10			
F11			
F11	Numerical	Incorrect	Number of questions answered incorrectly per chapter, per attempt, per question session
F12			
F13			
F14	Numerical	Attempt number	Number of times a question was answered by the student
F15	Numerical	Reading sessions	Number of times a student read the learning material (chapter)
F16	Numerical	Question facility index	Represents the question difficulty and is facility index of the students correct answers in first attempt in range between 0 and 1 (where 0 is very difficult and 1 is very easy).
F17	Numerical	Question number1	Question number in the question session
F18	Numerical	Question number2	Accumulated question number: total number of questions answered correctly for a specific chapter.

The features (in Table 1) have been transformed to an appropriate format (Vectors) for machine learning models by using the Vector Assembler component from the machine learning library (Mllib) in Spark.

3.4 Models

To predict the probability if a question will be answered correctly, we use six models: linear regression, logistic regression, gradient-boosted tree regression (Apache Spark Mllib), XGBoost (Chen, 2016) (Python library), feed-forward neural network (Keras deep learning library) and a rich context model (RCM) from Contextualization Service (Sotseko, 2016). We select two simple models (linear and logistic regression), and four more advanced models (two decision trees models, deep neural network and RCM) because they are most commonly used in regression problems (predicting probabilities) and taking contextual information into account (RCM). It is important to understand the student's current context (e.g., time of the day: morning, lunch, afternoon, evening, night; location, number of difficult questions answered correctly) in order to provide personalized learning tasks/quizzes. Furthermore, some of these models were used in predicting students' performance (Bucos, 2018; Shahiri, 2015; Zaidah, 2007).

Linear and logistic regressions are one of the simplest machine learning models used to predict one dependent variable based on the set of independent variables (Seber, 2012). Linear regression assumes that there is a linear relationship between dependent and independent variables. In our scenario the dependent variable is the answer to a question (correct/incorrect) and independent variables are features as described in Table 2. We use this model to check whether there is a linear relationship between the learning activity (quiz) and students' answers on the questions. Logistic regression is applied when the dependent variable is binary. In our prediction problem, linear and logistic regressions predict the probability from 0 to 1 if the question will be answered correctly. The following features are selected from Table 2: numerical features (F9-F18, except F15), one time related feature (F6), and one categorical feature (F5 transformed from categorical representation to numerical by using hour of the day from 0-24) based on decision that these models work with numerical data types. Features F1-F4 were used as labels in all machine learning models.

More advanced models such as decision trees (gradient boosted tree and improved version extreme gradient boosted tree (XGBoost)) help to reduce factors such as bias, variance, and dealing with unbalanced data (Cieslak, 2008; Chawla, 2004). Gradient boosted tree and XGBoost algorithm

have been shown great success in winning machine learning competitions such as Kaggle². To the best of our knowledge, in the literature review this model was not applied to student’s knowledge prediction. Therefore, we decide to test this model and to apply it in our study. The same set of features was selected and used as in linear regression model.

Deep neural networks become more popular in EDM tasks (Coelho, 2017; Guo, 2015). We use a multi-layer sequential feed-forward neural network. Different model parameters were tested, the following parameters were found empirically by minimizing the error: input layer with 11 neurons activation function ‘*relu*’, one hidden layer with 11 neurons and activation function ‘*relu*’, output layer with 1 neuron and activation function “*sigmoid*”, optimizer ‘*adam*’ and loss ‘*binary_crossentropy*’. We use the same set of features in linear regression model.

We use also rich context model models contextual information in a multidimensional vector space model (MVSM) to provide recommendations based on the current context of a user. This model can handle different data types therefore we include categorical features in predicting the answer correctness. The 16 features (except the reading time features F7, F8 and F15 that are left for future work) used to model student’s quiz records in RCM. RCM requires an *examples* set of vectors that represent the basis (or training set used in machine learning approach) and a *current context* set of vectors to obtain recommended result (or testing set used in machine learning approach). In this study, the quiz records are divided into *examples* and *current context* datasets with 7:3 ratios. The *examples* dataset is transformed to one-dimensional vector and placed in MVSM (e.g., Quiz Record X shown on Figure 2).

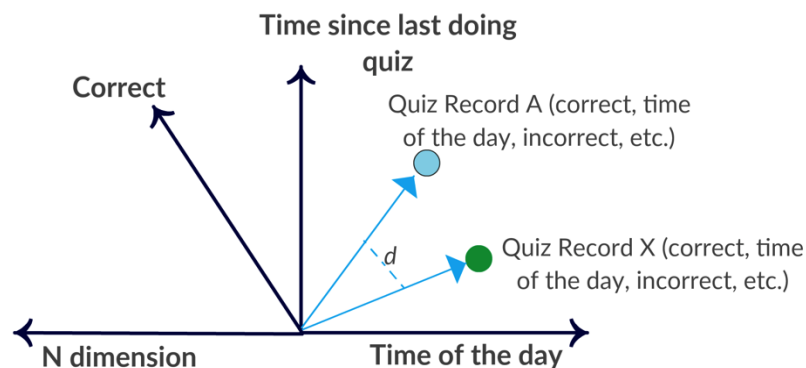


Figure 2. Representation of quiz records in RCM

The *current context* datasets are transformed into a one-dimensional vector (Quiz Record A in Figure 2) and Euclidean similarity metric (d) is used to calculate the distance to find the most similar quiz record in *examples* dataset (Quiz Record X). The most similar quiz record that has minimal distance defines if the student will answer correct or incorrect.

4. Evaluation

The evaluation in this study is conducted using the 10-fold cross validation approach (Kohavi, 1995) and the train-validation split approach provided by the Spark Mllib library for hyper-parameter tuning (Gounaris, 2018). The dataset used consists of 18092 quiz records from medical students that have studied using the Hypocampus web-based learning platform (Table 1). As mentioned earlier in this paper, the purpose of this study is to evaluate which model predicts best the probability that a student will answer the question correct. The evaluation results are shown in Table 3 with the following metrics: false positive rate (FP,%), false negative rate (FN,%), precision (Precision,%), recall (Recall,%), accuracy (Accuracy,%), F1-Score (F1,%) and Pearson correlation coefficient (R) between the predicted value by the algorithm (described in Section 3.4) and depended variable (answer on the question). In our answer prediction task: false positive are incorrect answered questions which have

² <https://www.kaggle.com>

been predicted as correct; false negative are correctly answered questions which were predicted as incorrectly answered questions; precision is a proportion of correct answers predicted; recall is a proportion of correctly answered questions which are predicted to be correctly answered; accuracy is a proposition of total number of answer predictions that were correct; F1-score is a weighted harmonic average of precision and recall; Pearson correlation coefficient shows how well the true value correlated with the predicted value, where 0 is not correlated and 1 is highly correlated.

Table 3

Evaluation results

Algorithm	FP,%	FN,%	Precision,%	Recall,%	Accuracy,%	F1,%	R
Linear Regression	62	4	82	95	81	88	0,43
Logistic Regression	72	4	80	96	79	88	0,35
Gradient-boosted tree	30	3	91	97	90	94	0,72
XGBoost	25	4	92	96	91	94	0,73
Neural Network	45	3	87	97	86	92	0,61
RCM	63	17	80	83	71	81	0,19

As shown in Table 3, the five machine learning algorithms performed well in predicting the probability that an answer will be correct, with an accuracy rate ranging between 81% and 91%. These results indicates that dataset of features very well correlates with independent variable (student's answer). The RCM model got the lowest accuracy (71%) and more false negative errors (17%) than machine learning approaches. This could be explained by the lack of using contextual information in the model such as student's location, age, gender, and others. The best algorithms in predicting the probability that a student will answer correctly are Gradient-boosted tree and XGBoost with around 90% accuracy and highest correlation (0,72-0,73) and lowest false positive and false negative errors (FP=25%, FN=3%). However, the false positive error is the biggest error in all algorithms. We manually analyzed the errors (FP and FN) and decided to balance the data in order to reduce FP errors. As mentioned earlier the original dataset contains 81% correct answers and 19% incorrect answers that make algorithms to predict more often that student will answer correctly. Therefore, we applied one of the well-known techniques for data balancing – synthetic minority oversampling technique (SMOTE) (Chawla, 2002). This technique adds incorrect answer records by finding the k-nearest-neighbors for minority class (incorrect answers) and randomly choosing one and using it to create a similar record. We applied SMOTE only to training dataset and validation dataset we did not change. Same evaluation metrics calculated for balancing the dataset and presented in Table 4 below.

Table 4

Evaluation results (balanced training data)

Algorithm	FP	FN	Precision	Recall	Accuracy	F1	R
Linear Regression	19	17	93	82	81	87	0,57
Logistic Regression	15	14	94	86	85	90	0,65
Gradient-boosted tree	14	10	95	90	88	92	0,71
XGBoost	13	10	95	90	89	93	0,73
Neural Network	24	11	92	89	85	90	0,61
RCM	5	23	94	77	85	84	0,72

Balancing the training data improves the accuracy 14% for the RCM and 6% for the logistic regression. Pearson correlation coefficient increased for all models, more than 20% of FP errors were reduced. Almost no change in accuracy for gradient-boosted tree, XGBoost, and neural network were obtained. This can be explained due to the fact that decision trees work well on unbalanced data and neural networks do not require balancing the data. Furthermore, the FP error is reduced and balanced (FN is increased) in gradient boosted tree and XGBoost. XGBoost algorithm still provides the best results even after balancing the training dataset. RCM received the smallest FP error (5%), biggest FN error (23%) and overall improvements in accuracy, F1-score, and Pearson correlation. Based on the obtained results

we can assume that RCM is not invariant to unbalanced data and requires to have balanced *example* data that is placed in MVSM as basis. Overall, this study shows that data generated in web-based learning platforms can be used to predict student's answer in order to estimate his/her knowledge on a subject.

5. Conclusion and Future Work

We applied different machine learning techniques to the problem of estimating students' knowledge by knowing the probability if the student will answer the question correctly. Six algorithms were applied including linear and logistic regression, gradient-boosted tree, XGBoost, deep neural network, and rich context model on a dataset consisting of medical students' answers on quizzes carried out at the Hypocampus web-based learning platform. The results show that the Gradient-boosted tree and the XGBoost algorithms outperform others by obtaining the overall prediction accuracy 90-91% and lowest false negative and false positive errors (4% and 25%). Additionally, the XGBoost algorithm performs well on unbalanced dataset with two classes that is shown in our case. The obtained results increase the prediction accuracy about 5% in comparison with other studies discussed in the related work section. Our results indicate that it is possible to predict the probability that a student will answer the question correct before doing a quiz by analyzing student's log data.

As part of our future research efforts, we plan to (a) add more features: reading time and time since last time read the chapter (F7, F8, F15), (b) add text type questions, and (c) apply one of the variants of deep knowledge tracing models (e.g. Bayesian neural network) and compare the obtained results with XGBoost algorithm; (d) create a measurement of students' knowledge based on the prediction probabilities results obtained from the XGBoost algorithm.

References

- Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In *International conference on intelligent tutoring systems*. 406-415.
- Bucos, M., & Drăgulescu, B. (2018). Predicting student success using data generated in traditional educational environments. *TEM Journal*, 7(3), 617.
- Bunkar, K., Singh, U. K., Pandya, B., & Bunkar, R. (2012). Data mining: Prediction for performance improvement of graduate students using classification. In *2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*. 1-5.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Chaudhry, R., Singh, H., Dogga, P., & Saini, S. K. (2018). Modeling Hint-Taking Behavior and Knowledge State of Students with Multi-Task Learning. *International Educational Data Mining Society*.
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1), 1-6.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785-794.
- Chen, C. M., Lee, H. M., & Chen, Y. H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, 44(3), 237-255.
- Chounta, I. A., Albacete, P., Jordan, P., Katz, S., & McLaren, B. M. (2017). The "Grey Area": A computational approach to model the Zone of Proximal Development. In *European Conference on Technology Enhanced Learning*. 3-16.
- Cieslak, D. A., & Chawla, N. V. (2008). Learning decision trees for unbalanced data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 241-256.
- Coelho, O. B., & Silveira, I. (2017, October). Deep Learning applied to Learning Analytics and Educational Data Mining: A Systematic Literature Review. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 28, No. 1, p. 143).
- Duong, H., Zhu, L., Wang, Y., & Heffernan, N. T. (2013). A prediction model that uses the sequence of attempts and hints to better predict knowledge: "Better to attempt the problem first, rather than ask for a hint". In *EDM*, 316-317.

- Galvez, J., Guzman, E., Conejo, R., & Millan, E. (2009). Student Knowledge Diagnosis Using Item Response Theory and Constraint-Based Modeling. In *AIED*. 291-298.
- Gounaris, A., & Torres, J. (2018). A methodology for spark parameter tuning. *Big data research*, 11, 22-32.
- Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting student's performance in educational data mining. In *2015 International Symposium on Educational Technology (ISET)*. 125-128.
- Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception*. 65-93.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Hodara, M., Jaggars, S., & Karp, M. J. M. (2012). Improving developmental education assessment and placement: Lessons from community colleges across the country. (Working Paper No. 51). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319 (5865), 966-968.
- Khajah, M. M., Huang, Y., González-Brenes, J. P., Mozer, M. C., & Brusilovsky, P. (2014). Integrating knowledge tracing and item response theory: A tale of two frameworks. In *CEUR Workshop Proceedings*.7-15.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*. Vol. 14 (2). 1137-1145.
- Lin, C., & Chi, M. (2017) Comparisons of BKT, RNN and LSTM for Predicting Student Learning Gains. In *AIED*.
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... & Xin, D. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research*, 17(1), 1235-1241.
- Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36.
- Pardos, Z. A., & Heffernan, N. T. (2011). KT-IDEM: introducing item difficulty to the knowledge tracing model. In *International conference on user modeling, adaptation, and personalization*. 243-254.
- Papoušek, J., Pelánek, R. (2015). Impact of adaptive educational system behaviour on student motivation. In *International Conference on Artificial Intelligence in Education*. 348-357.
- Pentreath, N. (2015). *Machine learning with spark*. Packt Publishing Ltd.
- Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3-5), 313-350.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in neural information processing systems*, 505-513.
- Reise, S. P., & Revicki, D. A. (Eds.). (2014). *Handbook of item response theory modeling: Applications to typical performance assessment*.
- Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis*. 329.
- Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422.
- Sotsenko, A., Jansen, M., Milrad, M., & Rana, J. (2016). Using a rich context model for real-time big data analytics in twitter. In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. 228-233.
- Wang, L., Sy, A., Liu, L., & Piech, C. (2017). Learning to Represent Student Knowledge on Programming Exercises Using Deep Learning. In *Proceedings of the 10th International Conference on Educational Data Mining*. 324-329.
- Wilson, K. H., Xiong, X., Khajah, M., Lindsey, R. V., Zhao, S., Karklin, Y., Van Inwegen, E.G., Han, B., Ekanadham, C., Beck, J.E., Heffernan, N. (2016). Estimating student proficiency: Deep learning is not the panacea. In *In Neural Information Processing Systems, Workshop on Machine Learning for Education*.
- Xiong, X., Zhao, S., Van Inwegen, E. G., & Beck, J. E. (2016). Going Deeper with Deep Knowledge Tracing. In *9th International Conference on Educational Data Mining*.
- Zaidah I., Daliela R. (2007). Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression. In *21st Annual SAS Malaysia Forum*.

Effects on Fostering Computational Thinking by Externalizing a Solution with Construction of a Problem-Solving Model

Kazuaki KOJIMA^{a*} & Kazuhisa MIWA^b

^a*Learning Technology Laboratory, Teikyo University, Japan*

^b*Graduate School of Informatics, Nagoya University, Japan*

*kojima@lt-lab.teikyo-u.ac.jp

Abstract: Literature on education has paid considerable attention to computational thinking (CT), thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent. Many attempts to develop CT in problem-solving education have been made in K-12 education. This study proposes an approach to foster CT in curricula for general undergraduate students: students construct rule-based computational models of problem solving. We empirically investigated effects of construction of a rule-based model on externalizing a solution in problem-solving in terms of problem decomposition, one aspect of CT. Undergraduate students described knowledge required to solve a simple problem in the format of rule-based models before and after they constructed models of the problem for a production system. Results indicate that model construction improved student decomposition in knowledge externalization of the solution.

Keywords: Computational thinking, learning by modeling, production system

1. Introduction

Literature on education has paid considerable attention to computational thinking (CT), which refers to thinking like a computer scientist. In fact, CT has great influence not only in the natural sciences of physics, chemistry, and biology, but also in psychology, economics, literature, and psychiatry. In other words, the activity of computer scientists is now spreading into broad disciplinary areas.

Although CT has not yet been clearly defined, most researchers seem to agree on the definition by Cuny, Snyder, and Wing (2010): *thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent* (e.g., Aho, 2012; Brennan, & Resnick, 2012; Yadav, Mayfield, Zhou, Hambrusch, & Korb, 2014). Such skill must not be reserved only for experts in information engineering, science, and other academic areas, but instead, must be applicable by anyone who engages in problem solving. In this respect, various efforts to implement CT training in education should be emphasized. Above all, many attempts to develop CT in problem-solving education have been made in K-12 education (e.g., Barr, & Stephenson, 2011; Bocconi, Chiocciariello, Dettori, Ferrari, Engelhardt, Kampylis, & Punie, 2012; Grover, & Pea, 2013; Yadav et al. 2014). Such CT training generally adopts programming as a tool to represent problem solutions in learning activities. Popular systems used in CT training are graphical programming environments and web-based simulation authoring tools (Grover, & Pea, 2013), which are easy to use for learners who are not information-engineering students.

Here, we propose an approach to foster CT in curricula for general undergraduate students. In fact, in an article on CT, Jeannette M. Wing (2006) insists that professors in computer science should teach university freshmen subjects such as “a way to think like a computer scientist” in all departments, not just the department of computer science. In our approach, students experience modeling like a computer scientist. Actually, they construct rule-based computational models of problem solving. Scientists in cognitive science have used computational models as research tools for in-depth

understanding of the human mind. Such model construction is expected to improve skills for externalizing knowledge required in problem solving, and that must foster some components of CT.

The current study investigated effects of construction of a rule-based model on the externalization of knowledge required to solve a simple problem. We empirically confirmed how undergraduate students externalize knowledge of a problem solution and whether construction of a computational model improves students' knowledge externalization. We particularly focused on decomposition of a problem solution as one aspect of CT, which is described in the next session.

The remainder of this paper is structured as follows. Section 2 briefly explains part of our approach's theoretical background. Section 3 describes the empirical investigation's method for confirming the research questions. Finally, Section 4 provides study results and discusses them.

2. Theoretical Background

Many researchers have continuously discussed various aspects of CT. Although CT does not yet have a clear definition, most researchers agree on some of its main components, for example, decomposition of problems (modularizing) and recognition and generalization of patterns, abstractions, and automation (algorithms) (*e.g.*, Barr, & Stephenson, 2011; Grover, & Pea, 2013; Krauss, & Prottzman, 2017).

In our approach, students construct computational models that can simulate problem solving for a production system. Therefore, they must disassemble solution processes in problem solving into separate operations and set conditions to adapt the operations to proper states. To implement operations in a computational model, students might cut certain details of the operations derived. Finally, students implement production rules under the specifications so that the production system reproduces the solution. These steps are regarded to involve the following four CT components proposed by Krauss and Prottzman (2017): *decomposition, pattern recognition, abstraction, and automation.*

Our preliminary study (Kojima, & Miwa, 2018) confirmed that model construction improved the pattern recognition in externalizing knowledge required for a simple problem when decomposition of the problem was supported. Students learned to describe more appropriate conditions for knowledge to solve the problem after they had experienced construction of a rule-based model. The current study focused on the skill of decomposition. Mvalo and Bates (2018) studied students' use of decomposition as a CT component in problem-solving tasks to design and troubleshoot computer networks using simulation software. In the study, focus group interviews to undergraduate and postgraduate students confirmed that they used decomposition in the tasks. However, this study reported only results of qualitative research in the domain of information engineering. To accommodate our goal to foster CT for general undergraduate students, we must precisely investigate how non-information-engineering students perform decomposition of knowledge. Therefore, we examined the following two research questions.

RQ1 How do undergraduate students externalize knowledge required to solve a problem within the format of rule-based models in terms of decomposition?

RQ2 Does construction of a computational model for the problem improve students' decomposition in knowledge externalization of the solution?

We have developed a framework for learning in which general undergraduate students create models of human cognitive processes. This framework uses a production system for novices called DoCoPro (an anywhere production system) (Miwa, 2008; Miwa, Nakaike, Morita, & Terai, 2009; Nakaike, Miwa, Morita, & Terai, 2009). We adapt learning by constructing models to a learning framework for fostering CT.

Figure 1 shows a screenshot of DoCoPro. A student creates a model of solving a problem by inputting the initial states in the working memory in the left frame, editing if-then format rules in the right frame, and simulating problem-solving processes by executing the model with the controller in the upper frame. For allowing novices to experience model construction, DoCoPro limits its constructs. Students have only to learn about if-then rules, matching, and some built-in functions (*e.g.*, functions to test whether two values are equal and to add an assertion to the working memory). It has no functions to perform simulation effectively and to represent human cognitive functions helpful in scientific research. Instead, it helps students examine rules through trial and error by providing functions to test the rules in a variety of ways. Furthermore, DoCoPro has instructional texts that help students learn how to create a

model with an example of a simple block-stacking problem in addition to basic concepts of a production system.

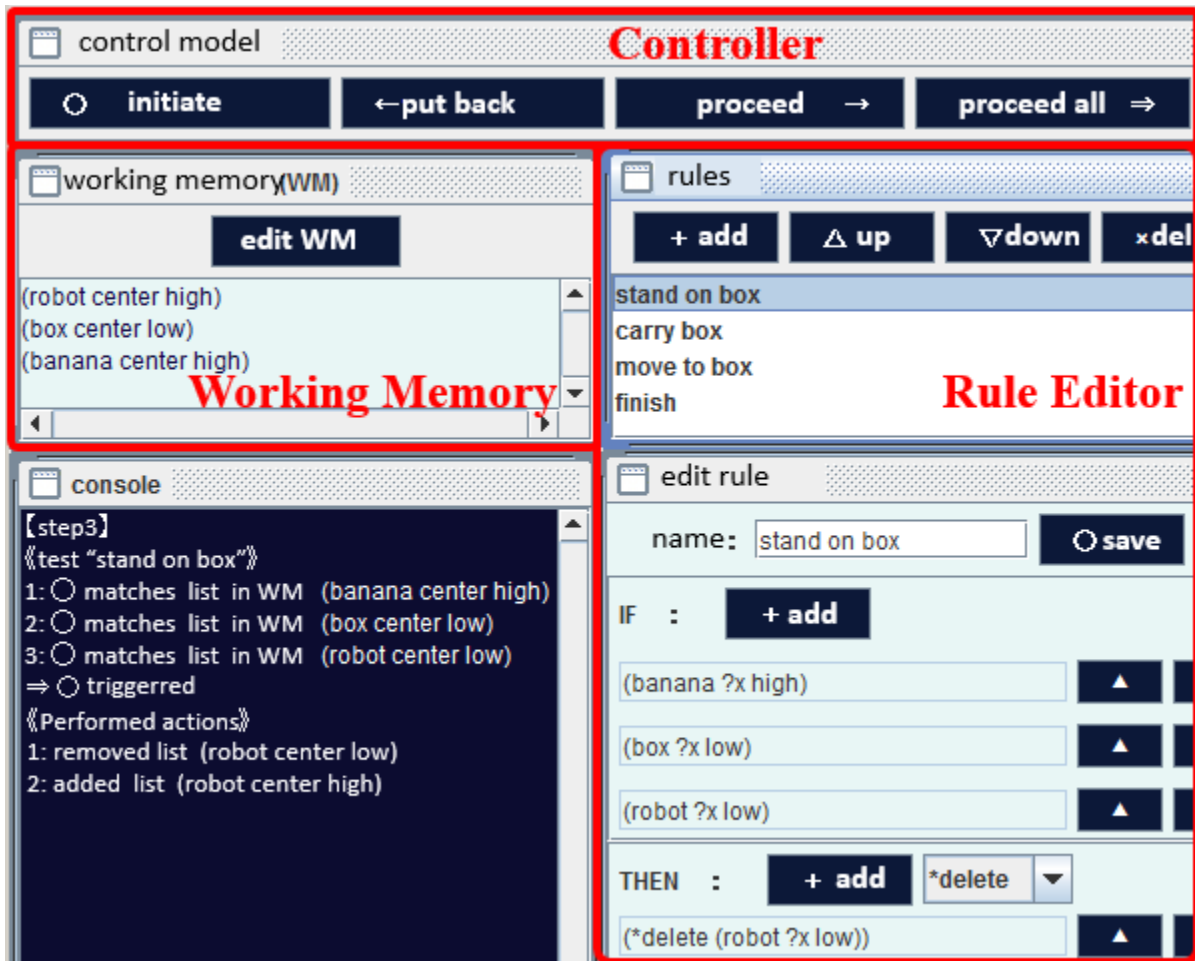


Figure 1. Part of Screenshot of DoCoPro.

3. Method

3.1 Procedures & Materials

We conducted lectures in which undergraduate students constructed models of problem solving in a cognitive science class. Although most students had experienced programming in other classes, they had not experienced training from experts in information engineering. In the cognitive science class, two lectures were conducted for model construction. In the first lecture, students learned about a production system with instructions from the lecturer and the first part of the instructional texts of DoCoPro. They actually experienced creating an if-then rule with DoCoPro.

Students were then presented a robot and banana problem, an altered version of the famous toy problem monkey and banana. Figure 2 illustrates initial and goal states of the problem. We used this quite simple problem because it requires problem decomposition but even novices are expected to succeed in model construction. When this problem was presented to students, they received Figure 2 and the following descriptions about the solution: “To have the robot get the banana, have the robot move to the same position as the banana. The robot can move into a high position by standing on the box. The robot can also carry the box.” Representation of the initial state (a) for DoCoPro was also presented as follows.

(robot door low)
 (box window low)
 (banana center high)

Students were asked to describe if-then rules necessary to solve this problem with natural sentences. We refer to this task as a pretest. Students were instructed to design general rules adaptable to various initial states. They completed the remainder of the instructional texts before the second lecture, which was 2 weeks after the first lecture.

In the second lecture, each student engaged in constructing a model of the robot and banana problem with DoCoPro. Students were again instructed to design general rules, and compose a model that can appropriately function regardless of the order of rules. After model construction, students again described the rules of the problem by using sentences (a posttest).

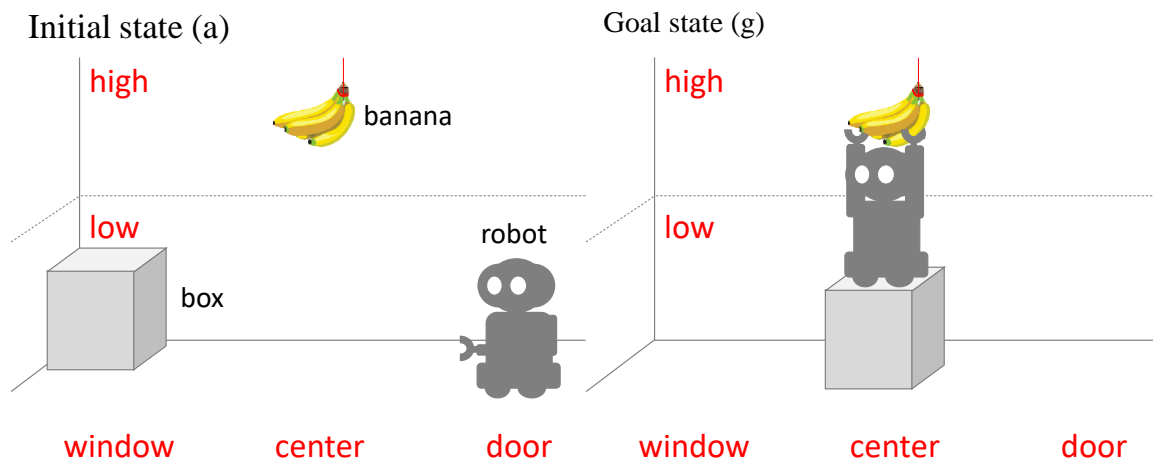


Figure 2. Initial and goal states in the robot and banana problem.

3.2 Data Analysis

We assessed operations in rules of models described by students in pre- and posttests. The best sequence of actions (rules) that can solve this problem includes the following four separate operations.

1. Robot moves to box (change the horizontal position of the robot to that of the box)
2. Robot carries box under banana (change the horizontal positions of the robot and box to that of the banana)
3. Robot stands on box (change the vertical position of the robot to high)
4. Robot gets banana (finish)

According to what operations were described and how they were incorporated in rules, we checked whether each student model had each of the following features.

- Normative Decomposition (ND) included four operations, and each operation was incorporated in different rules. A model with five rules comprising the four rules and one additional rule to finish the reasoning was also regarded as ND because it is natural to separate a functional rule only to halt problem-solving processes.
- Enhancing Operations (EO) included one or two rules incorporating operations that can successfully enhance a model. Actually, they were “robot moves to banana in a low position” and “robot gets down from box.”
- Surplus Decomposition (SD) included one or two operations that must not be separated from an operation in the best sequence. Actually, they were “robot lifts box” and “robot put box.” They are not independent because they are necessarily performed right before/after “robot carries box.”
- Lacking Operations (LO) means omitting one or more operations in the best sequence.
- Invalid Decomposition-Combining (ID-C) incorporated multiple separate operations into a single rule as an action in the best sequence. In a model with this feature, for example, one rule included combined operations such as “robot moves to box and carries it under banana.”
- Invalid Decomposition-Decombining (ID-D) broke an operation in the best sequence into two or more rules, except the case to break the rule to finish into “robot gets banana” and “halt the reasoning.” In a model with this feature, multiple rules included the same operation with different

parameters such as “robot moves to window (if box is at window side)” and “robot moves to center (if box is at center).”

- Invalid Operations (IO) included one or more operations that violated problem conditions or that could not be interpreted. In violated operations, for example, the box was supposed to move autonomously without the robot, and the robot can reach the banana without carrying the box itself. In another example, one rule directly describes the goal state (g) without the operations on the best sequence. Such a model cannot reproduce the solution process.

ND is regarded as a successful model, and SD, LO, ID-O, ID-D, and IO as failed because any of them can prevent a model from reaching the goal state with the best sequence through inappropriate decomposition or excessive operations. Although these seven features were independently assessed, ND is exclusive from the five failed features. EO can be held simultaneously with any others, and the five failed features can be held simultaneously with each other.

To examine RQ1 in terms of decomposition, we checked whether students described models of ND or those with features of the five failed models in the pretest. For RQ2, we compared pre- and posttests.

4. Results and Discussion

Figure 3 indicates proportions of students whose models had each of the seven features in the pre- and posttests. As the graph shows, NDs were few, and many student models had features of the five failed features in the pretest, whereas ND increased, and LO and IO decreased in the posttest. We compared numbers of each feature between the two tests by using the chi-square test; results indicated significant differences in ND ($\chi^2(1) = 26.01, p < .01$), EO ($\chi^2(1) = 3.94, p < .05$), LO ($\chi^2(1) = 46.10, p < .01$) and IO ($\chi^2(1) = 26.01, p < .01$). No significant differences were found in SD ($\chi^2(1) = 1.47, n.s.$), ID-C ($\chi^2(1) = 0.79, n.s.$) and ID-D ($\chi^2(1) = 2.66, n.s.$). These results confirmed that models with ND increased and those with lack of operations decreased after model construction with DoCoPro.

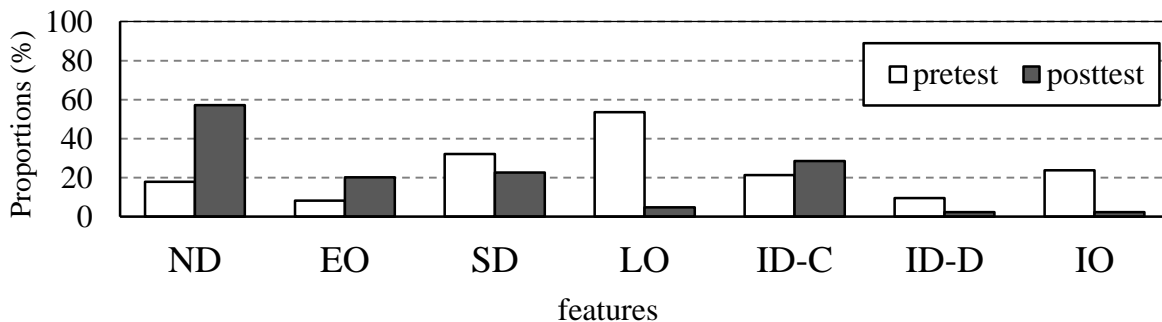


Figure 3. Proportions of students using each feature in pre- and posttests.

Those results indicated that in the pretest, most student models did not normatively decompose the solution of the robot and banana problem. In other words, undergraduate students could not necessarily externalize knowledge to solve the problem in terms of decomposition (RQ1). And ND increased in the posttest, indicating that construction of a model improved decomposition in knowledge externalization (RQ2). These facts reveal that construction of a rule-based model can foster one aspect of CT.

In the posttest, LO decreased. Students who omitted some operations of the best sequence in the pretest described them in the posttest. Most omitted operations were “finish,” and update the position of the robot¹ in “robot moves to box.” These operations may be implicit in human problem solving because people can solve this problem without awareness of them. In constructing a computational model, however, omitting such information causes error feedback. Such feedback must have improved

¹ In tests and model construction with DoCoPro, students wrote a name for each rule. Some student rules had names indicating “robot carry box,” but had operations including only a description such as “add ‘box is under banana.’”

student models in terms of lack of operations. This indicates that model construction can remove ambiguity in human representation of a problem's solution.

On the other hand, ID-C rather increased in the posttest although the difference was not significant. As explained above, models with this feature had duplicate operations with different parameters. Many students who described such models failed in model construction with DoCoPro in the second lecture. Their DoCoPro models fail to reach the goal state. Therefore, model construction by such students can be regarded as an inappropriate learning activity. Although we cannot precisely discuss this point due to page limitation, it alerted us to the necessity of learning support to guide appropriate model construction. Thus, one important our future work is investigating whether support for constructing an appropriate model enhances the effects and, if so, designing and implementing such support.

References

- Aho, A. V. (2012). Computation and computational thinking. *The Computer Journal*, 55(7), 832-835.
- Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? *ACM Inroads*, 2(1), 48-54.
- Bocconi, S., Chiocariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kampylis, P., & Punie, Y. (2016). Developing computational thinking: approaches and orientations in K-12 education. In G. Veletsianos (Ed) *Proceedings of EdMedia 2016* (pp. 13-18). NC: AACE.
- Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 AERA* (pp. 1-25).
- Cuny, J., Snyder, L., & Wing, J. M. (2010). Demystifying computational thinking for non-computer scientists. Retrieved May 15, 2019, from <http://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf>.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.
- Kojima, K., Miwa, K. (2018). Preliminary study on fostering computational thinking by constructing a cognitive model. In Y.-T. Wu, N. Srisawasdi, M. Banawan, J. C. Yang, M. Chang, L.-H. Wong, & M. T. Rodrigo (Eds) *Workshop Proceedings of ICCE 2018* (pp. 265-270). Taoyuan, Taiwan: APSCE.
- Krauss, J., & Prottzman, K.: *Computational thinking and coding for every student: the teacher's getting-started guide*. CA: Corwin (2017)
- Miwa, K. (2008). A cognitive simulator for learning the nature of human problem solving. *Journal of Japanese Society for Artificial Intelligence*, 23(6), 374-383.
- Miwa, K., Nakaike, R., Morita, J., Terai, H. (2009). Development of production system for anywhere and class practice. In S. C. Kong, H. Ogara, H. C. Arnseth, C. K. K. Chan, T. Hirashima, F. Klett, J. Lee, C. C. Liu, C. K. Looi, M. Milrad, A. Mitrovic, K. Nakabayashi, S. L. Wong, & S. Yang (Eds) *Proceedings of ICCE 2009* (pp. 91-99). Jhongli, Taiwan: APSCE.
- Mvalo, S., & Bates, C. (2018). Students' understanding of computational thinking with a focus on decomposition in building network simulations. In B. McLaren, R. Reilly, S. Zvacek, J. Uhomobhi (Eds) *Proceedings of CSEDU 2018* (pp. 245-252), vol. 1. Setubal, Portugal: SciTePress.
- Nakaike, R., Miwa, K., Morita, J., Terai, H. (2009). Development and evaluation of a web-based production system for learning anywhere. In S. C. Kong, H. Ogara, H. C. Arnseth, C. K. K. Chan, T. Hirashima, F. Klett, J. Lee, C. C. Liu, C. K. Looi, M. Milrad, A. Mitrovic, K. Nakabayashi, S. L. Wong, & S. Yang (Eds) *Proceedings of ICCE 2009* (pp. 127-131). Jhongli, Taiwan: APSCE.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.
- Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education*, 14(1), 5.

Towards Adaptive Provision of Examples During Problem Solving

Faiza TAHIR^{a*}, Antonija MITROVIC^a & Valerie SOTARDI^b

^a*Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand*

^b*School of Educational Studies and Leadership, University of Canterbury, New Zealand*

*faiza.tahir@pg.canterbury.ac.nz

Abstract: Intelligent Tutoring Systems (ITSs) are effective in supporting learning, as shown in numerous studies. The goal of our project is to develop an adaptive strategy that would be capable of identifying situations during problem solving in which the student would benefit from worked examples. As a first step towards developing such a strategy, we conducted a pilot study in the context of SQL-Tutor, a mature ITS that teaches database querying. The participant could ask for a worked example whenever he/she wanted during problems solving. After each example, the participant specified whether the example was useful, and whether additional examples were needed. Participants' facial expressions and eye gaze were recorded. The findings show that the participants generally found examples useful, although in some cases they stated additional examples would be beneficial. The analysis of the eye gaze shows that students compared provided examples to their own solutions. Affect analysis shows that negative emotions reduced while engagement increased when participants viewed examples, and immediately after examples.

Keywords: intelligent tutoring system, problem solving, worked examples, eye tracking, affective modeling

1. Introduction

Intelligent Tutoring Systems (ITSs) have been proven to be very effective in supporting learning (Kulik & Fletcher, 2016; Mitrovic, 2012; VanLehn, 2011). The main activity in ITSs is problem solving, where the student receives help from the ITS adaptively, based on his/her actions and knowledge. On the other hand, there is a long tradition of research on learning from worked examples (WEs), starting from 1950s (Atkinson et al., 2000). A worked example contains the problem statement and a step-by-step solution with accompanying explanations. Atkinson et al. (2000) suggested the importance of worked examples in early stages of skill acquisition. Learning can also be increased when WEs are combined with self-explanation (Große & Renkl, 2007), problem solving (Cooper & Sweller, 1987), faded examples, (Renkl & Atkinson, 2003), or erroneous examples (Große & Renkl, 2007).

Examples have also been found beneficial when incorporated into ITSs. ELM-PE is one of the first ITSs to incorporate examples and their explanations (Burow & Weber, 1996). SE-Coach (Conati, Larkin, & VanLehn, 1997) guided students to self-explain examples; on the basis of student explanations and student model, it estimated the student understanding of a particular example. EA-Coach (Muldner & Conati, 2007) provided examples adaptively, based on learners' characteristics and example utility. Another study revealed the positive effects of providing WEs adaptively by fading their steps in a cognitive tutor (Salden et al., 2009). A study with SQL-Tutor compared learning from problems only, WEs only, or alternatively provided examples and problems to learners, found that a mixture of WE and problem solving resulted in best learning outcomes (Najar & Mitrovic, 2014). In follow-up studies, Najar and colleagues (2015) showed that adaptive selection of learning activities resulted in highest learning gains. Later on, erroneous examples were introduced in SQL-Tutor and proved to be helpful for advanced learners (Chen, Mitrovic & Mathews, 2019). Another study used a concept-based similarity approach to select most similar examples for the learner, when the learner fails to complete a Java program (Hosseini & Brusilovsky, 2017).

Most of these studies focused either on adaptive strategies for presenting WEs and/or problems, or on adaptive provision of example steps. However, there is a lack of research on adaptive strategies for providing examples to students when they need help during problems solving. In order to fill this gap, we designed and conducted a pilot study with SQL-Tutor (Mitrovic, 2003), the goal of which was to observe when and how students use worked examples during problem solving.

We start by presenting the worked example version of SQL-Tutor used in the pilot study, and then describe the procedure in Section 3. Section 4 presents the findings, while Section 5 presents conclusions.

2. Experimental Setup

The version of SQL-Tutor used in the pilot study contained ten problems. The screenshot in Figure 1 shows the problem-solving environment of SQL-Tutor. At the top of the page, there are several buttons allowing the student to change the database, select a problem, look at the history of the session or his/her student model, run the query, ask for help or exit the system. For each problem, there was one WE that was isomorphic to the problem, using the same database and same domain principles. Figure 1 shows a WE, which includes the problem statement, the solution accompanied with an explanation. After the explanation, the student was required to specify which clause of the Select statement he/she had difficulty with, and then to specify whether the example was useful, and whether additional examples were needed. The three questions were mandatory. In the study, we used the Tobii eye tracker to record the participant's eye gaze, and iMotions to record facial expressions.

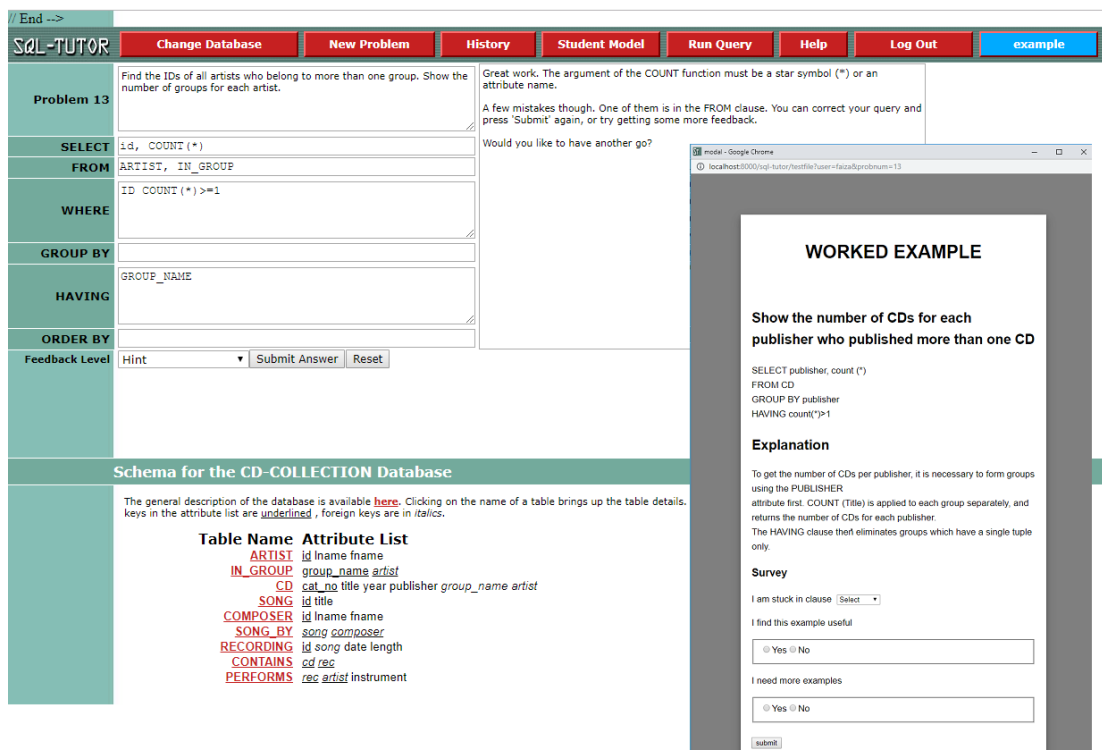


Figure.1 Screenshot of the worked example mode of SQL-Tutor

3. Procedure

We recruited seven undergraduate and three postgraduate students (two females, eight males), who were all studying Computer Science. Six participants were domestic students, while the remaining four were international (three Asian, and one Latin American). Five participants were aged 18-23, three 24-29 and two 30-35. All participants were familiar with SQL, and some of them have worked with

SQL-Tutor before the study. Each student had an individual session (40 minutes long), and was awarded a \$20 voucher for their participation.

At the beginning of the study, the participants provided informed consent, and filled a short questionnaire, in order to collect basic information about participants and their level of familiarity with SQL-Tutor. The participant sat in front of the Tobii screen, and the standard Tobii calibration was completed. The calibration test took 6 seconds, and the experiment started only if results were excellent. Otherwise, the position of participant was readjusted and recalibration took place. The experimenter sat to the other side, and monitored the participant’s face and eye gaze captured by both iMotions and Tobii. This monitoring ensured that during experiment full face of participant was captured so iMotions could record the facial features properly. The participants were instructed to solve at least five problems in SQL-Tutor, and to ask for examples as needed.

iMotions recorded participants’ facial expressions, which needed to be post-processed first, and later converted into action units and emotions by using the Affectiva AFFDEX facial expression analysis engine. After post processing, only those recordings with the AFFDEX sampling rate quality higher than 80% were included in the analyses (i.e. in 80% of samples it was possible to identify facial features). Affectiva AFFDEX generated probabilistic estimates for the seven emotions (anger, disgust, surprise, sadness, joy, fear and contempt) based on macro-expressions (lasting 0.5-4 seconds) of each participant. We selected the amplitude-based thresholding technique to focus on the strongest emotion.

4. How Much have Participants used Examples?

Five participants have not used SQL-Tutor prior to the study. Two participants used SQL-Tutor a lot, while the remaining three had limited experience with the system. Table 1 shows how many participants attempted and completed each problem, asked for examples, and how much time was spent on average on the problem/example. The *Example* column specifies the number of participants who viewed examples. The participants mostly solved the problems in the provided order, from the easiest to the hardest. On average, participants attempted 6 problems (sd = 1.89). The four most difficult problems were attempted much less often, and no one completed problems 9 and 10. The participants completed 62% of the problems they attempted, and viewed examples in 59% of the cases. For problems 1-5, as the problem complexity grows, the example use increases. In problems 2, 4, 5 and 10, participants viewed the examples more than once. When they viewed examples for the first time, they spent on average a minute viewing them. Upon the second and third viewing, this time decreased to 10-20 seconds only. The average time per example is proportional to the average time on problem.

Table 1

Problem, Example and Feedback Use; Time in minutes for problems, and in seconds for examples

Problem	Attempted by	Completed by	Time/ problem	Example	Time/ example	Feedback
1	9	9	1.48 (0.95)	2	40 (7.07)	1
2	8	7	4.18 (2.02)	7	41 (9.72)	3
3	9	7	1.6 (1.34)	1	25 (0)	2
4	10	6	5.58 (3.18)	7	72 (30)	4
5	9	4	6.5 (3.79)	6	46 (54)	5
6	5	3	3.46 (2.62)	3	38 (7.63)	2
7	2	1	2.1 (0.14)	2	16.5 (12)	0
8	3	1	2.05 (0.07)	2	26 (20)	1
9	3	0	2.2 (0.52)	3	30 (16)	0
10	3	0	7.3 (5.23)	3	43 (15.2)	0

The *Feedback* column of Table 1 shows the number of participants who have explicitly required specific levels of feedback (such as hint, partial/complete solution) while solving problems. More participants have used feedback for the easier problems (1-5) than for the rest of problems. This trend is opposite to how participants used examples.

Table 2

Participants' Opinions on Examples

Example	Viewed by	Useful	More examples
1	2	1	2
2	7	7	1
3	1	1	1
4	7	5	2
5	6	4	2
6	3	2	0
7	2	1	1
8	2	2	0
9	3	3	0
10	3	2	1

Table 2 shows participants' responses to the three questions given with WEs. In 78% of the cases participants found them useful, and in 36% of these cases, they wanted more WEs. For complex problems (problems 7-10), when completion rate was low (below 20%), the participants found examples very useful (80% of the cases), even when they have not completed those problems. This shows that regardless of success in problem solving, the participants found the examples useful. Please note that our study was voluntary, and therefore there was no need for students to complete all problems.

5. Eye Gaze Analysis

We analyzed the eye tracking data to determine how the participants read worked examples. Such analysis allows us to understand whether the participants use WEs appropriately. Each WE is isomorphic to the problem, and we expected students to compare the solution provided in the WE to their solution. The area of interest (AOI) was defined to cover the whole example (i.e. title, solution and explanation). The metrics included in eye tracking analysis are: (1) *Time in AOI*, i.e. the total time spent looking at the AOI; (2) *Visits*, i.e. the number of times the participant's eye gaze returns to the AOI; (3) *Fixation count*, showing the total number of fixations within the AOI; (4) Duration of the first fixation on the AOI; and (5) the average fixation duration in AOI. Table 3 shows the metrics for the ten examples (including multiple viewings) averaged over all participants who viewed those examples.

The *Time in AOI* column provides the average time spent by participants while examining a WE. The average number of visits to the AOI seems to increase as problems become more complex. As the number of example steps grows in later examples, the participants looked more often towards the problem solving area and schema.

Table 3

Averages (Standard Deviations) for Eye Tracking Metrics. Times are Reported in Seconds

Example	Time in AOI	Visits	Fixation count	First fixation duration (s)	Fixation duration
1	40 (2.8)	13 (1.4)	43 (6.3)	.2 (.07)	.19 (0)
2	33 (18.5)	9 (9.06)	38 (30)	.14 (.05)	.21 (.030)
3	24 (0)	5 (0)	7 (0)	.24 (0)	.23 (0)
4	59 (30)	27 (25)	132 (127)	.25 (.073)	.24 (.03)
5	59 (61.9)	18 (18.5)	111 (116.9)	.2 (.094)	.24 (.05)
6	51 (20.5)	23 (2.12)	82 (21)	.23 (.063)	.2 (.04)
7	23 (0)	13 (0)	51 (0)	.22 (0)	.25 (0)
8	49 (24.9)	32 (14)	107 (94)	.46 (.37)	.24 (.05)
9	34 (4.7)	10 (6.5)	55 (32)	.19 (.053)	.22 (.04)
10	47 (16.3)	20 (9.5)	107 (52)	.31 (.31)	.22 (.02)

The average fixation count shown in Table 3 is highest in examples 4, 5, 8 and 10. The highest fixation count shows that these participants did not just glance over those examples, but studied them thoroughly, not only the first time but also for the second or third viewing. The high fixation count and average duration of the first fixation on more complicated examples strengthens the above findings that as the number of example steps grows, more fixations were recorded.

6. Affect Analysis

Affectiva AFFDEX analyzes facial expressions and reports the values of seven emotions: anger, sadness, surprise, disgust, joy, contempt and fear, based on Ekman's (1999) categorization of emotions. However, these are general emotions, not the emotions specific to learning (Baker et al., 2010; Craig et al., 2004). Woolf and colleagues (2009) suggested mappings between Ekman's basic emotions to learning-related emotions: joy mapped to excitement, anger mapped to frustration, surprise mapped to boredom, and fear mapped to anxiety.

In line with the above mentioned research, we considered anger, joy, fear and surprise. We additionally included engagement, which is also crucial for learning (Craig et al., 2004; D'Mello, Picard, & Graesser, 2007). We observed some general trends. At the start of each problem, the dominant emotion was surprise, and once the problem was solved, the dominant emotion was joy. When participants received feedback from SQL-Tutor (upon submitting their solutions), the level of surprise was higher. In those situations when participants were able to solve the problem after receiving feedback, again the level of joy was increased. However, if they were not able to solve the problem, we noticed higher levels of anger, showing the participants' frustration.

Another event of interest is when students asked for examples. We analyzed the emotions for three different time intervals: (1) one minute before example use, (2) during example use, and (3) one minute after example use. Firstly, during one minute before participants asked for examples, the dominant emotions were anger and surprise, which seem to suggest that participants asked for examples when they were frustrated. Engagement increased and surprise decreased during or after working with examples. Fear was the least detected emotion; it decreased while and after working with examples and increased slightly when they were working on examples. Joy increased when they were working with examples, and immediately after, when the participants were able to complete problems after viewing examples. On the other hand, if the example did not help the participant solve the problem, we observed increased values for anger and surprise. In some of those cases, the participants asked for the example for the second time, and after that abandoned the problems. This is consistent with findings reported in the literature showing that frustration may lead to boredom, in which case learners lose interest in learning activities.

In summary, we found that participants asked for examples when the levels of anger (i.e. frustration) and surprise (i.e. anxiety) were elevated. Working with examples reduced such negative emotions and increased joy. After viewing examples, when participants turned again to problem solving, the intensity of negative emotions was low, but gradually increased if they were unable to solve the problem. The level of engagement increased for all participants during and after viewing examples. Therefore, examples have positive impact on participants' affective states, which will be helpful in learning with SQL-Tutor.

7. Conclusions

This paper presented the pilot study the goal of which was to analyze how participants use, study and feel about worked examples in their problem solving journey with SQL-Tutor. The results show that participants used examples extensively, particularly when the complexity of problems increased. Most participants agreed on the usefulness of examples and a few required more examples. This indicates the demand for examples during problem solving, regardless of success in problem solving. The eye gaze analysis revealed that participants tried to understand example structure by comparing examples with their solutions. Lastly, the positive impact of examples on participants' emotions is as examples reduced participant's negative emotions, and increased engagement and up to some extent joy.

The presented findings illustrate the need for and effectiveness of WEs, supported by cognitive and affective states of participants. These findings provide a starting point for developing an adaptive strategy for providing WEs adaptively, during problem solving. A limitation of our study is the small sample size. We plan to collect more data about how and when students use example in the forthcoming study in a large database course, which will enable us to develop and evaluate the adaptive strategy in follow-up studies.

Acknowledgements

We thank Jay Holland for helping with the study, as well as our participants.

References

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of educational research*, 70(2), 181-214.
- Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Human-Computer Studies*, 68(4), 223-241.
- Burow, R., & Weber, G. (1996). *Example explanation in learning environments*. International Conference on Intelligent Tutoring Systems, pp. 457-465.
- Chen, X., Mitrovic, A., & Matthews, M. (2019). Learning from Worked Examples, Erroneous Examples and Problem Solving: Towards Adaptive Selection of Learning Activities. *IEEE Transactions on Learning Technologies*. 10.1109/TLT.2019.2896080
- Conati, C., Larkin, J., & VanLehn, K. (1997). A computer framework to support self-explanation. In *Proc. Int. Conf. Artificial Intelligence in Education* (Vol. 39, pp. 279-276).
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347.
- Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media*, 29(3), 241-250.
- D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22(4), 53-61.
- Ekman, P. (1999). Facial expressions. *Handbook of cognition and emotion*, 16(301), e320.
- Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17(6), 612-634.
- Hosseini, R., & Brusilovsky, P. (2017). A study of concept-based similarity approaches for recommending program examples. *New Review of Hypermedia and Multimedia*, 23(3), 161-188.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42-78.
- Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, 13(2-4), 173-197.
- Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User modeling and user-adapted interaction*, 22(1-2), 39-72.
- Muldner, K., & Conati, C. (2007). Evaluating a Decision-Theoretic Approach to Tailored Example Selection. *Proc. IJCAI*, pp. 483-488.
- Najar, A. S., & Mitrovic, A. (2014). Examples and tutored problems: is alternating examples and problems the best instructional strategy? *Research & Practice in Technology Enhanced Learning*, 9(3), 439-459.
- Najar, A.S., Mitrovic, A., & Neshatian, K. (2015). Eye tracking and studying examples: how novices and advanced learners study SQL examples. *Journal of computing and information technology*, 23(2), 171-190.
- Najar, A. S., Mitrovic, A., McLaren, B. (2015) Examples and tutored problems: adaptive support using assistance scores. *Proc. 24th Int. Joint conference on Artificial Intelligence*, pp. 4317-4323.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational psychologist*, 38(1), 15-22.
- Salden, R. J., Alevan, V. A., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: redundant or synergistic forms of support? *Topics in Cognitive Science*, 1(1), 203-213.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Woolf, B., Woolf, B., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *Int. Journal of Learning Technology*, 4(3-4), 129-164.

Can EEG signal predict learners' perceived difficulty?

Aman KUMAR^a, Pankaj CHAVAN^b & Ritayan MITRA^{b*}

^a*Department of Computer science and Engineering, Thapar Institute of Engineering and Technology, India*

^b*IDP in Educational Technology, Indian Institute of Technology Bombay, India*

*rmitra@iitb.ac.in

Abstract: This study presents an approach to predict learner's perceived difficulty using features extracted from electroencephalography (EEG) data. We demonstrate how EEG signals can be used effectively to estimate learner's perceived difficulty of learning content. Student self-reports of perceived difficulty and EEG data were gathered from 9 participants who watched a video lecture. A machine learning model with random forest classifier achieved a maximum accuracy of 75.24% in estimating perceived difficulty. Furthermore, the model predicted the difficulty level of the entire video lecture for individuals fairly well. Our results have implications for intelligent tutoring systems which aim at providing the learner with an adaptive and personalized learning environment.

Keywords: Electroencephalography (EEG), DEBE feedback, perceived difficulty, intelligent tutoring systems (ITS), machine learning

1. Introduction

Assessing and monitoring the mental states of learners is important in the teaching-learning process. Detecting the mental states of the learner by collecting appropriate data and providing a personalized experience for every learner has always been the holy grail of adaptive learning (Shawky & Badawi, 2018). Intelligent tutoring systems (ITS) are known for their focus on providing the learner with an adaptive and personalized learning environment. The learner model contains knowledge about the learner's cognitive-affective states and guides individualized learning trajectories (Nkambou, Mizoguchi, & Bourdeau, 2010).

There are several approaches to create learner models. A learner's behavioural data such as keystrokes (Bixler & D'Mello, 2013), clickstream data (Baker et al., 2008; D'Mello, 2017) etc. can be used to infer learning processes. For example, keystroke data was used to estimate engagement of learners engaged with writing task (Bixler & D'Mello, 2013), and clickstream data to model whether the learner is engaged in gaming the system (Baker et al., 2008) etc. Another approach is to use a variety of psychophysiological measures to measure different mental states. Eye tracking has been demonstrated to be a good indicator of cognitive workload (Schultheis & Jameson, 2004), attention (Conati, Merten, Muldner, & Ternes, 2005), engagement (D'Mello, Olney, Williams, & Hays, 2012) and also boredom (Jaques, Conati, Harley, & Azevedo, 2014). Facial emotion recognition has been used to measure frustration (Sidney et al., 2005) and perceived difficulty of learners (Whitehill, Bartlett, & Movellan, 2008). A multimodal approach has also been found to be effective in such applications for monitoring cognitive and affective states within an ITS (Lane & D'Mello, 2019). Pressure-sensitive mouse and chair, and galvanic skin response are some other less commonly used psychophysiological sensors that have been used (Arroyo et al., 2009; Kapoor, Burleson, & Picard, 2007). Multimodal approaches have usually yielded good accuracy in predicting mental states (Arroyo et al., 2009; Kapoor et al., 2007). A relatively less common data in this context is that from electroencephalogram or EEG, which measures brain waves. It has also shown promise in capturing changes in the attention and cognitive workload of learners in ITS (Chaouachi, Jraidi, & Frasson, 2015).

As evident in the aforementioned discussion, some commonly investigated parameters in ITS research are cognitive workload, attention, engagement, frustration, boredom etc. The knowledge of

learners' perceived difficulty is also an important (Lane & D'Mello, 2019), albeit relatively less studied variable. A learner may *perceive* the content difficult to understand (regardless of the true difficulty of the topic), and his perception may differ from other students. Even then, such perception have been shown to decrease interest and increase boredom directing the focus more towards the feeling of the negative affect which can further lower the attention and motivation, eventually resulting in poor performance (Pekrun, Goetz, Titz, & Perry, 2002). There have been several attempts to measure perceived difficulty in learners (Pham & Wang, 2018; Whitehill et al., 2008) and adjust instructional material accordingly.

The DEBE framework has been recently proposed as a systemized way to collect continuous, fine-grained feedback from students on their levels of perceived difficulty (and affective states) during a lecture (Mitra & Chavan, 2019). We use student-self reports in the form of DEBE feedback to train an EEG machine learning algorithm that can be used to predict levels of perceived difficulty when interacting with a video lecture. In this study, we investigate the potential of a low-cost consumer-grade EEG device, MUSE™, in predicting perceived difficulty in learners.

2. Method

2.1 Experimental set-up

The setup of this experiment was designed keeping in mind that it should be relevant to, and closely resemble real-world conditions as opposed to providing elementary cognitive tasks that are often used in psychology studies (e.g., simple arithmetic to measure cognitive load). For this purpose, we obtained a video recording of 47 minutes duration for a lecture from the Advanced Heat Transfer course offered at the university. Ten students, aged 22-29, taking this course participated in the study. However, we have used the EEG data of 9 participants for analysis due to the presence of high noise and missing values (>25%) in one of the samples. The students were asked to come to the laboratory individually and watch the lecture video. The Institute Ethics Committee approved the study (IEC Approval Letter_Proposal No. IITB-IEC/2018/004). The students watched the video lecture and provided DEBE feedback on four parameters, namely, whether the lecture was easy, difficult, engaging or boring? The students could press the buttons as often they felt any of the four states in real-time while watching the video lecture. Details of the experimental setup and related work have been provided in (Chavan, Gupta, & Mitra, 2018). This feedback was unprompted and the students could click whenever they felt any of the four cognitive/affective states. We used a MUSE™ EEG device having 4 electrodes at AF7, AF8, TP9 and TP10 according to the 10-20 international system, sampling at 256 Hz frequency to record the EEG signals. For the purpose of this study, we chose to focus only on the instances when the students clicked either difficult or easy (engaging and boring clicks were not analyzed) with the understanding that this represents their perceived difficulty of the topic.

2.2 Data pre-processing

EEG data is noisy, complex and suffers from the curse of high dimensionality. Hence, this step aims at cleaning the EEG signal, managing its complexity by filtering out less relevant information from the signal and reducing dimensionality by selecting a subset of relevant channels. Therefore, we decided to look at only the frontal lobes that have been shown to be active during higher order processing (Zarjam, Epps, Chen, & Lovell, 2013). The data from the two temporal lobes were discarded. Besides the lower relevance of those areas for the task, they are also susceptible to more artefacts such as jaw clenches (Kappel, Looney, Mandic, & Kidmose, 2017). A 60 Hz notch filter was applied uniformly to remove the power line artefact. The cleaned data were then chunked into one-minute windows preceding each difficulty or easy click. Each such segment was divided into 5-sec epochs and all 12 epochs thus formed were categorized as either easy or difficult.

EEG data analysis is usually confined to either the frequency or the time domain. However, this dualistic approach of data analysis completely discards the information contained in one or the other domain. An alternative approach would be to retain information from both domains for the analysis. Discrete wavelet transform (DWT) is one such analysis that uses information from both domains. The

DWT method is widely used for the time-frequency analysis in EEG signals, also due to non-stationary characteristics of EEG signals and DWT does not assume signals to be stationary. Wavelet Packet Decomposition (WPD) is a wavelet transform method where the signal is passed through more filters than the discrete wavelet transform (DWT) resulting in much more extensive decomposition and offers richer signal analysis. Hence, we have used WPD for our analysis. The WPD method decomposes the signal using successive high pass and low pass filters, and this time-scale representation is generated by dilation and translation of a mother wavelet. As the choice of mother wavelet depends on the application, researchers have mentioned that Daubechies order-4 (db4) is the most suitable mother wavelet for EEG signal analysis (Adeli, Zhou, & Dadmehr, 2003). Hence, we applied level-5 Wavelet Packet Decomposition using Debauchies-4 mother wavelet to decompose the signal and then extracted approximate delta (0-4Hz), theta (4-8Hz), alpha (8-16Hz), beta (16-32Hz) and gamma (32-64Hz) bands.

2.3 Machine learning

Feature extraction aims at describing the EEG signal by a few relevant values called “features”. Features were extracted from the EEG signals to construct features vectors (samples) for input into the machine learning models. Such features should capture the information embedded in the EEG signals that is relevant. In passive brain-computer interface (BCI) applications, entropy and energy are one of the most effective and commonly used features. Furthermore, researchers have found these to be statistically significant in determining cognitive states (Zarjam et al., 2013). Therefore, energy and Shannon entropy were extracted from each of the five frequency bands (alpha, beta, etc.) from both the frontal channels giving rise to a feature vector with 20 values ($2*5*2=20$). After extracting the feature vectors from the epochs, we created datasets corresponding to difficult and easy feedback epochs for each individual. The individual datasets were normalized (z-normalization) to account for individual differences and then merged to a single dataset of feature vectors. Initially, we had 778 instances of the difficult class and 384 instances of the easy class but, to avoid overrepresentation of the difficult class, we selected random 384 samples out of 778 samples for the difficult class. Therefore, our dataset comprised 384 instances each for difficult and easy class each.

Finally, we compared performances of various ML classifiers, namely, support vector machine (SVM), decision tree, artificial neural network (ANN) and random forest, in order to optimize and achieve the best results. We used hyperparameter tuning to achieve optimal accuracy. To test and validate the models thoroughly, we used k-fold cross-validation ($k = 10$). In k-fold cross-validation, the dataset is partitioned into k equal-sized non-repeating complementary subsets. Then, out of k subsets, k-1 subsets are used as training set, and 1 subset is retained as the validation set. This process is repeated k times, with each of the k subsets used exactly once as the validation set. All reported values are averages of the 10 iterations of this cross-validation.

3. Results and discussion

3.1 Performance of classifiers

We implemented SVM using three different kernels: linear, polynomial (degree = 3) and gaussian radial basis function (rbf). Further, for each kernel, we tuned the penalty parameter (C) of the error term. The best accuracy was 62.23% with the rbf kernel and $C = 10$. Several ANN architectures were implemented with Rectified Linear Unit (ReLU) and “logistic” as the activation function for the hidden layers and output layer, respectively. A maximum accuracy of 64.37% was achieved with two hidden layers, each having 64 neurons. The decision tree classifier with two splitting criterions, gini impurity (Gini) and information gain (entropy), achieved a maximum accuracy of 66.11% with splitting criteria = information gain (entropy). Finally, with a random forest classifier, we achieved a maximum accuracy of 75.24% with splitting criteria = Information gain (entropy) and the number of decision tree classifiers = 50.

3.2 Sensitivity analysis for random forest classification

We performed a sensitivity analysis by varying the length of the time window for data extraction preceding a feedback click and the length of epoch used to divide it further (Table 1). In this paper, we have achieved results with time window length = 60 seconds and epoch length = 5 seconds. Decreasing the epoch size increases the accuracy till 1 sec, after which the accuracy reduces, indicating a tradeoff between the size of dataset and the noise introduced by reducing the epoch size. For epoch sizes greater than or equal to 2 sec window size of 1 minute seem to be an ideal choice as accuracy falls on both sides. However, for smaller epoch sizes, increasing window size only decreases accuracy, possibly indicating the fact that the algorithm is sensitive to the noise introduced as a result of reducing epoch size with increasing window size.

Table 1

Best classification accuracies (in percent) with Random Forest classifier with varying length of the time window for data extraction and epoch length.

		Time window (sec)		
		30	60	120
Epoch (sec)	0.5	75.26	73.86	71.95
	1	76.41	75.68	72.19
	2	74.79	76.10	73.96
	5	72.72	75.24	72.10
	10	67.78	72.38	69.40

3.3 Predicting perceived difficulty from EEG data

To further test the potential of the ML model, we used the trained model to estimate the difficulty level for the entire lecture for each individual participant. The EEG data for the entire duration of the lecture was divided into contiguous epochs of 5 seconds. Each epoch was processed and features were extracted to prepare the feature vector. Subsequently, each epoch was classified using the trained random forest classifier, which had performed the best in our analysis. At this point, it is important to note that although accuracy gives an estimate of the performance of the ML algorithm it essentially uses 5 sec epochs of one minute windows prior to all clicks to make a prediction of the category of an epoch. It does not give us an understanding of what the classifier would predict for those epochs that are further away from the clicks - either after the click or more than one minute away from one. This section attempts to address this issue so that we can form another representation of the utility of such predictions.

The model predictions tallied fairly well with the observations. We see most clicks happen at predictable times (Fig 1 a, b and c). However, the number of false positives appear to be the problem when the predicted value appear to be either difficult or easy, but no clicks were observed. This could be due to a variety of reasons. The null hypothesis would be that the algorithm is indeed not suitable for more accurate predictions. One alternative hypothesis is that the student might be experiencing difficulty without recognizing it and hence does not click. The student could also be clicking only after feeling a particular state for a reasonable yet variable duration. The difference with these two possibilities being only in whether the student is aware of the cognitive-affective state s/he is in as the action (of clicking) would be identical.

Despite the possible shortcomings of the temporal accuracy of the predictions, it is encouraging to find that for participants with many difficult (easy) clicks the algorithm is indeed showing a prolonged state of difficulty or ease (Fig. 1b and c). It is to be noted that the training set included only 384 epochs each of easy and difficult segments (with possible overlaps between epochs). Therefore, the total training interval was equal to or less than 64 minutes (384*2*5 secs). The total predicted interval was $47*9 - 64 = 359$ minutes (47 minute lecture duration for 9 students). Also, the ML algorithm was trained on all positives (interval preceding either difficult or easy clicks) but the additional data for prediction (359-64 = 295 minutes) were all negatives (participants did not click either difficult or easy). This makes Figs. 1b and c even more interesting because it seems a participant clicking exclusively

difficult/easy tend to have long intervals of predicted difficulty/ease which could imply that, at the very least, the algorithm successfully predicts high average difficulty (ease) values for candidates who tend to click difficult (easy) several times during the lecture.

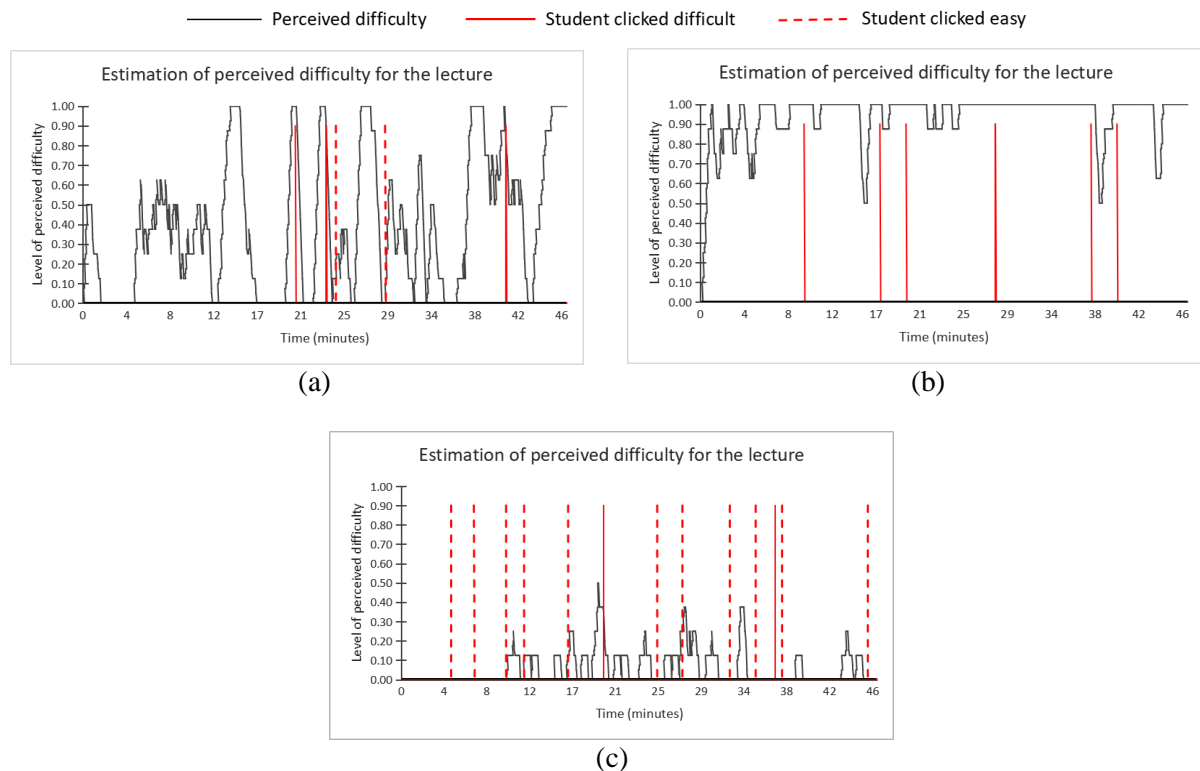


Figure 1. Model predicted difficulty levels for three participants. Y-axis values (0=easy and 1=difficult) are averages of 40 sec or 8 epochs. For example, if 4 out of 8 epochs were predicted to be difficult, then the corresponding predicted difficulty level would be 0.5.

4. Conclusion and limitations

We have demonstrated the viability of a low-cost EEG device in predicting perceived difficulty of learners. A random forest ML algorithm was able to predict perceived difficulty in participants with an accuracy of 75.24%. Furthermore, when presented with new data from the rest of the lecture the model predictions tallied closely with the clicking behavior – students who clicked difficult (easy) more often were predicted to have high (low) perceived difficulty.

In spite of the promising results, our research has some limitations. The machine learning model was trained with data from only 9 participants. We need a considerably larger dataset to train a robust and reliable classifier that can accurately classify EEG signals into the corresponding cognitive state. It is also likely that the accuracy would increase further with increasing size of the dataset. Another limitation of this study is that there was no way to validate the mental state of the participants when there was no click. Although we did find a consistently high or low value of perceived difficulty (Figs. 1b and c, respectively) for appropriate participants, we still have no validation for the interim gaps when there were no clicks. Another limitation of this study is the possibility that the EEG data were somehow affected by the clicking itself and could have introduced artefacts that were unrelated to the construct under investigation, namely, perceived difficulty. We are so far unable to rule out such a possibility.

Acknowledgements

This work was supported by an IRCC, IIT Bombay grant (17IRCCSG013) to Ritayan Mitra.

References

- Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*, *123*(1), 69–87.
- Arroyo, I., Cooper, D. G., Burlison, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *AIED*, *200*, 17–24.
- Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.
- Bixler, R., & D’Mello, S. (2013). Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 225–234. ACM.
- Chaouachi, M., Jraidi, I., & Frasson, C. (2015). Adapting to learners’ mental states using a physiological computing approach. *The Twenty-Eighth International Flairs Conference*.
- Chavan, P., Gupta, S., & Mitra, R. (2018). A novel feedback system for pedagogy refinement in large lecture classrooms. *International Conference on Computers in Education*, 464–469. Philippines.
- Conati, C., Merten, C., Muldner, K., & Ternes, D. (2005). Exploring eye tracking to increase bandwidth in user modeling. *International Conference on User Modeling*, 357–366. Springer.
- D’Mello, S. (2017). Emotional learning analytics. *Handbook of Learning Analytics*, 115.
- D’Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377–398. <https://doi.org/10.1016/j.ijhcs.2012.01.004>
- Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. *International Conference on Intelligent Tutoring Systems*, 29–38. Springer.
- Kapoor, A., Burlison, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, *65*(8), 724–736.
- Kappel, S. L., Looney, D., Mandic, D. P., & Kidmose, P. (2017). Physiological artifacts in scalp EEG and ear-EEG. *Biomedical Engineering Online*, *16*(1), 103.
- Lane, H. C., & D’Mello, S. K. (2019). Uses of Physiological Monitoring in Intelligent Learning Environments: A Review of Research, Evidence, and Technologies. In T. D. Parsons, L. Lin, & D. Cockerham (Eds.), *Mind, Brain and Technology: Learning in the Age of Emerging Technologies* (pp. 67–86). https://doi.org/10.1007/978-3-030-02631-8_5
- Mitra, R., & Chavan, P. (2019). DEBE feedback for large lecture classroom analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 426–430. ACM.
- Nkambou, R., Mizoguchi, R., & Bourdeau, J. (2010). *Advances in intelligent tutoring systems* (Vol. 308). Springer Science & Business Media.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students’ self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105.
- Pham, P., & Wang, J. (2018). Adaptive Review for Mobile MOOC Learning via Multimodal Physiological Signal Sensing-A Longitudinal Study. *Proceedings of the 2018 on International Conference on Multimodal Interaction*, 63–72. ACM.
- Schultheis, H., & Jameson, A. (2004). Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioral Methods. In P. M. E. De Bra & W. Nejdl (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems* (Vol. 3137, pp. 225–234). https://doi.org/10.1007/978-3-540-27780-4_26
- Shawky, D., & Badawi, A. (2018). A reinforcement learning-based adaptive learning system. *International Conference on Advanced Machine Learning Technologies and Applications*, 221–231. Springer.
- Sidney, K. D., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. *Affective Interactions: The Computer in the Affective Loop Workshop At*, 7–13.
- Whitehill, J., Bartlett, M., & Movellan, J. (2008). Automatic facial expression recognition for intelligent tutoring systems. *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 1–6. IEEE.
- Zarjam, P., Epps, J., Chen, F., & Lovell, N. H. (2013). Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Computers in Biology and Medicine*, *43*(12), 2186–2195.

Identifying Significant Indicators of Eye-movement and EEG-based Attention to Predict Reading Performance

Song LAI^a, Bingbing NIU^a, Jiaqi LIU^a & Fati WU^{a*}

^a*School of Educational Technology, Faculty of Education, Beijing Normal University, China*

*wft@bnu.edu.cn

Abstract: It is important to extract students' reading data to predict their reading performance. This study aims to identify significant indicators of eye-movement and EEG-based attention and to test their predictive effectiveness on reading performance. Data were collected from 56 undergraduate students who read an illustrated science text about geography. Out of 21 reading indicators, 16 were found to have a significant correlation with reading performance. The multiple regression model suggested that *Whole time*, *Text-diagram*, *Test-text*, *Medium attention*, and *High attention* were significant indicators. They predicted 62.5% of the variation in students' reading performance.

Keywords: Reading performance, significant indicators, eye-movement, EEG-based attention

1. Introduction

Due to the rapid development of information and communication technologies, digital reading has become a dominant trend (Ogata et al., 2015). The analysis of reading behaviors is important to understand readers' reading processes. The analytic results contribute to the revision of learning materials, provision of learning interventions, identification of less effective learning strategies, and extraction of more effective learning strategies.

In web-based learning, students are required to read learning materials before performing related tasks. To successfully comprehend materials, students repeatedly interact with them. One class of interaction behaviors in the field of Human-Computer Interaction is eye-movement. Eye-movement provides a natural and efficient way to observe students' behaviors from gaze (Klami, 2010). According to the eye-mind hypothesis, eye tracking can identify what is attracting students' attention and subconscious behaviors (Just & Carpenter, 1980). Consequently, eye-tracking data can be applied to analyze students' areas of interest (AOIs), visual search processes, and information processing (Rayner, 2009; Sun et al., 2017). Researchers have found that eye-movement indicators, such as mean fixation duration and saccades (Jian, 2017), significantly correlate with learning performance in reading.

Furthermore, when reading learning materials, students' brains generate plenty of electrical activities, recorded as waveforms using electroencephalogram (EEG). EEG reflects the inherent features of brain waves. Brainwave frequencies are closely related to attention state (Prinzel et al., 2001; Sirca, Onorati, Mainardi, & Russo, 2015). Apparently, EEG can determine changes in attention state. EEG-based attention is regarded as a psychological process comprised of focus and concentration, which can improve cognition speed and accuracy (James, 1983). Sustained attention has close relationship with learning performance (Steinmayr, Ziegler, & Träuble, 2010).

The goal of this study is to identify significant indicators and to explore predictive effectiveness of these indicators on students' reading performance by combining eye-movement and EEG-based attention data. All the pertinent reading data concerning eye-movement and EEG-based attention were extracted to make a bivariate correlation analysis with students' reading performance. From a total of 21 potential explanatory reading indicators, 16 indicators with significant univariate relationship with reading performance were chosen for inclusion in a multiple regression analysis. *Whole time*, *Text-diagram*, *Test-text*, *Medium attention*, and *High attention* were the indicators that significantly predicted reading performance, explaining over 60% of the variation in reading performance. The

results support the viewpoint that few reading indicators are able to accurately predict reading performance. Hence, the provision of reading materials that improve the level of learning attention should be of high priority during the design and practice of online reading.

2. Related Work

The surge of internet promotes the revolution and development of human learning style. As one of important symbols in the digital age, digital reading occupies a high proportion in the current learning scene, from paper to electronic, from single-media form to multimedia forms. In reading process, students produce massive interactions reflecting their engagement, which has an indelible impact on final performance (Liu, Chen, Zhang, & Rao, 2018). Hence, reading interaction data can be employed to determine the level of students' reading performance. This knowledge can help instructors to provide appropriate guidance for students in different reading states.

Eye-movement, revealing the allocation of visual attention in information search, is typically a reading interaction. Researchers have found that eye-movement data were able to estimate different levels of reading comprehension (J. Li, Ngai, Leong, & Chan, 2016; Sanches, Augereau, & Kise, 2018). Also, the associations between eye movements and reading performance were explored. For example, Everatt and Underwood (1994) found that gaze durations accounted for 9% in reading comprehension scores. S. C. Chen et al. (2014) demonstrated that eye-movement behaviors, especially the mean fixation duration and re-reading time in proportion, could successfully quantify students' performance. Similarly, Peterson et al. (2015) indicated that eye fixation and fixation sequence features were good predictors to assess learning performance. Moreover, features from eye-movement were extracted to predict reading performance by using machine learning approaches, whose results presented relatively high prediction effectiveness (Khedher & Frasson, 2016; Rajendran, Carter, & Levin, 2018). However, eye-movement data base on the external behaviors, which ignore students' internal cognitive states. By contrast, reading data based on physiological signals, such as EEG, are more reliable.

EEG measurements, recording electrical activity along the scalp, are correlated with students' goal-directed attention allocation revealed by their eye movements (Gwizdka, Hosseini, Cole, & Wang, 2017). There are strong correlations between individual differences in reading rate and brain activity, and reading rate can be predicted well by measurements of brain activity (Demb, Boynton, & Heeger, 1997). From the cognitive psychology perspective, EEG instantly shows the attention level (Ghassemi, Moradi, Doust, & Abootalebi, 2009; X. Li et al., 2011). Attention has a positive correlation with learning. The higher the level of attention, the more effective the learning. Also, C. M. Chen and Huang (2014) suggested that sustained attention and reading comprehension were strongly correlated, showing that sustained attention to learning materials is the prerequisite for effective learning. All of the aforementioned studies validated the effective predictive ability of their reading indicators and identified critical variables to predict reading performance accurately. However, notably few studies have been done to examine the effects on reading performance by combining eye movements and EEG-based attention. The combination may be more effective in reading performance prediction.

3. Methods

3.1 Participants

After preprocess, this study considered data of 56 undergraduate students. 24 of them were male, 32 were female, and their age ranged from 21 to 23. Students majored in non-geography and took fundamental geography courses in middle schools, so they already possessed some prior knowledge to address the geography science problem. All participants had normal or corrected-to-normal vision.

3.2 Materials

An illustrated science text was provided for participants to read, shown in Fig. 1. The article topic was the principle of tornado, consisting of a title, text section, illustration section and test section. The text

section included three paragraphs: the first briefly depicted tornadoes; the second presented the process of tornado formation during airflow motion; and the third introduced types of tornadoes. The illustration section related to Paragraph 2 of the text described the processes of airflow movement in detail. The test section included three test questions related to the article topic. Answers to the three questions were scored 0 to 6 according to their degree of correctness and completeness.



Figure 1. Six AOIs (title, paragraph 1, paragraph 2, paragraph 3, illustration and test) of the reading material. The participants in this study did not see the black frames.

3.3 Data Source

Data from 56 participants were recorded by the Tobii T60 eye tracker and the Neurosky mobile EEG headset. After calibration, participants were instructed to read carefully the material in less than 600s approximately. After finished reading, participants immediately completed the three questions. For these questions, the answers were scored by two independent raters who were blind to the purpose of the study. For each question, inter-rater reliability was evaluated with the Cohen's Kappa coefficient. The inter-rater reliability Kappa ranged from 0.891 to 0.975, showing fair to very good reliability. The score of each question was identified using the average score of two raters. Finally, the test grade was calculated as the sum of scores for all questions. Mean grades were 9.905 ($SD = 4.126$). Additionally, we first studied the whole reading material as one AOI and then divided it into three AOIs: text, diagram, and test, similar to Jian (2017). The collected features included 17 eye-movement indicators and 4 EEG-based attention indicators. EEG-based attention values, ranging from 0 to 100, were averaged to produce the attention value and categorized to three different types as Low (value under 40), Medium (value between 40 and 60), and High (value above 60). Details are shown in Table 1.

4. Results

To investigate eye-movement and EEG-based attention indicators that significantly correlated with students' reading performance, Pearson correlation analysis was performed. Then, to identify significant indicators that predicted student reading performance, inferential statistics were employed. The inferential analysis was a forward stepwise multiple regression, run on SPSS 20.0 with level of significance of .05.

The results about Pearson correlation coefficient (PCC) were presented in Table 1. There were 16 indicators (12 eye-movement indicators and 4 EEG-based attention indicators) with a statistically significant correlation ($p < 0.05$). Both *Text rate* and *Low attention* were negatively correlated with students' test grades. By contrast, others had a positive correlation. Within the significant subset of the reading indicators, 9 demonstrated a moderate effect size (PCC = 0.40-0.59). The remaining 7 indicators had a weak effect size (PCC = 0.20-0.39). Obviously, *Mean attention* (PCC = 0.536) and *High attention* (PCC = 0.592) have better effect size than all eye-movement indicators.

Although correlation coefficients are of great value in identifying the relationship of two indicators, correlation is simply a way to describe how two indicators vary together and cannot control for the other indicators that affect the dependent indicator, thereby giving false relationships. By contrast, linear regression gives coefficients when controlling for the other indicators, capturing in a better way the effect of independent indicators on dependent indicators (Lai, Sun, Wu, & Xiao, 2019; Zacharis, 2015). More importantly, a stepwise regression is a robust and valid method to find the best set of independent indicators that significantly predict student reading performance. Hence, this study employed a forward stepwise multiple regression, in which indicators that are not statistically significant in relation to the predictive power of the model are removed. From the set of significantly correlated eye-movement and EEG-based attention indicators, 16 potentially significant indicators were identified for inclusion in a multiple regression analysis. As presented in Table 2, *Whole time* ($B = -0.018$, $p < 0.01$), *Text-diagram* ($B = 0.014$, $p < 0.01$), *Test-text* ($B = 0.024$, $p < 0.01$), *Medium attention* ($B = 0.032$, $p < 0.001$), and *High attention* ($B = 0.025$, $p < 0.001$) were significant in predicting reading performance. The variance of student reading performance explained by the best fitting model was 62.5%. This showed that the 5 predictors contributed significantly to the predictive model. Moreover, the model was validated via 5-fold cross-validation with PCC, concordance correlation coefficient (CCC), mean absolute error (MAE), and root mean square error (RMSE) used as metrics for the fit. The experiments were conducted in WEKA 3.8. The experimental results showed that the regression model with a PCC of 0.621 ($p < 0.01$), CCC of 0.616, MAE of 2.969, and RMSE of 3.666, provided good prediction effectiveness. This confirmed the robustness of the model.

Table 1

Eye-movement and EEG-based attention indicators

	Attribute name	Description	PCC
Eye-movement	Whole time	Total reading time in whole article	0.498***
	Whole duration	Mean fixation duration in whole article	0.081
	Text time	Total reading time in text section	0.465***
	Text fixation	Number of fixations in text section	0.399**
	Text rate	Rate of total reading time	-0.322*
	Text duration	Mean fixation duration in text section	0.082
	Diagram time	Total reading time in diagram section	0.423**
	Diagram fixation	Number of fixations in diagram section	0.445**
	Diagram rate	Rate of total reading time	0.379**
	Diagram duration	Mean fixation duration in diagram section	0.070
	Test time	Total reading time in test section	0.320*
	Test fixation	Number of fixations in test section	0.349**
	Test rate	Rate of total reading time	0.066
	Test duration	Mean fixation duration in test section	0.112
	Text-diagram	Transitions of text to diagram	0.500***
Test-text	Transitions of test to text	0.415**	
Test-diagram	Transitions of test to diagram	0.297*	
EEG-based attention	Mean attention	Attention value in average	0.536***
	Low attention	Number of low attention value	-0.282*
	Medium attention	Number of medium attention value	0.443**
	High attention	Number of high attention value	0.592***

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2

Multiple regression analysis on reading performance

	B	SE	Beta	R ²
Whole time	-0.018	0.005	-0.633**	0.625
Text-diagram	0.014	0.005	0.350**	
Test-text	0.024	0.009	0.305**	
Medium attention	0.032	0.007	0.567***	
High attention	0.025	0.004	0.712***	

Note: ** $p < 0.01$, *** $p < 0.001$.

5. Conclusion

The current study was conducted to explore the significant eye-movement and EEG-based attention indicators in reading and their predictive effectiveness on reading performance to build a predictive model. Students' reading performance is highly related to their engagement level, so measures that reflected the degree of engagement are specifically employed to predict reading performance. Eye-movement and EEG-based attention data are some of the most frequently examined engagement indicators in reading. However, few studies have applied the combination of eye-movement and EEG-based attention to predict students' reading performance. In this light, this study used the combination as predictors, including 21 reading indicator variables. A bivariate correlation analysis of these indicators identified 16 of them to be significantly associated with reading performance. The multiple regression model revealed that 62.5% of the variance in students' reading performance was explained by just five indicators: *Whole time*, *Text-diagram*, *Test-text*, *Medium attention*, and *High attention*. As expected, EEG-based attention indicators (*Medium attention* and *High attention*) presented stronger effect size and significance than eye-movement indicators (*Text-diagram* and *Test-text*). This indicated that EEG-based attention indicators, displaying the level of mental effort, were stronger predictors in the construction of reading performance prediction. The findings suggested that students can have trainings about how to improve their own engagement level or search for meaningful information during reading process to foster deep understanding of the reading material that would further improve their reading performance.

There are a number of limitations that may affect the overall generalizability of this study. First, the study is based on a small sample of students at a single university. Future studies may collect a larger data set from multiple universities to build more robust model of student reading performance prediction. Second, due to the short reading material displayed on a single screen, students had no click operations. Hence, no clickstream data, which may be effective to improve the accuracy of reading performance prediction, was obtained. Future studies may present longer reading material with additional pages to collect clickstream data. Finally, deep learning approaches might be considered to construct more predictive models.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document, especially the members of the AddictedtoLearning team at the Beijing Normal University for their invaluable help which made this work possible.

References

- Chen, C. M., & Huang, S. H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance. *British Journal of Educational Technology*, 45(5), 959-980.

- Chen, S. C., She, H. C., Chuang, M. H., Wu, J. Y., Tsai, J. L., & Jung, T. P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Computers and Education*, 74, 61-72.
- Demb, J. B., Boynton, G. M., & Heeger, D. J. (1997). Brain activity in visual cortex predicts individual differences in reading performance. in *Proceedings of the National Academy of Sciences*, 94(24), 13363-13366.
- Everatt, J., & Underwood, G. (1994). Individual Differences in Reading Subprocesses: Relationships Between Reading Ability, Lexical Access, and Eye Movement Control. *Language and Speech*, 37(3), 283-297.
- Ghassemi, F., Moradi, M. H., Doust, M. T., & Abootalebi, V. (2009). Classification of sustained attention level based on morphological features of EEG's independent components. In *2009 ICME International Conference on Complex Medical Engineering* (pp. 1-6).
- Gwizdka, J., Hosseini, R., Cole, M., & Wang, S. (2017). Temporal dynamics of eye-tracking and EEG during reading and relevance decisions. *Journal of the Association for Information Science and Technology*, 68(10), 2299-2312.
- James, W. (1983). *The principles of psychology*. New York: Holt.
- Jian, Y. C. (2017). Eye-movement patterns and reader characteristics of students with good and poor performance when reading scientific text with diagrams. *Reading and Writing*, 30(7), 1447-1472.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329.
- Khedher, A. B., & Frasson, C. (2016). Predicting user learning performance from eye movements during interaction with a serious game. In *EdMedia 2016 World Conference on Educational Media and Technology* (pp. 1504-1511).
- Klami, A. (2010). Inferring task-relevant image regions from gaze data. In *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)* (pp. 101-106).
- Lai, S., Sun, B., Wu, F., & Xiao, R. (2019). Automatic personality identification using students' online learning behavior. *IEEE Transactions on Learning Technologies*, Early Access.
- Li, J., Ngai, G., Leong, H. V., & Chan, S. C. F. (2016). Your Eye Tells How Well You Comprehend. In *Proceedings of International Computer Software and Applications Conference* (pp. 503-508).
- Li, X., Hu, B., Dong, Q., Campbell, W., Moore, P., & Peng, H. (2011). EEG-based attention recognition. In *2011 6th International Conference on Pervasive Computing and Applications* (pp. 196-201).
- Liu, Y., Chen, J., Zhang, M., & Rao, C. (2018). Student engagement study based on multi-cue detection and recognition in an intelligent learning environment. *Multimedia Tools and Applications*, 77(21), 28749-28775.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *Doctoral Student Consortium (DSC) - Proceedings of the 23rd International Conference on Computers in Education (ICCE 2015)* (pp. 401-406).
- Peterson, J., Pardos, Z., Rau, M., Swigart, A., Gerber, C., & McKinsey, J. (2015). Understanding student success in chemistry using gaze tracking and pupillometry. In *International Conference on Artificial Intelligence in Education* (pp. 358-366).
- Prinzel, L. J., Pope, A. T., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Prinzel, L. J. (2001). *Empirical Analysis of EEG and ERPs for Psychophysiological Adaptive Task Allocation*. NASA/TM-2001-211016.
- Rajendran, R., Kumar, A., Carter, K. E., Levin, D. T., & Biswas, G. (2018). Predicting Learning by Analyzing Eye-Gaze Data of Reading Behavior. In *Proceedings of the 11th International Conference on Educational data Mining* (pp. 455-461).
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
- Sanches, C. L., Augereau, O., & Kise, K. (2018). Estimation of reading subjective understanding based on eye gaze analysis. *PLoS ONE*, 13(10), E0206213.
- Sirca, F., Onorati, F., Mainardi, L., & Russo, V. (2015). Time-varying spectral analysis of single-channel EEG: Application in affective protocol. *Journal of Medical and Biological Engineering*, 35(3), 367-374.
- Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement? *Learning and Individual Differences*, 20(1), 14-18.
- Sun, B., Lai, S., Xu, C., Xiao, R., Wei, Y., & Xiao, Y. (2017). Differences of online learning behaviors and eye-movement between students having different personality traits. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education (MIE 2017)* (pp. 71-75).
- Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *Internet and Higher Education*, 27, 44-53.

Promoting Reflection on Question Decomposition in Web-based Investigative Learning

Yoshiki SATO ^{a*}, Akihiro KASHIHARA ^a, Shinobu HASEGAWA ^b, Koichi OTA ^b,
Ryo TAKAOKA ^c

^a *The University of Electro-Communications, Japan*

^b *Japan Advanced Institute of Science and Technology, Japan*

^c *Yamaguchi University, Japan*

* yoshiki.sato@uec.ac.jp

Abstract: In Web-based investigative learning, learners are expected to construct wider and deeper knowledge by navigating a great number of Web resources/pages. In elaborately investigating an initial question, learners are expected to decompose an initial question into related question to be further investigated. However, it is difficult for learners to conduct question decomposition in concurrence with their knowledge construction. In our previous study, we have proposed a model of Web-based investigative learning, and developed the system named interactive Learning Scenario Builder (iLSB for short). Although iLSB could promote self-directed investigative learning, learners often decompose a question into unrelated sub-questions. This suggests the necessity of promoting reflection on question decomposition by diagnosing the appropriateness of question decomposition. Toward this issue, we have proposed a method for diagnosing the appropriateness of question decomposition with Linked Open Data (LOD). In this paper, we describe an adaptive prompting with diagnosed results for reflection on question decomposition. This paper also reports a case study whose results suggest the potential for promoting reflection on improper question decomposition.

Keywords: Web-based investigative learning, Self-regulated learning, Reflection, Diagnosis, Adaptive prompting

1. Introduction

The Web allows learners to investigate any question to learn from a great number of Web resources in a self-directed way [Hill and Hannafin, 1997]. In the Web-based investigative learning process, learners are expected to construct wider and deeper knowledge from their point of view by selecting and navigating Web resources/pages suitable for learning, and integrating the contents learned at the navigated resources/pages [Henze and Nejd, 2001].

On the other hand, learners tend to search a limited number of Web resources/pages for investigating a question, which often results in an insufficient investigation. In elaborately investigating an initial question, learners are expected to deepen and widen the question by identifying related questions to be further investigated during navigation and knowledge construction [Hill and Hannafin, 1997]. This corresponds to decomposing the initial question into related ones as sub-questions, which would give rise to wider and deeper knowledge construction.

In addition, the Web do not provide learners with learning scenarios which implies questions to be investigated and their sequence such as a table of contents in an instructional textbook. In learning with Web resources, therefore, learners are expected to create learning scenarios by decomposing a question into sub-questions. Such learner-created scenarios could play a crucial role in self-regulating their navigation and knowledge construction process [Azevedo and Jennifer, 2004].

However, it is difficult for learners to create learning scenario in concurrence with question decomposition and knowledge construction [Land and Susan M, 2000]. In our previous work, we have proposed a model of Web-based investigative learning and developed a system named interactive Learning Scenario Builder (iLSB for short). iLSB provides scaffolds for conducting the investigative learning process as modeled [Kashihara and Akiyama 2017]. On the other hand, learners often

decompose a question into unrelated sub-questions, which could be caused by insufficient reflection on question decomposition. This suggests the necessity of promoting reflection on question decomposition by diagnosing the appropriateness of question decomposition. But, it is a challenging issue in self-directed learning since we could not define the suitability of relationships between questions in advance.

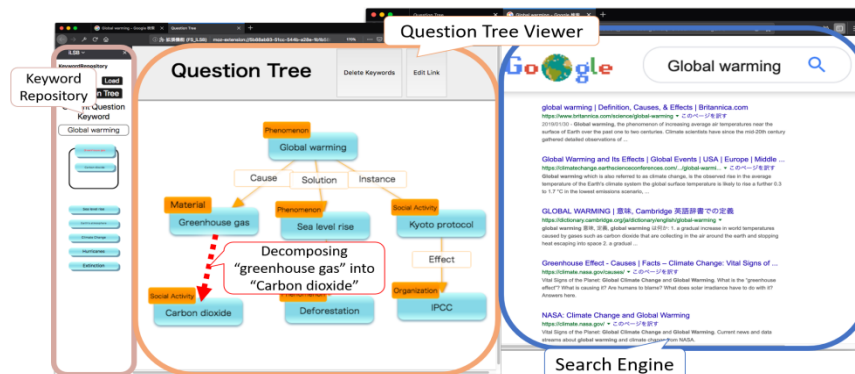


Figure 7: User Interface of iLSB

Toward this issue, we have proposed a method for diagnosing question decomposition with Linked Open Data (LOD) [Sato et al., 2019]. In this paper, we describe adaptive prompting for reflection on question decomposition with diagnosed results, which induces learners to revise their learning scenario. This paper also reports a case study whose purpose was to ascertain whether adaptive prompting could promote reflection on question decomposition and learning scenario revision. The results suggest that it allows learners' reflection on their question decomposition.

2. Web-based Investigative Learning

Let us first describe the model of Web-based investigative learning and iLSB. We then explain the issue of how to prompt learners to reflect on their own question decomposition.

2.1 Model of Web-based Investigative Learning and iLSB

This model includes three cyclic phases: (a) search for Web resources, (b) navigational learning, and (c) question decomposition. In phase (a), learners are expected to select Web resources suitable for investigating an initial question and to navigate across them using a search engine. In phase (b), the learners are expected to navigate the Web pages in the selected resources and to construct their knowledge by extracting keywords to represent the contents learned in the pages and by making the relationships among them. In phase (c), the learners are expected to find out some related sub-questions to be further investigated about the initial question, which could be obtained from the keywords extracted in phase (b). This corresponds to decomposing the initial question into sub-questions. Each sub-question is investigated cyclically in the next phases (a) and (b).

The question decomposition results in a tree called question tree, in which questions as nodes are represented with keywords (called q-keywords). This tree includes part-of relationships between the question and the sub-questions whose root represents the initial question. The question tree represents a learning scenario.

In addition, we have developed iLSB as an add-on for Firefox, which scaffolds the investigative learning process as modeled. Figure 1 shows the user interface of iLSB. iLSB provides learners with functions according to the three cyclic phases: search engine, keyword repository, and question tree viewer. Learners are expected to construct knowledge by means of keyword repository, and to decompose a question into sub-questions by means of question tree viewer.

2.2 Issue and Purpose

In Web-based investigative learning, it is important for learners to reflect on their investigative learning process. On the other hand, learners often decompose into sub-questions unrelated to the initial question even if they use iLSB, which is caused by insufficient reflection on question decomposition. A common approach to promote learners' reflection is to prompt them to review their own activities. Although such prompting could stimulate their metacognitive activities [Bannert, M& Reimann. P, 2012], it is often conducted in an ad hoc way [Narciss 07].

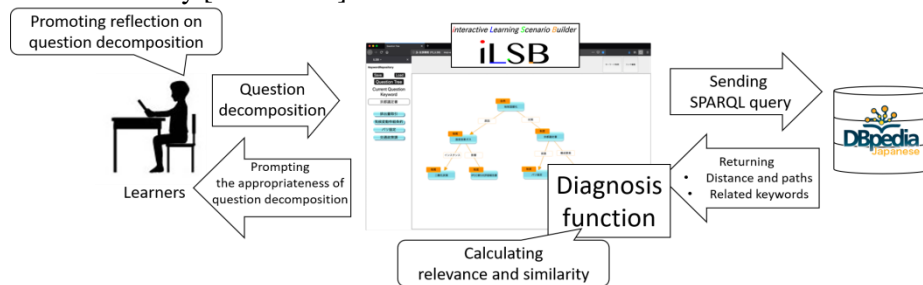


Figure 8: Framework of Question Decomposition Diagnosis

The main issue addressed in this paper is how to promote reflection on question decomposition in an adaptive way. Our approach is to diagnose the question decomposition conducted by learners, and to prompt learners to reflect on it according to diagnosed results. This allows learners to create an appropriate learning scenario. However, it is difficult to verify whether the relationships between a question and the sub-questions decomposed during self-directed investigation are appropriate, since we could not prepare a valid relationship with any question in advance. The adaptive prompting can be accordingly viewed as a challenging issue. We have proposed the diagnosis method with LOD so far [Sato et. al., 2019]. In this paper, we aim to confirm whether presenting diagnosed results could promote learners to reflect on their question decomposition.

3. Diagnosis of Question Decomposition

Let us next demonstrate the method for diagnosing and prompting the appropriateness of question decomposition with LOD.

3.1 LOD (Linked Open Data)

LOD (Linked Open Data) is a set of structured data interlinking with related ones on the Web. In this work, we use DBpedia Japanese whose data are extracted from Japanese Wikipedia [DBpedia Japanese, 2016]. The data in DBpedia Japanese are expressed as RDF (Resource Description Framework), which consists of three entities known as triples: subject, predicate, and object. Such RDF data are extracted/operated by sending SPARQL queries to DBpedia Japanese.

3.2 Framework of diagnosis

Figure 2 shows the framework for diagnosing the appropriateness of question decomposition by means of DBpedia Japanese. The diagnosis is implemented as a function of iLSB.

In this framework, learners are first expected to investigate an initial question according to the model of Web-based investigative learning, and to decompose a question into sub-questions. iLSB then sends SPARQL queries to DBpedia Japanese, and obtain paths between q-keywords and keywords related to each q-keyword. From obtained paths and related keywords, iLSB then calculates the relevance and similarity between q-keywords. Following calculated relevance and similarity, iLSB finally decides the appropriateness of question decomposition as one of three levels: appropriate, weak appropriate and unknown. The diagnosed results are presented to learners, which is expected to prompt them to reflect on the appropriateness of their question decomposition.

3.3 Diagnosis Procedure

We have designed the procedure for diagnosing the appropriateness of decomposition into any q-keyword i in a question tree as shown in Figure 3 [Sato et. al., 2019]. This procedure decides the appropriateness level of question decomposition by calculating the relevance and similarity between q-keywords. The relevance is calculated as three levels: relevant, weak relevant, and unknown. The similarity is also calculated as three levels: similar, weak similar, and unknown.

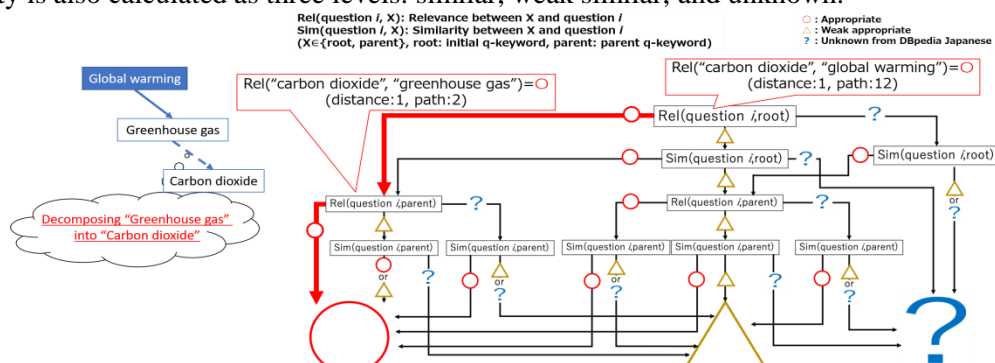


Figure 9: Diagnosis Procedure

This procedure first calculates the relevance and similarity between q-keyword i and the root q-keyword. It then calculates the ones between q-keyword i and the parent q-keyword. Depending on the calculated levels of relevance and similarity, as shown in Figure 3, it finally decides the appropriateness level as one of the three levels.

In calculating the relevance between q-keywords, iLSB first sends a SPARQL query to DBpedia Japanese, and obtain paths between q-keywords in DBpedia Japanese. Depending on the distance and the number of paths, the relevance is determined based on thresholds. In case the distance is 1, the relevance level is suggested as relevant. In case the distance is 2 and the number of paths is more than 30, the relevance level is suggested as weak relevant. In case the distance is more than 3 or the number of paths is less than 30, the relevance level is suggested as unknown.

In calculating the similarity between q-keywords, iLSB first sends the SPARQL query to DBpedia Japanese in order to obtain keywords related to each q-keyword. iLSB then creates two sets each of which consists of words to be extracted from the obtained keywords by means of morphological analysis. iLSB finally calculates the overlap coefficient which indicates the similarity between the two sets. The similarity level is determined based on thresholds depending on the calculated overlap coefficient. In case the overlap coefficient is more than 0.3, the similarity of the keywords is suggested as similar. In case the overlap coefficient is between 0.1 and 0.3, it is suggested as weak similar. In case the overlap coefficient is less than 0.1, it is suggested as unknown.

In our previous work, we had a case study for evaluating the validity of the designed procedure shown in Figure 3 [Sato et. al., 2019]. In this study, we compared the appropriateness of question decomposition diagnosed with this procedure and with the one diagnosed manually. The results suggest that the accuracy of diagnosis with this procedure toward manual diagnosis was 77.8%, which seems high.

3.4 Prompts with Diagnosed Results

iLSB calculates the appropriateness of question decomposition with the designed diagnosis procedure whenever learners decompose a question into sub-questions. After their question decomposition, iLSB provides prompts on their demand in the question tree viewer, which include the appropriate levels diagnosed. The learners are expected to reflect on the decomposed questions particularly diagnosed as weak appropriate and unknown with the prompts. They are then expected to re-investigate the question and sub-questions to revise their question tree.

4. Case Study

4.1 Purpose and Procedure

We have conducted a case study whose purpose was to ascertain whether prompts with diagnosed results could promote reflection on question decomposition.

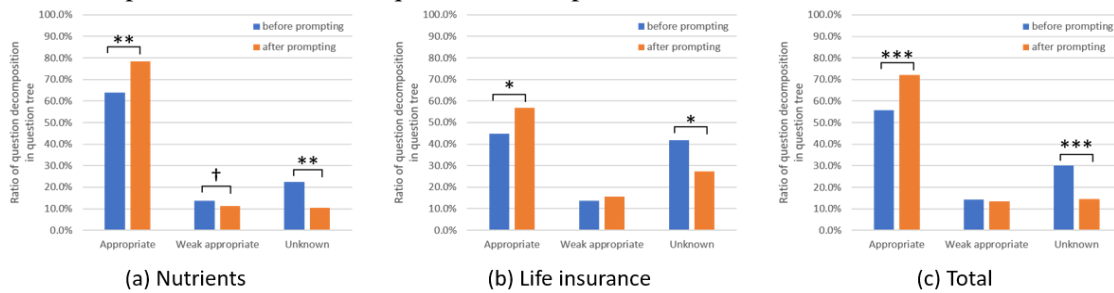


Figure 10: Ratios of Diagnosed Results

Participants were 16 graduate and undergraduate students in science and technology. Half of the participants were given “What is nutrients?” and the others were given “What is life insurance?” as an initial question. The participants were first required to use iLSB to investigate the initial question without diagnosis for 30 to 60 minutes. After their investigation, iLSB diagnosed each question decomposition in the question trees the participants created, and presented it as prompts on the question tree viewer. If necessary, they could re-investigate to revise their question trees by means of iLSB by means of diagnosis within 30 minutes. After the tree revision, the participants were finally required to answer a five-point scale questionnaire for assessing the effectiveness of prompts. In order to ascertain whether prompts could promote reflection on question decomposition, we set the following hypothesis:

Hypothesis: The ratio of question decomposition diagnosed as appropriate increases, and the one diagnosed as unknown decreases.

4.2 Results

Figure 4 shows the ratios of diagnosed results in question trees obtained from each initial question and the total ratios of diagnosed results obtained from the two initial questions. In order to ascertain whether these results support the hypothesis, we compared the ratios of diagnosed results before prompting and the ones after prompting. Regarding each initial question in Figure 4, the question decomposition diagnosed as appropriate significantly increased (Nutrients: $t(15)=-3.99$, $p<.01$, Life insurance: $t(15)=-2.21$, $p<.05$) as the results of one-sided t-test. The one diagnosed as unknown also significantly decreased (Nutrients: $t(15)=3.89$, $p<.01$, Life insurance: $t(15)=2.68$, $p<.05$). These results support the hypothesis. As for the total ratios, the question decomposition diagnosed as appropriate significantly increased ($t(15)=-3.90$, $p<.001$), and the one diagnosed as unknown significantly decreased ($t(15)=4.52$, $p<.001$).

Table 1 shows the results of the questionnaire. All average points obtained were over 3.5 which seems high. As for Q1, Q2 and Q4 asking about the effectiveness of reflection with prompts, the average points were over 4.0 that was higher than the others, and the variances were all under 1.00. As for Q5 asking about their self-directed investigation and Q3 asking about how easy it was to understand prompts, in addition, the average points were over 3.6 which seems high. On the other hand, the variances were 1.35, which seems large in answer among the participants.

4.3 Discussion

Overall, the results of this case study suggest that iLSB could adaptively promote reflection on question decomposition on the Web, which is ill-structured and difficult to scaffold for learners. Let us first discuss the effectiveness of prompts for reflection. The results shown in Figure 4 indicate a significant increase in question decomposition diagnosed as appropriate and a significant decrease in the one diagnosed as unknown by means of prompts. This suggests prompts could promote reflection on question decomposition, which is also supported by the results of Q1, Q2 and Q4 in Table 1.

In addition, the results of Q3 in Table 1 suggest that it is easy to understand prompts presented as the three levels. We accordingly asked the participants who selected “1” or “2” the reason why they had difficulty in understanding. They answered that they could not know why iLSB diagnosed question decomposition as unknown. This suggests the necessity for providing some evidence for prompts.

Table 1: Averages and Variances of Questionnaire Results

Question number	Question with five-point scale	Average	Variance
Q1	How helpful were prompts for reflection? (5: Very helpful 1: Not at all)	4.0	0.98
Q2	How effectively prompts improved question tree? (5: Very effectively..... 1: Not at all)	4.1	0.86
Q3	Was it easy to understand prompts indicating three levels? (5: Very easy to understand .. 1: Not easy to understand at all)	3.6	1.35
Q4	Did prompts allow you to reflect on the contents learned? (5: Yes, they did. 1: No, they didn't at all.)	4.0	0.69
Q5	How much did you feel that prompts prevented you from self-directed investigative learning? (5: Not at all 1: Very prevented)	3.8	1.35

As for self-directedness with prompts, the average of Q5 in Table 1 suggests that prompts could not prevent learners from self-directed investigative learning, but it seems to depend on learners since the variance was large. This suggests that learners who have more question decomposition diagnosed as unknown tend to feel more restriction on their self-directed investigative learning process.

5. Conclusion

This paper has addressed the issue how to promote reflection on question decomposition in an adaptive way. Toward this issue, we have proposed an adaptive prompting that includes the results of diagnosing the appropriateness of question decomposition with LOD.

This paper has also reported a case study for ascertaining whether the adaptive prompting could be effective for learners to reflect on question decomposition. The results suggest that prompts could promote reflection on the question decomposition. On the other hand, we found out the necessity of providing learners with some evidence for prompts presented, which is one of our future work.

Acknowledgements

The work was supported in part by JSPS KAKENHI Grant Number 17H01992.

References

- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of educational psychology*, 96(3), 523.
- DBpedia Japanese (2016), <http://ja.dbpedia.org/>
- Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational technology research and development*, 45(4), 37-64.
- Henze, N. and Nejdil, W. (2001). Adaptation in Open Corpus Hypermedia. *International J. of Artificial Intelligence in Education*, 12(4), 325-350
- Kashihara, A., & Akiyama, N. (2016). Learning Scenario Creation for Promoting Investigative Learning on the Web. *The Journal of Information and Systems in Education*, 15(1), 62-72.
- Land, Susan M (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, 48(3), 61-78.
- Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, 23, 1126-1144.
- Sato, Y, et al. (2019). Diagnosis with Linked Open Data for Question Decomposition in Web-based Investigative Learning. *Foundations and Trends in Smart Learning*, 103-112.

Modeling Video Viewing Styles with Probabilistic Mode Switching

Hiroaki KAWASHIMA^{a,b,*}, Kousuke UEKI^b & Kei SHIMONISHI^b

^a*School of Social Information Science, University of Hyogo, Japan*

^b*Graduate School of Informatics, Kyoto University, Japan*

*kawashima@sis.u-hyogo.ac.jp

Abstract: During video lectures, learners may have attentional modes such as “follow a lecturer's guide (speech and pointers),” “look ahead of spoken parts and actively check slide content,” and “roughly browse a slide.” The dynamic change of these modes is useful to characterize personal and/or temporary viewing styles. This paper presents a method to analyze video viewing styles through gaze behavioral data by using a probabilistic generative model with a latent mode variable. In our experiments, we show that the model can infer viewers' temporal mode patterns and successfully characterize task-dependent viewing situations.

Keywords: Eye-tracking, Video lectures, Viewing styles, Probabilistic gaze-behavior model

1. Introduction

While clickstream logs on video lectures (e.g., MOOCs) are promising sources to analyze learning behavior (Kim et al., 2014), those logs mainly contain users' intentional activities, such as page jumps, and it cannot be directly used to understand detailed learners' reaction to video content (e.g., slide information, lecture's speech, and pointing actions). Gaze data in video lectures, on the other hand, enable us to analyze learners' behavior deeply enough to infer which consisting regions attract/confuse viewers (Nguyen & Liu, 2016) and how much viewers followed a lecturer's speech (Sharma, Jermann, & Dillenbourg, 2014).

To understanding learners' performance and skills, distinctive spatio-temporal gaze patterns is often analyzed in gaze studies (Mangaroska, Sharma, Giannakos, Tr tteberg, & Dillenbourg, 2018). This gives us insight that such gaze patterns partially reveal viewers' internal process, and in other words, there may exist useful “intermediate representation” between gaze patterns and human cognitive process. In Kawashima, Ueki, & Shimonishi (2019), a minimal model for viewers' attentional modes during video lectures, such as “roughly grasp slide content,” “actively follow slide content,” and “follow a lecturer's guide (speech and pointers)” is introduced. While these modes are not necessarily directly related to human attentional modes, inferred modes can be treated as useful intermediate representation to characterize viewers' situations. In this paper, we extend the work by introducing probabilistic mode switching model to analyze and visualize learners' viewing styles in video lectures.

2. Multi-mode Gaze Model

The model proposed in Kawashima et al. (2019) is a multi-mode generative model, where a viewer's gaze sequence is assumed to be affected by the three component submodels. We first briefly describe the model in this section and introduce its extension in Section 3 to model a variety of viewing styles.

2.1 Model

Assume that raw gaze data are converted into a sequence of areas-of-interest (AOIs), which we refer to as *gaze regions*, based on the velocity of eye movements (Salvucci & Goldberg, 2000). Suppose that $r_1, \dots, r_{k-1}, r_k, \dots, r_K$ is a gaze sequence of a viewer, where $r_k \in \{R_1, \dots, R_N\}$ is a region ID at time step k , and R_1, \dots, R_N are region IDs in a slide. Note that *time* (or *time step*) k denotes an ordered number

of AOI switches. Meanwhile, we use t_k to describe actual *media time* (physical time whose origin is the start time of a slide) in the video at time step k .

Mode 0 (base distribution): Attracted by salient regions such as high contrast or important terms, modeled by region distribution $P(r_k = R_i | m_k = 0) = a_i$,

Mode 1 (region order): Follow slide content by considering the meaning of content information, modeled by transitional probabilities $P(r_k = R_j | r_{k-1} = R_i, m_k = 1) = b_{ij}$,

Mode 2 (lecture's guide): Follow a lecturer's guide such as spoken words and pointers, modeled by time-dependent region distribution $P(r_k = R_i | t_k = t, m_k = 2) = c_{it}$.

Then, the influence ratio of the above submodels (modes) is assumed to change dynamically:

$$P(r_k | \cdot) = P(m_k = 0 | \cdot)P(r_k | m_k = 0) + P(m_k = 1 | \cdot)P(r_k | r_{k-1}, m_k = 1) + P(m_k = 2 | \cdot)P(r_k | t_k, m_k = 2),$$

where (\cdot) denotes the previous gaze regions r_{k-1}, r_{k-2}, \dots and timestamps t_k, t_{k-1}, \dots

The probability $P(m_k = m)$, $m \in \{0, 1, 2\}$ describes the influence ratio of submodel m , whose sum is $\sum_{m=0}^2 P(m_k = m) = 1$. Based on this model, typical situations can be described as follows: When a viewer follows the lecturer's guide (spoken words and pointers) at time k , the value of $P(m_k = 2)$ becomes higher than $P(m_k = 0)$ and $P(m_k = 1)$. Meanwhile, when a viewer follows the slide content actively by ignoring the lecturer's guide, $P(m_k = 1)$ becomes higher. In a time period of mind wandering, $P(m_k = 0)$ get higher due to not following either the content or the lecturer.

Note that the concept of mode $m = 2$ is closely related to the with-me-ness proposed in Sharma et al. (2014) in terms that the corresponding submodel tries to describe specific spatio-temporal points that a lecturer attracts gaze of learners.

2.2 Model Training

Suppose that a collection of gaze sequences $r_{seq}^{(v)}$ ($v = 1, \dots, n$) from n viewers of the same slide is given in a model training phase. Using this gaze data, the model parameters a_i , b_{ij} , and c_{it} is estimated based on the maximum-likelihood estimation. This estimation is not straightforward due to the hidden variables m_k , whose posterior needs to be estimated simultaneously. That is, this model training can be viewed as clustering of gaze patterns to three modes, and therefore the expectation-maximization (EM) algorithm can be utilized.

The algorithm repeats the E- and M-steps iteratively. In the E-step, the posterior $P(m_k^{(v)} | r_{seq}^{(v)})$ for each time step k is computed using the model with current parameters during iterations. Then, all the parameters are updated in the M-step, where additive smoothing (in our experiments, additional 0.1 count) was used to avoid the zero-frequency problem.

3. Probabilistic Mode-Switching Model

3.1 Inference of Mode Sequences with a Probabilistic Mode Transition

Given a newly observed gaze data r_{seq} , a *mode posterior* $P(m_k = m | r_{seq}) = \gamma_k(m)$ ($m = 0, 1, 2$, $k = 0, 1, 2, \dots, K$) can be obtained with a similar procedure as the E-step in the training phase. The posterior can be considered as the inferred ratio of the influence of each submodel. An inferred sequence of $\gamma_k(m)$ contains useful information of when and which mode the viewer took during the video viewing situations. For example, the value of $\gamma_k(2)$ is expected to be large for time k when the viewer focuses on the lecturer's talk (guide information).

To statistically analyze inferred mode sequences, we do not directly use the patterns of the sequences but utilize the temporal structure behind the patterns. Specifically, we extend the model described in the previous section by introducing mode switching model with the following *mode-transition probability*:

$$P(m_k = q | m_{k-1} = p) = A_{pq},$$

where $p, q \in \{0, 1, 2\}$. Note that this extension corresponds to the model of state transition of hidden Markov models (HMMs), while its states (modes) and emission probabilities are specifically designed to describe viewers' situations as explained in Section 2.1.

3.2 Model Training

Now the extended model has a parameter set $\{a_i, b_{ij}, c_{it}, A_{pq}\}$. Since the goal of this study is to characterize video viewing styles, we introduce an assumption that these parameters can be divided into two: viewer-independent (shared) parameters $\theta_{shared} = \{a_i, b_{ij}, c_{it}\}$ and viewer-dependent parameters $\{A_{pq}^{(v)}\}$. By using the shared parameters, modes and their switching can be analyzed in the common space. Note that we add superscript (v) to explicitly denote that the mode-transition probabilities depend on viewer v .

For better convergence of the model parameters, we also divide the training into two steps. In the first step, viewer-independent parameters θ_{shared} are estimated without mode switching model (introduced in Section 2). Then, viewer-dependent parameters $\{A_{pq}^{(v)}\}$ are estimated for each of viewer v to characterize the structure of viewer v 's mode transition. In this second step, all the viewer-independent parameters are fixed. As for initial mode probabilities $P(m_1)$, we used equal probabilities 0.5 for $m_1 = 0, 2$, and 0 for $m_1 = 1$.

3.3 Comparing Mode-Switching Structures

Once each viewer's mode-switching behavior is encoded by the model above, dissimilarity between two models can be introduced. This dissimilarity can be considered as a pseudo distance between two viewing patterns. Since the extended model is analogous to HMMs, we here use a distance measure proposed in Juang & Rabiner (1985), which utilizes Kullback-Leibler (KL) divergence.

Let $\theta^{(v_1)} = \theta_{shared} \cap \{A_{pq}^{(v_1)}\}$ and $\theta^{(v_2)} = \theta_{shared} \cap \{A_{pq}^{(v_2)}\}$ be the parameter sets of viewer v_1 and v_2 , respectively. Then, the divergence-related value can be computed by using log likelihood:

$$D(v_1 || v_2) = \frac{1}{K^{(v_1)}} \left[\log P(r_{seq}^{(v_1)} | \theta^{(v_1)}) - \log P(r_{seq}^{(v_1)} | \theta^{(v_2)}) \right],$$

where $K^{(v_1)}$ is the length of gaze sequence $r_{seq}^{(v_1)}$. Considering the non-symmetric property of KL divergence, the pseudo distance can be defined as the following average:

$$\text{Dist}(v_1, v_2) = (D(v_1 || v_2) + D(v_2 || v_1))/2.$$

4. Experiments

In this experiment, we verify the proposed model in terms of its capability of characterizing and visualizing video viewing styles. To focus on the evaluation of the proposed model itself, we conducted laboratory experiments with designed settings. Specifically, we prepared not only (a) a normal video-viewing situation but two additional artificial situations: (b) with a sub task and (c) with an edited content (static slides with no sound). Since the ground truth of viewers' internal states cannot be obtained, we considered that these designed situations highly bias their internal states and simulate some extreme situations such as mind wandering (Mills, Bixler, Wang, & D'Mello, 2016; Hutt et al., 2017) or ignoring a lecturer's guide. Self-reporting or think aloud protocol is another option to obtain ground truth of internal process, but we do not take this option to avoid an additional task affecting gaze behavior.

4.1 Experimental Settings

33 university students were recruited to conduct the lab-setting research. Each of the students was explained a summary of our research and signed an informed consent form upon the arrival. The explanation of the research objective was abstract enough to avoid affecting their gaze behavior.

4.2 Tasks and Content

Video-viewing tasks were assigned to each of the 33 participants. They were asked to (i) watch a video on a monitor and (ii) answer several written questions related to the content of the watched video. The post-questions were prepared to make each participant concentrated enough on the video content and to measure the degree they could follow the content. Confidence of each answer were also collected in five-point Likert scale, and the prior-knowledge of each question was also asked to verify that most of the participants did not know the question-related content before the experiment.

As described in the beginning of this section, the participants were divided randomly into the following three groups (11 participants each):

- (a) Group 1 (normal): No additional task was assigned.
- (b) Group 2 (sub task): Additional task of mental calculation (repeatedly subtract 3 from 1000) was assigned from slide 2.
- (c) Group 3 (no guide): No additional task but an edited video (a sequence of static slides with no sound) was displayed from the middle (slide 3-); the length of the presence of each slide was also edited according to the density of the content.

As for the sub task, participants were asked to vocalize each result of the mental calculation, but the vocalized values were not checked. The confidence of the answers to the post-questions was low enough in most of the questions, and the scores of Group 2 were lower than Group 1 for all the questions.

The topic of the content was “inferential statistics” from JMOOC gacco (<https://gacco.org/>), and the length of the video content, consisting of 4 slides, was about 10 minutes. The screen of the video consisted of a slide and a lecturer. The slide was partially and temporarily overlaid by the lecturer's arm and pointers. In this experiment, the 3rd slide was used, which contained only texts.

4.3 Data acquisition

Tobii X120 eye tracker was used to measure participants' gaze points on a screen with sampling rate 60 Hz. We here used this lower sampling rate because we focus on analyzing fixations rather than saccadic eye movements. Each participant was asked to sit in front of the screen where chin rest was used to reduce measurement noise as much as possible. While the used device was robust against head movements to some extent, we decided to use this setting to focus on the verification of the proposed model and algorithm. In the identification step of fixations, lack of data less than 150 ms was considered as instantaneous noise and interpolated using surrounding data.

While automatic region segmentation is possible to define regions, in this experiment, we manually segment regions based on the meaning of words for the sake of avoiding region segmentation errors. Region IDs were basically numbered from top left to bottom right, which roughly coincided with the order of reading the slide content.

4.4 Results

The two-step training described in Section 3 was applied to the acquired gaze sequences on the slide 3: The shared parameters θ_{shared} were estimated by normal-task data in the first step, and mode-transition probabilities were obtained for each of the viewers in Group 1-3 in the second step. During the second step, each viewer's mode posterior sequence was also computed.

Figure 1 shows examples of mode sequences from each of the three tasks (Group 1-3). It can be seen that mode 1 and mode 2 are dominant in (a) normal task and (b) sub task while only mode 1 is dominant in (c) no guide. Compared to Group 1 (normal task), mode 0 appears more frequently in Group 2 (sub task). These trends can be seen consistently in most of other viewers' sequences.

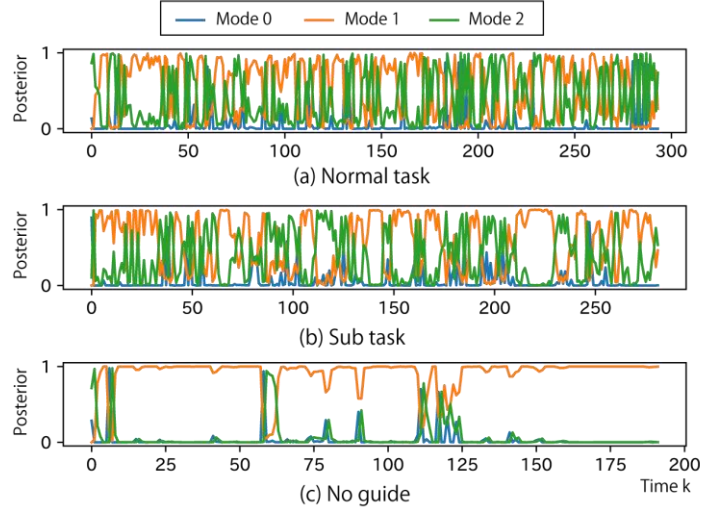


Figure 1. Examples of mode-posterior sequences in the three situations with (a) normal task (Group 1), (b) sub task (Group 2), and (c) no guide (Group 3).

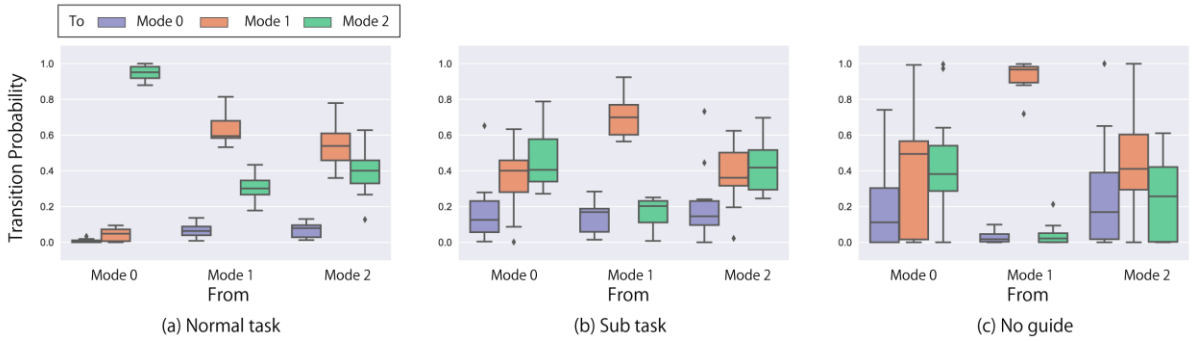


Figure 2. Statistics of mode-transition probabilities estimated through the switching-mode model in the three situations with (a) normal task (Group 1), (b) sub task (Group 2), and (c) no guide (Group 3).

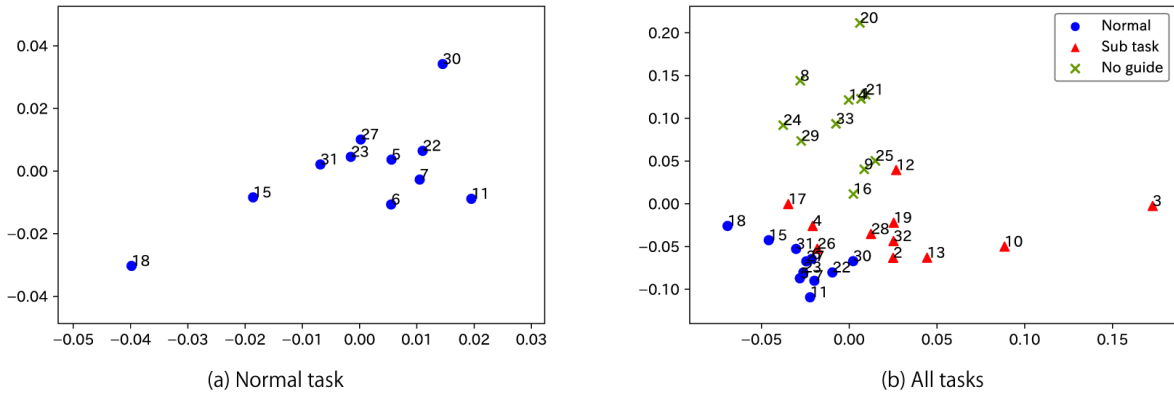


Figure 3. Results of multi-dimensional scaling of dissimilarities of trained models (the numbers correspond to viewers' IDs): (a) dissimilarity structure among normal-task (Group 1) viewers are depicted in 2-d space. (b) dissimilarity structure of all viewers in Group 1-3 are visualized in 2-d space. Colors of marker corresponds to the groups (assigned tasks).

To analyze the difference of viewing styles in Group 1-3, we plot the statistics of parameter distributions of mode-transition probabilities $\{A_{pq}\}$ in Figure 2. In normal-task group (Group 1), there were less transitions to mode 0. In addition, the probability of self-loop of mode 0, which corresponds to the trend of duration length of the mode, also took very small values. This can be interpreted that in normal-task situations, viewers tended to take mode 1 or mode 2 most of the time by actively following

slide content and/or the lecturer's guide compared to sub-task situations. On the other hand, as can be seen in Figure 2 (c), Group 3 (no guide) had high self-loop probability of mode 1, which means they mostly took mode 1 (follows slide content). This is natural since there was no lecturer's guide (e.g., speech).

To demonstrate how the model can be used to discriminate video viewing styles, we visualized the structure of dissimilarities of trained models from each of the viewers. Figure 3 (a) shows the plot of Group 1 structures. Here, multi-dimensional scaling (MDS) was used to visualize the dissimilarity structure in the 2-d Euclidean space. From this figure, we can find which viewer's behavior was different from the others. For example, viewer 18, plotted far from the remaining, had longer duration of mode 0 compared to the others. Figure 3 (b) is also the visualization with MDS but all the viewers' models were used. Here, the difference of the tasks can be clearly observed in this visualization.

5. Conclusion

This paper proposed a method to characterize and visualize video viewing styles during video lectures using a probabilistic mode-switching model. The model consists of three designated submodels to extract and interpret key features in video viewing gaze behaviors, while it has similar structure as HMMs. Through the analysis of mode-transition probabilities and visualization, the proposed model successfully distinguishes the difference of gaze behavior in different tasks. While the assigned tasks were artificially designed in this study, we believe that the proposed technique can be used to find "in which period" each learner has non-typical behavior. Then, this information may help us to design personalized feedback to assist learners by providing information for self-assessment or further study and also to improve learning materials. Another interesting question is how the extracted viewing styles are related to learners' performance, which should be investigated in the future work with larger data.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19H04226, JST PRESTO Grant Number JPMJPR14D1, and the Telecommunications Advancement Foundation.

References

- Juang, B. H. & Rabiner. L. R. (1985). A Probabilistic Distance Measure for Hidden Markov Models. *AT & T Technical Journal*, 64 (2), 391–408.
- Kawashima, H., Ueki, K., & Shimonishi, K. (2019). Model-Based Analysis of Gaze Data During Video Lectures. *3rd Multimodal Learning Analytics Across Spaces (CrossMMLA) Workshop, the Companion Proceedings of LAK19*, 526–533.
- Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller., R. C. (2014). Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos. *International Conference on Learning @ Scale (L@S)*, 31–40.
- Mangaroska, K., Sharma, K., Giannakos, M., Tr tteberg, H. & Dillenbourg, P. (2018). Gaze Insights into Debugging Behavior Using Learner-Centred Analysis. *ACM International Conference on Learning Analytics & Knowledge (LAK)*, 350–359.
- Mills, C., Bixler, R., Wang, X., & D'Mello, S. K. (2016). Automatic Gaze-Based Detection of Mind Wandering during Narrative Film Comprehension. *9th International Conference on Educational Data Mining*, 30–37.
- Hutt, S., Hardey, J., Bixler, R., Stewart, A., Risko, E., & D'Mello, S. K. (2017). Gaze-based Detection of Mind Wandering during Lecture Viewing. *10th International Conference on Educational Data Mining*, 226–231.
- Nguyen, C. & Liu, F. (2016). Gaze-based Notetaking for Learning from Lecture Videos. Paper presented at *CHI Conference on Human Factors in Computing Systems*, 2093–2097.
- Salvucci, D. D. & Goldberg, J. H. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. *Symposium on Eye Tracking Research & Applications (ETRA)*, 71–78.
- Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-me-ness": A Gaze-Measure for Students' Attention in MOOCs. *International Conference of the Learning Sciences (ICLS)*, 1017–1022.

Externalization Support for Hypotheses Creation Process of Discovery Learning in Biology

Takuya WADA^{a*}, Toya OTAGAKI^b & Tomoko KOJIRI^c

^a Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan

^b Keihoku Junior and High School, Toyo's Affiliated School, Japan

^c Faculty of Engineering Science, Kansai University, Japan

*ddb04005@edu.osakafu-u.ac.jp

Abstract: Discovery learning is a learning method for understanding knowledge through repeated process of observation, hypothesis generation and verification. However, some students cannot generate hypotheses or verify generated hypotheses because such activities are usually done implicitly. In order to engage in the activities, to be aware of insufficiency of the activities by grasping the process of them explicitly is effective. This research proposes the discovery learning support system of biology by providing the interfaces for externalizing the processes.

Keywords: Discovery learning, learning support system, externalization, hypotheses generation support, hypotheses verification support, biology

1. Introduction

In recent years, various styles of active learning are introduced in the primary/secondary school education. One of the styles is a fieldwork type class. An example of such class in biology is that students go to ecological gardens such as zoos and botanical gardens to discover the natures of animals or plants before learning about them. Such learning method for discover the knowledge from the experience is called discovery learning. In discovery learning, students observe objects and make hypotheses based on information obtained from observation. In making hypotheses, they need to check the validity of their hypotheses and modify them. In order to discover important natures as hypotheses, they at least need to repeat these processes until derived hypotheses can explain all observation results.

Balim et al. (2009) stated that discovery learning has an effect on enhancing interests of students by grasping phenomena from various aspects. However, current discovery learning held in a fieldwork type class, especially in Japan, emphasizes only on practicing the class in the field, such as outside of the school. Since it is difficult for teachers to handle the activities of all students at once, they are often not able to check the hypotheses that students made and also the processes of how they derived them.

Most of studies supporting fieldwork activities in biology focus on promoting observations by asking questions that can be solved by observation (Hwang et al. 2008, Ohsugi et al. 2016). These studies do not support activities for generating hypotheses from observation. Nishio et al. (2016) proposed a system that supports learners to organize the observation results for making presentations. However, this system does not promote the process of deriving hypotheses. Although these studies could partially support discovery learning, there are almost no studies that promotes activities of generating hypotheses from the observation in the fieldwork type class.

This study aims at proposing a system that encourages students to proceed the processes of discovery learning, such as deriving candidates of hypotheses and checking their validities. One of the reasons that students cannot actively conduct these activities is that they usually do it implicitly and do not intentionally promote them. If they are aware of insufficiency of the activities, they can execute the process of discovery learning actively. For the awareness, students need to explicitly grasp their understanding contents corresponding to results of each activity in discovery learning. This study constructs a system where students can externalize their understanding contents in activity during the discovery learning. It also provides the information which triggers students' activity.

2. Activity in Discovery Learning of Biology and its Supporting Framework

This research focuses on the discovery learning in biology where students try to discover the natures of animals. The objective of such learning is to find the unique characteristics of target animals, such as unique behavior and body parts. “A mane of lions” and “a wash gesture of raccoons” are some of the examples. Therefore, hypotheses are the unique body parts and behaviors that can be grasped from the observation. In order to derive the hypotheses, following four activities are needed.

Step 1 “Observation”: To observe body parts and behaviors of target animals from various viewpoints,

Step 2 “Creating hypotheses”: To detect the remarkable body parts and behaviors from the observation results as hypotheses,

Step 3 “Verifying rationality of hypotheses”: To find the correlation between unique body parts and behaviors,

Step 4 “Verifying uniqueness of hypotheses”: To compare body parts and behaviors with other animals.

Observation is a step where students obtain information for creating hypotheses. Since most of unique characteristics are seen in detailed body parts and behaviors, observing from not only large-scale viewpoints but also detailed viewpoints is desirable. Candidates of unique characteristics are chosen from observation results in the step for creating hypotheses. The body parts of animals are developed to allow them to act in favor of survival, so unique behavior should have corresponding unique body parts that are used to do. Therefore, to check the correlation between detected body parts and behaviors is one of the methods for verifying the validity of the uniqueness of them. In addition, unique body parts and behaviors are those that cannot be seen in other kinds of animals. Therefore, to compare the body parts of behaviors with other animals also promote to verify the uniqueness.

Since these four steps are needed in the discovery learning of the biology, this study proposes the support system that has the functions for supporting externalization of each steps and also for promoting them. This system consists of three interfaces; observation record interface, body and behavior correlation interface and animals comparison interface. Observation records interface focuses on the steps 1 and 2, and allows students to record what they discovered through observation and create hypotheses from them. Body and behavior correlation interface is for executing step 3. It allows students to make relations between observed body parts and behaviors. Animals comparison interface is for step 4. It provides an interface for comparing observation results of two different animals. By using the body and behavior correlation interface and the animal comparison interface, students are able to check whether the body parts or behaviors are unique to the target animal.

3. Prototype System of Discovery Learning Support System

We have implemented the discovery learning support system by HTML and Java Script.

Observation record interface consists of two part. One is the interface for recording body parts shown in Figure 1(a) and the other is the interface for recording behavior shown in Figure 1(b). In body/behavior display area, inputted body parts or behaviors are represented by circles. When the new body parts or behaviors have been discovered, students need to select body parts or behaviors that the newly discovered one is included by body/behavior display area, and input the name of the discovered one in the body/behavior addition area. If students find the attributes of the body parts and behaviors, they can input it from modifier addition area and check it on modifier display area.

Figure 1(c) shows body and behavior correlation interface. In tree structure display area, tree structures of body parts and behaviors are displayed. Blue nodes and directed links shown in the upper half part of the tree structure display area are the tree structure of the body parts, and orange nodes and directed links shown in the lower half part are that of the behaviors. The tree structure of the behaviors is represented in the opposite way to that of the body part, namely the root node of the behaviors is in the bottom while that of the body part is in the top. Purple nodes show modifiers and are connected by links to nodes of body parts or behavior. By selecting a node of the body part and a node of the behavior and push the add correlation button, yellow links can be added and selected nodes are correlated.

Figure 1(d) shows the animals comparison interface. It shows the results of the body and behavior correlation interfaces for two animals. The upper half displays the current target animal. When the name

of the animal to compare is inputted in the animal selection box, the observation results of the selected one is appeared in the lower half of the interface. By clicking difference detect button, common nodes of the two animals are detected and are colored by blue, and other nodes are painted by red. By observing the colors, the uniqueness of the body parts of the behaviors of the current target animal is grasped. When the body part and the behavior of the hypothesis are found as unique, the hypothesis is verified as appropriate. If either of one is found as not unique, other behavior / body part should be found to form the hypothesis.

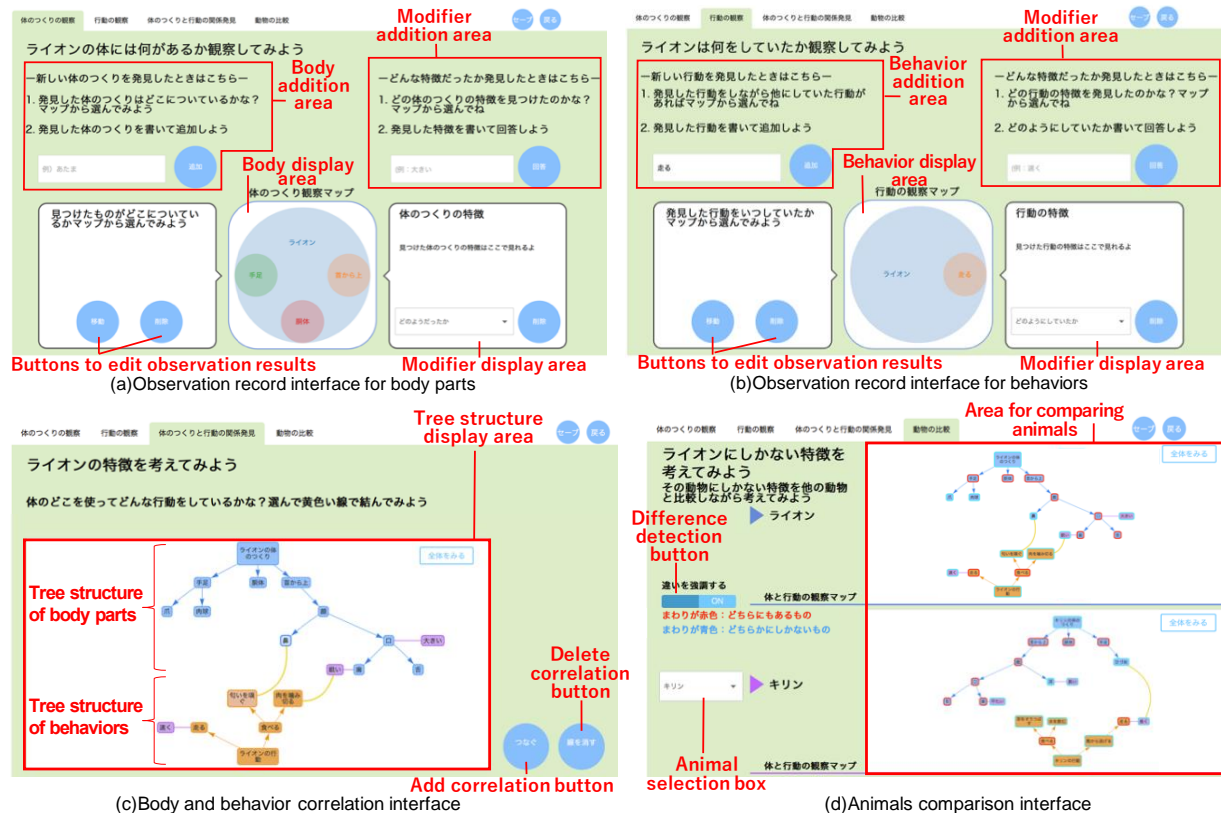


Figure 1. Interfaces of the system

4. Conclusion

This study proposed a system of activating the process of the discovery learning of biology. The system provides the interface for externalizing the process, such as observation, hypotheses creation, and verification. In the future, we need to evaluate the effectiveness of the system through experiment with primary/secondary school students.

Acknowledgements

The work was supported in part by JSPS KAKENHI (No. 18H03346).

References

- Balim, A. G. (2009). The Effects of Discovery Learning on Students' Success and Inquiry Learning Skills. *Eurasian Journal of Educational Research*, 35(1), 1-20.
- Hwang, G. J., Tsai, C. C., & Yang, S. J. H. (2008). Criteria, strategies and research issues of context-aware ubiquitous learning. *Journal of Educational Technology & Society*, 11(2), 81-91.
- Nishio, K., & Kashihara, A. (2016). Retrip: A Learning Environment for Augmenting and Presenting Knowledge from Field Trip. *Procedia Computer Science*, 96(1), 1304-1313.
- Ohsugi, T., Nakanishi, W., Tainaka, M., Inoue, T., Itoh, Y., Iwai, R., Kagawa, K., Matsushita, M., Hori, M., & Ogino, M. (2016). Design of Migration Behavior in the Zoo for Facilitating Autonomic Observation. *IPSJ Journal*, 58(11), 1765-1775 (in Japanese).

Correlating Working Memory Capacity with Learners' Study Behavior in a Web-Based Learning Platform

Alisa LINCKE^{a*}, Daniel FELLMAN^b, Marc JANSEN^{ta,c}, Marcelo MILRAD^a, Elias BERGE^d, Bert JONSSON^b

^a *Department of Computer Science and Media Technology, Linnaeus University, Sweden*

^b *Department of Psychology, Umeå University, Sweden*

^c *University of Applied Sciences Ruhr West, Germany*

^d *Hypocampus AB, Sweden*

*alisa.lincke@lnu.se

Abstract: Cognitive pre-requisites should be taken into consideration when providing personalized and adaptive digital content in web-based learning platforms. In order to achieve this it should be possible to extract these cognitive characteristics based on students' study behavior. Working memory capacity (WMC) is one of the cognitive characteristics that affect students' performance and their academic achievements. However, traditional approaches to measuring WMC are cognitively demanding and time consuming. In order to simplify these measures, Chang et al. (2015) proposed an approach that can automatically identify students' WMC based on their study behavior patterns. The intriguing question is then whether there are study behavior characteristics that correspond to the students' WMC? This work explores to what extent it is possible to map individual WMC data onto individual patterns of learning by correlating working memory capacity with learners' study behavior in an adaptive web-based learning system. Several machine learning models together with a rich context model have been applied to identify the most relevant study behavior characteristics and to predict students' WMC. The evaluation was performed based on data collected from 122 students during a period of 2 years using a web-based learning platform. The initial results show that there is no linear correlation with learners' study behavior and their WMC.

Keywords: working memory capacity, learner's study behavior, personalized learning, machine learning

1. Introduction

A number of learning systems have been developed aiming to support personalized and adaptive learning based on learners' cognitive characteristics (Chang et al., 2015; Graf, 2010; Van Merriënboer & Ayres, 2005). Two of those key cognitive characteristics are working memory and executive functions associated with working memory. Working memory has a limited capacity and refers to the ability to store and manipulate information simultaneously (Baddelev, 2012). Traditionally, WMC has been measured by complex working memory tasks such as the operation span, reading span, and counting span (Unsworth, et al., 2005). What these tasks have in common is that one must keep in mind the sequences of unrelated items (i.e., the storage component) while subsequently performing an intervening task (i.e., the processing component). However, in online learning platforms, the possibility of measuring WMC for all students using cognitive demanding and time-consuming tasks is limited. Mapping individual WMC onto individual patterns of learning is a difficult task and requires expert knowledge from different disciplines. Together with researchers and experts in Cognitive Science, Medical Science and Computer Science, we investigated to what extent it is possible to map individual WMC data onto individual patterns of learning by correlating working memory capacity with learners' study behavior in a web-based learning system called Hypocampus¹.

¹ <https://www.hypocampus.se/>

The rest of the paper is organized as follows: Section 2 provides a description of related work in this field while section 3 describes the approach we are using to predict WMC based on students' study behavior data. Section 4 briefly described extraction of user study behavior characteristics. Section 5 presents the evaluation results of our experiments together with conclusions.

2. Related Work

A search of the literature revealed few studies which have attempted to automatically measure student's WMC based on students' study behavior (Chang et al., 2005; Qinghong et al., 2014). Chang et al. (2005) suggest an approach to calculate students' WMC based on students' navigational behavior patterns and learning style. Another researcher (Qinghong et al., 2014) integrated test questions into educational systems in order to measure student's cognitive level. Based on these test questions, the system provides recommendations of different learning resources with a level of difficulty appropriate to the student's WMC. The drawback of this approach, however, is that students are required to perform additional questionnaires to their study program, which might be time consuming and distract them from their original learning activities. Overall, these studies indicate a great potential of identifying students' WMC based on learners' study behavior data (such as behavior patterns, browsing patterns, game tasks). However, it is still not clear how previous researchers modeled users' "study behavior" and which study behavior characteristics help to identify students' WMC. The rationale for the present study is two-folded. First, the evidence for working memory and WMC being predictive of school performance is extensive (de Smedt et al., 2009; Alloway & Alloway, 2010). Secondly, we can notice an increasing amount of efforts to introduce online learning platforms to support different educational processes. Students using these online platforms generate a massive amount of data. Big data tools and artificial intelligence techniques provide new opportunities to measure learners' cognitive characteristics. Thus, there is an opportunity to further investigate how students WMC can be predicted automatically in online learning systems.

3. Our Approach for Predicting WMC

In order to measure WMC, we asked 122 students to perform an operation span to measure students' WMC. These WMC values are used as the baseline for training and validating the proposed models. When students use the Hypocampus platform during their studies, their user data is stored (named as "logfiles" in Figure 1) in a systematic way. In order to predict individual differences in WMC (shown as "Test results" on Figure 1) as a function of study behavior (extracted from "Log files" as shown on Figure 1), we applied several machine learning models such as multiple linear regression, logistic regression, artificial intelligent tools (AI) and a rich context model (Sotsenko et al., 2016) . Lastly, we used these trained models for predicting students' WMC value as shown in Figure 1.

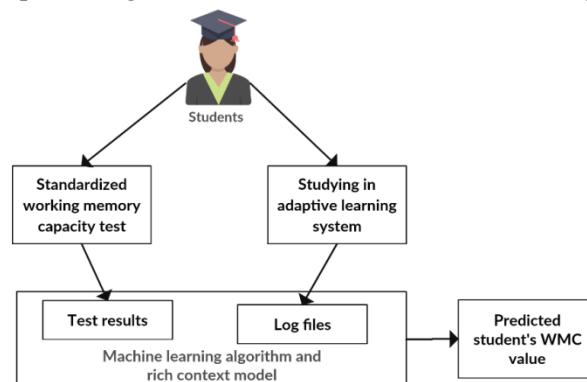


Figure 11. The general idea of our approach.

4. Extracting User Study Characteristics

In order to describe students' study behavior and select relevant characteristics for estimating their WMC we first decide to (a) *identify and use sequential study behavior pattern* in the analyses; afterwards, we (b) *excluded the repetitions* on a basis of having only first attempt answers; finally, we (c) *analyzed the repetitions* in order to find "remembering/forgetting" characteristic. In total 64 user study behavior characteristics were identified and used to correlate with student's individual WMC values.

5. Results and Conclusions

We applied four approaches to predict the WMC value based on the learners' study behavior data (64 user study behavior characteristics): a *multiple linear regression* (MLR), a *logistic regression* (LOR), a *neural network* (NN) *regression*, and a *rich context model* (RCM) (Sotsenko et al., 2017). The validation was performed on dataset from 122 students using a 5-fold cross-validation approach. We used the root mean square error (RMSE) to evaluate the models (as shown in Table 1).

Table 1.

RMSE values for sequential study behavior characteristics.

Algorithm	RMSE
Multiple Linear Regression	20
Logistic Regression	21
Neural Network Regression	18
RCM	19

Overall all models performed with similar RMSE in range between 18-21. These results show that more relevant user study characteristics should be added/found in order to improve the results. Additionally, considering the size of our sample data (N=122), we suggest that it should be further tested and validated with larger datasets.

References

- Alloway, T.P., Alloway, R.G.: Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of experimental child psychology* 106(1), 20-29 (2010).
- Chang, T.W., Kurcz, J., El-Bishouty, M.M., Kinshuk, Graf, S.: Adaptive and personalized learning based on students' cognitive characteristics. *Ubiquitous Learning Environments and Technologies*, 77-97 (2015).
- De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., Ghesquière, P.: Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology* 103(2), 186-201 (2009).
- Graf, S.: Using cognitive traits for improving the detection of learning styles. In: *Database and Expert Systems Applications Workshop - DEXA*, 74-78 (2010).
- Qinghong, Y., Dule, Y., Junyu, Z.: The research of personalized learning system based on learner interests and cognitive level. In: *9th International Conference on Computer Science and Education - ICCSE*, pp. 522-526. (2014).
- Sotsenko, A., Jansen, M., Milrad, M., Rana, J.: Using a rich context model for real-time big data analytics in twitter. In: *Future Internet of Things and Cloud Workshops IEEE*, 228-233 (2016).
- Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W.: An automated version of the operation span task. *Behavior research methods* 37(3), 498-505 (2005).
- Van Merriënboer, J.J., Ayres, P.: Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development* 53(3), 5-13 (2005).

Simulatable Open Learner Models of Core Competencies for Setting Goals for Course Performance

Chih-Yueh CHOU^{a*}, Wen-Chieh CHIH^a, Shu-Fen TSENG^b & Zhi-Hong CHEN^c

^a*Department of Computer Science and Engineering, Yuan Ze University, Taiwan, R.O.C.*

^b*Department of Information Management, Yuan Ze University, Taiwan, R.O.C.*

^c*Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan, R.O.C.*

*cychou@saturn.yzu.edu.tw

Abstract: A competency-based curriculum involves courses for cultivating core competencies required for specific professions to enable students to take courses to cultivate their core competencies. This paper presents a curriculum level and competency-based learning analytics dashboard system with simulatable open learner models (OLMs) of core competencies for assisting students in setting goals for course performance. At first, the system provides students with their OLMs of core competencies based on their taken courses and grades. After that, students are asked to set their goals for course grades of all their courses during the current semester. The system displays the simulation of students' future OLMs if they achieve their goals for course grades this semester. Students can re-set their goals and conduct the simulation of OLMs again until they satisfy the simulation results. The system also enables students to set detailed goals for attendance, assignment hand-in rate, midterm exam, and final exam in order to achieve their goals for course grade. A preliminary evaluation was conducted. Most students agreed that the simulatable OLMs assisted them in understanding the influence of their goals for course grades on core competencies and in setting goals for course grades. Most students also stated that setting detailed course goals prompted them to achieve the goal for course grade.

Keywords: Open learner models, core competencies, goal setting, competency-based curriculum, learning analytics dashboard system

1. Introduction

Higher education has increasingly adopted competency-based curricula to cultivate students' core competencies required for specific professions (Burke, 1989). A competency-based curriculum involves an outcome-based approach for defining students' outcome of specific core competencies, designing and implementing a curriculum for cultivating these core competencies, evaluating students' outcome, and reflecting on and regulating the curriculum, faculty teaching, and student learning. Researchers developed a curriculum level and competency-based learning analytics dashboard system, named VACC (Visualized Analysis of Core Competencies), to build and display students' open learner models (OLMs) of core competencies to assist students in reflecting on their core competencies and setting goals regarding taking additional courses (Chou et al. 2017). OLMs are inspectable, co-operative, editable, or negotiated visualized representations of system diagnosis of student learning (Bull, 2004) and are designed for promoting students' metacognitive processes, such as awareness, reflection, and self-assessment (Bull & Kay, 2013; Chou et al. 2015; Mitrovic & Martin, 2007). This paper presents a function of simulatable OLMs added in VACC for assisting students in setting goals for course performance.

2. Simulatable OLMs of Core Competencies for Setting Goals for Course Performance

A function of simulatable OLMs of core competencies was added in VACC for assisting students in setting goals for course performance. When students log in to VACC and choose the function of simulatable OLMs of core competencies, the system lists the courses they have taken during the current semester and asks them to set their goals for course grades. For example, Figure 1 shows that a student took the course CS554 and set 90 as his/her goal for course grade. After setting goals for course grades of all taken courses, students can push the button of “simulation” to inspect the simulation of future OLMs if they achieve their goals for course grades this semester.



Figure 1. Setting goals for course grade and simulation of future OLM

Figures 2 illustrate a simulation of OLM of quality of a student. The OLM of quality displays the student’s grade point average (GPA) of courses relevant to each core competency. Blue area indicates the current OLM and orange line indicates the simulation of future OLM if the student achieves his/her goals for course grades in all taken courses this semester. The simulation shows that the student’s future OLM of quality will slightly decrease in six core competencies. If the student does not satisfy the simulation result, the student can re-set his/her goals for course grades and conduct the simulation again. Figure 3 illustrates a simulation of OLM of ranking of quality. The ranking is calculated by comparing the student to all graduates, ranging from 0 (the best) to 1 (the worst). The simulation shows that the student’s future OLM of ranking of quality will increase in two core competencies but decrease in five core competencies. The result indicates that graduates averagely performed better than the student’s goals for course grades on the courses related to the five core competencies. The student can increase his/her goals for course grades on the courses related to the five core competencies and conduct the simulation again.



Figure 2. Simulation of OLM of quality

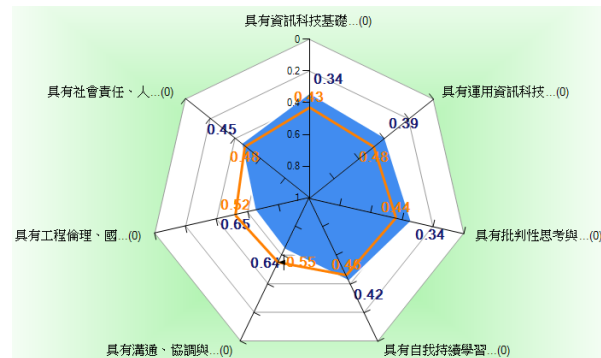


Figure 3. Simulation of OLM of ranking of quality

After setting goals for course grades, students are asked to set detailed course goals to achieve the grade goals for the taken courses. Setting detailed goals for courses includes goals for attendance, assignment hand-in rate, midterm exam, and final exam. For example, Figure 4 shows that the student set 95% as his/her goal for attendance, 100% as his/her goal for assignment hand-in rate, 80 as his/her goal for midterm exam, and 80 as his/her goal for final exam to achieve his/her goal for grade that was set as 90. Detailed course goals are used to prompt students to study to achieve their goals for course grade.

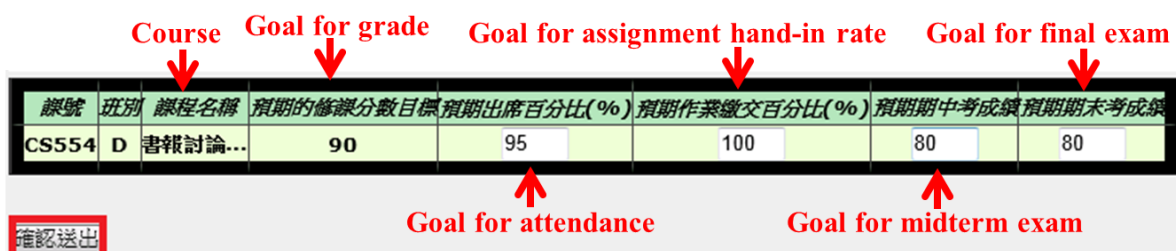


Figure 4. Setting detailed goals for a course

3. Evaluation

An evaluation was conducted to evaluate the simulatable OLMs. The participants were 77 undergraduate students. They were asked to log in to VACC to inspect their OLMs of core competencies, set their grade goals for their taken courses, and conduct the simulation of future OLMs. Then they set detailed course goals to achieve their goals for course grades. At last, they were asked to fill out a questionnaire with five 5-Likert scale items. Table 1 lists the questionnaire results. Most students agreed (strongly agree and agree) that the simulatable OLMs assisted them in understanding the influence of their goals for course grades on core competencies (75%, item #1) and in setting goals for course grades (70%, item #2). Most students (72%) expressed that they will re-set their goals for course grades if the simulation result of future OLMs is not good (item #3). Most students (78%) stated that they set their goals for course grades after much deliberation (item #4). Most students (70%) agreed that setting detailed course goals prompted them to achieve the goal for course grade (item #5).

Table 1. *Results of questionnaire*

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
#1 The simulatable OLMs assisted me in understanding the influence of my goals for course grades on core competencies.	15 (19%)	43 (56%)	17 (22%)	1 (1%)	1 (1%)
#2 The simulatable OLMs assisted me in setting goals for course grades.	15 (19%)	39 (51%)	19 (25%)	2 (3%)	2 (3%)
#3 I will re-set goals for course grades if the simulation of future OLMs is not good.	14 (18%)	41 (53%)	18 (23%)	3 (4%)	1 (1%)
#4 I set my goals for course grades after much deliberation.	17 (22%)	43 (56%)	14 (18%)	2 (3%)	1 (1%)
#5 Setting detailed course goals prompted me to achieve the goal for course grade.	18 (23%)	36 (47%)	20 (26%)	1 (1%)	2 (3%)

4. Summary

This paper presents a learning analytics dashboard system with simulatable OLMs of core competencies for assisting students in setting goals for course grades. The system enables students to inspect their OLMs of core competencies, set goals of course grades in their taken courses, and view the simulation of future OLMs of core competencies to understand the influence of their goals for course grades on core competencies. Students can re-set their goals for course grades and conduct the simulation of future OLMs until they satisfy the simulation result. The system also enables students to set detailed goals in order to achieve their goals for course grade. The results of questionnaire showed that most students expressed that the simulatable OLMs assisted them in setting goals for course grades.

References

- Bull, S. (2004). Supporting Learning with Open Learner Models. In *Proceedings of 4 th Hellenic Conference in Information and Communication Technologies in Education*, 1-15.
- Bull, S., & Kay, J. (2013). Open learner models as drivers for metacognitive processes. In *International handbook of metacognition and learning technologies* (pp. 349-365). Springer, New York, NY.
- Burke, J. W. (Ed.). (1989). *Competency based education and training*. Psychology Press.
- Chou, C. Y., Lai, K. R., Chao, P. Y., Lan, C. H., & Chen, T. H. (2015). Negotiation based adaptive learning sequences: Combining adaptivity and adaptability. *Computers & Education*, 88, 215-226.
- Chou, C. Y., Tseng, S. F., Chih, W. C., Chen, Z. H., Chao, P. Y., Lai, K. R., Chan, C. L., Yu, L. C. & Lin, Y. L. (2017). Open Student Models of Core Competencies at the Curriculum Level: Using Learning Analytics for Student Reflection. *IEEE Transactions on Emerging Topics in Computing*, 5(1), 32-44.
- Mitrovic, A., & Martin, B. (2007). Evaluating the effect of open student models on self-assessment. *International Journal of Artificial Intelligence in Education*, 17(2), 121-144.

Investigating Functional Fixedness among Novice Student Programmers

May Marie P. TALANDRON-FELIPE* & Kent Levi A. BONIFACIO*

Central Mindanao University, Philippines

*mmptalandron@cmu.edu.ph, klab@cmu.edu.ph

Abstract: Functional Fixedness (FF) refers to a type of fixation which hinders an individual to use an object in an unfamiliar, atypical, and new ways aside from what had been previously learned about the object's functionality. In a problem-solving activity, FF may delay or impede the process of finding a solution due to the inability to utilize the available materials and options presented. In this study, the incidence of FF in the context of computer programming was investigated. The students were given programming problems that require them to utilize a specific Java pre-defined method in a different manner from its basic and usual implementation. Results showed that all students exhibited at least 1 incidence of FF although no significant relationship was found between FF and their poor performance in solving the problems. The students explained that they got stuck in the known usage and it was difficult for them to think of other ways to use it in order to solve the problems.

Keywords: Functional Fixedness, Computer Programming, Novice Programmer

1. Introduction

Functional Fixedness (FF) was first hypothesized by Duncker (1945) which refers to the tendency of being fixed to specific uses or functionality of an object based on prior knowledge, how the objects were introduced, or the level of familiarity with the object. This fixation may then hinder the usage of the object in a different, atypical, or new manner which may be essential in a problem-solving task.

In the context of programming, researchers linked FF to cognitive issues in teaching students how to code functions (Mutsuddi & Frame, 2018) and loops (Smith, Paradise, & Smith, 2000). Schwill (1994) once recommended to teach programming using an object-oriented paradigm in the introductory level because it reflects the fundamental cognitive process of human by relating functions and usage to an object. He added that it should be done in ways that students could explore how to utilize object analogies of programming concepts beyond their basic definition. In coding user-defined functions, students show fixation with the function structure so when presented with sample integer functions, there was a difficulty coding user-defined functions for string. Similarly, in teaching the usage of for-loops using a series of examples where loops were used to sum elements, the students were fixated on the specific usage (i.e. summing elements) that when asked to write a looping statement for a different purpose, they tend to get stuck. However, these studies did not present an actual programming experiment where students could exhibit FF.

This study aims to revisit the concept of functional fixedness and investigate its incidence among student programmers taking up an introductory programming course and its implication on the students' problem-solving performance on programming problems.

2. Functional Fixedness

One trait that distinguishes humans from other creatures is that humans are creative beings by nature (Ward, Smith, & Vaid, 1997). This trait contributes to the ability of humans to find solutions to problems in addition to general intelligence. However, some psychological phenomena could impede the extent of creativity and one of them was described by cognitive psychologists as fixation. One reason for fixation is functional fixedness which refers to the inability to use an object in a different manner other than what had been previously known about it (Duncker, 1945). Several studies (Flavell,

Cooper, & Loisselle, 1958; Glucksberg & Danks, 1968; Jensen, 1960) further showed that functional fixedness occurs when the function of a key object is explicitly demonstrated, taught, or explained prior to the presentation of a problem that requires the utilization of the object in a way that was not taught.

One of the several experiments conducted to support this phenomenon was called the “box” problem (Duncker, 1945; Frank & Ramscar, 2003). The subject was presented with different materials including candles, matches, thumbtacks and boxes of varying sizes. The task was to fix the candles on the wall and light them. The solution is to tack the boxes to the wall using the thumbtacks and these boxes would serve as the bases for the candles. Then the candles are lit and fixed on the boxes with some wax. For the first group, the materials were presented in a way that the boxes were used as container for the candles, thumbtacks, and matches. For the second group, the materials were just scattered on the table. More subjects from the second group solved the problem than from the first group. Duncker explained that this is because the subjects from the first group perceived the boxes just as containers and it was difficult for them to lose the function associated with it.

In the context of computer programming, Perkins and Martin (1986) have previously explored this among novice programmers and argued that FF is a matter of negative knowledge transfer wherein knowing “how” sometimes impairs rather than support one’s performance in the application of a solution in an unfamiliar context. Another related phenomenon to FF called the Einstellung Effect (EE) which refers to an individual’s bias towards a familiar, working solution (Luchins, 1942; Luchins & Luchins, 1961) has been investigated among programming students (Obispo, Castro, & Rodrigo, 2018). EE is also a type of fixation and surprisingly, the study showed that, in a single programming activity, EE had a positive effect on the performance of the student programmers.

However, FF is different from EE in that FF deals with separate functionality of the objects and not a collective solution such in the case of EE. And since there have been no recent empirical studies to investigate the implications of FF on the performance of student programmers, this warrants further investigation.

3. Experiment

The participants of this investigative study were thirty (30) freshmen students taking up Bachelor of Science in Information Technology (BSIT) from a state university in southern Philippines. They were all considered as novice programmers based on their programming grades and also because they were only on their first year in the BSIT program. They were presented with the 3 Java pre-defined methods: Java String `indexOf()` and `startsWith()` as well Java Integer `parseInt()` which they have to use to solve 3 programming problems. These methods had already been taught to them in prior lectures in their programming classes.

The use of both `Integer.parseInt()` and `string.startsWith()` are straightforward but the participants must realize the `string.indexOf()` could be used in a different way other than just finding an index of a character or string. They were instructed that they will use these methods along with the basic program control structures in order to solve the following programming problems:

Problem 1: A string `s1` contains various integers separated by spaces. `s1 = “50 1 822 3 6 49 5 61 74 87”`
The task is to save the individual integers delimited by space into an integer array called `myArr`.

Problem 2: Ask the user to input a full name. Then, ask the user to input a prefix composed of 3 characters. The program should then check if the last name from the first input starts with the prefix.

Problem 3: Ask the user to input a string. Check if the string includes the substring “love”.

There could be several approaches to solve each of the three problems using the given methods. After the experiment, the students were then interviewed on how they felt about the programming activity where they discussed their solutions and their thoughts about using the given methods.

4. Results

The incidence of FF was based on whether the student was able to utilize `indexOf()` to solve the problem. One FF point was given if the student failed to use of `indexOf()` in order to solve the programming problem. Program codes submitted by the participants for each problem was analyzed in order to measure their performance and to determine the incidence of FF. Aside from looking at the

program logic, the codes were also checked for syntax and runtime errors to account for other possible reasons in cases where the student was unable to solve the problem. A point was given only if the program was able to execute and perform the task required in the problem and the `indexOf()` method was used as part of the solution. Only 6 students were able to earn points from successfully solving the programming problems but all students exhibited at least 1 incident of functional fixedness.

When the relationship between the FF incidence and student performance in terms of solving the problems was investigated, no significant relationship was found $r(30)=-0.19$, $p=0.51$. However, after submitting their program codes, the students were interviewed about the programming activity and the concept of functional fixedness was explained to them. The students reasoned that they got stuck in the known examples demonstrating the basic and usual usage of the `indexOf()` method and they cannot think of other ways to use it that will help solve the problem.

In conclusion, FF does exist among novice computer programmers because students are not yet adept in the creative ways on how these pre-defined methods could be used in various programming purposes aside from those that they have previously seen from examples. Although no significant relationship was found with performance in the experiment, the fact that a total of only 6 students out of 30 were able to solve at least 1 problem because others cannot think of how to use a previously known Java pre-defined method to solve the problem means that demonstrating and encouraging creativity is essential as a pedagogical technique in classes for novice student programmers. The authors would like to recommend further experiments and analysis that may involve intermediate programmers, more pre-define methods, and other programming components.

Acknowledgements

The authors would like to thank the students from the Institute of Computer Applications who participated in the experiment.

References

- Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58(5).
- Flavell, J. H., Cooper, A., & Loisselle, R. H. (1958). Effect of the number of pre-utilization functions on functional fixedness in problem solving. *Psychological Reports*, 4(3), 343–350.
- Frank, M. C., & Ramscar, M. (2003). How do presentation and context influence representation for functional fixedness tasks? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25.
- Glucksberg, S., & Danks, J. H. (1968). *Effects of discriminative labels and of nonsense labels upon availability of novel function*. 7, 72–76.
- Jensen, J. (1960). On functional fixedness: Some critical remarks. *Scandinavian Journal of Psychology*, 1(1), 157–162.
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54(6), i.
- Luchins, A. S., & Luchins, E. H. (1961). Einstellung effect in social learning. *The Journal of Social Psychology*, 55(1), 59–66.
- Mutsuddi, A., & Frame, D. (2018). Can Cognitive Science make it Easier for Novice Students to Learn to Code Functions? *Proceedings of the 19th Annual SIG Conference on Information Technology Education*, 149. International World Wide Web Conferences Steering Committee.
- Obispo, J. R. C., Castro, F. E. V. C. G., & Rodrigo, M. M. T. (2018). Incidence of Einstellung Effect among Programming Students and its Relationship with Achievement. *Proceedings of the 1st Information Computing Education*. Presented at the Information Computing Education, Philippines.
- Perkins, D. N., & Martin, F. (1986). Fragile knowledge and neglected strategies in novice programmers. *At Empirical Studies of Programmers, 1st Workshop, Washington, DC*, 213–229.
- Schwill, A. (1994). Cognitive aspects of object-oriented programming. *IFIP WG 3.1 Working Conference "Integrating Information Technology into Education"*.
- Smith, D. K., Paradice, D. B., & Smith, S. M. (2000). Prepare your mind for creativity. *Communications of the ACM*, 43(7), 110–116.
- Ward, T. B., Smith, S. M., & Vaid, J. (1997). Conceptual Structures and Processes in Creative Thought. In *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 1–27). Washington, D.C.: American Psychological Association.

MAIN CONFERENCE C2
COMPUTER-SUPPORTED COLLABORATIVE LEARNING
(CSCL) AND LEARNING SCIENCES

- FULL PAPER -

RECIPROCAL KIT BUILD APPROACH FOR PEER-TO-PEER COMMUNICATION: RELATIONSHIP BETWEEN SIMILARITIES ON KNOWLEDGE, TRANSFER OF KNOWLEDGE, AND AFFECTIVE RESPONSES..... 101
LIA SADITA, TSUKASA HIRASHIMA, YUSUKE HAYASHI, PEDRO GABRIEL FONTELES FURTADO, KASIYAH JUNUS, & HARRY BUDI SANTOSO

FLARE-FORK: A PEDAGOGY FOR EXPANDING PROBLEM AND SOLUTION SPACE FOR DESIGN PROBLEM SOLVING 111
SOUMYA NARAYANAN, SAHANA MURTHY

TOWARDS AN INTEGRATED FRAMEWORK OF GROUP AWARENESS SUPPORT FOR COLLABORATIVE LEARNING IN SOCIAL MEDIA 121
LISA OLLESCH, SVEN HEIMBUCH & DANIEL BODEMER

GROUP AWARENESS INFORMATION TO SUPPORT ACADEMIC HELP-SEEKING 131
CHRISTIAN SCHLUSCHE, LENKA SCHNAUBERT & DANIEL BODEMER

- SHORT PAPER -

CHARACTERIZATION OF DIFFERENT INSTANTIATIONS OF MATHEMATICAL BLINDNESS..... 141
ERROL MATTHEW C. GARCIA, MARIA ALVA Q. ABERIN

HOW TO MEASURE THE COLLABORATIVE PROBLEM-SOLVING COMPETENCY BASED ON CONVERSATIONAL AGENT 146
GE WANG, BIAN WU, YILING HU

INTRODUCTION OF EDUCATIONAL TECHNOLOGY ENGAGEMENT MODEL 152
IBRAHIM HAKKI BULUT & OMER DELIALIOGLU

COGNITIVE GROUP AWARENESS TOOLS: VERSATILE DEVICES TO GUIDE LEARNERS TOWARDS DISCREPANCIES 158
LENKA SCHNAUBERT, SVEN HEIMBUCH, MELANIE ERKENS & DANIEL BODEMER

SCAFFOLDING COMPUTER-SUPPORTED COLLABORATIVE LESSON DESIGN: A SPIRAL MODEL 165
WENLI CHEN, JESMINE S H TAN

HOW SHARED CONCEPT MAPPING FACILITATES EXPLANATION ACTIVITIES IN COLLABORATIVE LEARNING: AN EXPERIMENTAL INVESTIGATION INTO LEARNING PERFORMANCE IN THE CONTEXT OF DIFFERENT PERSPECTIVES 172
SHIGEN SHIMOJO & YUGO HAYASHI

A CROWD-PROGRAMMING APPROACH FOR COMPUTATIONAL THINKING EDUCATION..... 178
CHRISTOPHER KRIZANOVIC, NGUYEN-THINH LE & NIELS PINKWART

USING KNOWLEDGE FORUM TO SUPPORT THE DEVELOPMENT OF STEAM LITERACIES 185
LEANNE MA & THELMA AKYEA

- POSTER PAPER -

SURVEYING THE RELATIONSHIPS BETWEEN STUDENTS' EPISTEMIC CURIOSITY AND THEIR ONLINE ACADEMIC HELP SEEKING BEHAVIORS IN HIGHER EDUCATION..... 191
KUN-HUNG CHENG

THE ROLE OF TECHNOLOGY IDENTITY AMONG STUDENTS IN RURAL AREAS USING A WEB-BASED TUTORING SYSTEM.....	194
MAY MARIE P. TALANDRON-FELIPE	
PROPOSING CURRICULUM AND LEARNING ENVIRONMENT DEVELOPMENT FOR GLOBAL LIBERAL ARTS EDUCATION INCORPORATING FUTURE WORK SKILLS	197
TOSH YAMAMOTO, JULING SHIH, CHRIS PANG, BENSON ONG	
PATTERNS OF STUDENT LEARNING BEHAVIOUR WITHIN TECHNOLOGY-SUPPORTED SOCIALISED LEARNING CONTEXTS.....	200
HANYUNING LIN, MATHEW R PRYOR	
INVESTIGATING THE EFFECTS OF COGNITIVE STYLE ON BLENDED MUSEUM LEARNING	203
TIEN-YU HSU, HSIN-YI LIANG, JUN-MING CHEN	
DEVELOPMENT OF AN INTERACTIVE LEARNING MODULE FOR VISUALIZING SELF-REGULATED LEARNING SKILLS	206
HARRY B. SANTOSO & LUTHFI K. PUTRA	
INVESTIGATING STEM-BASED LEARNING PACKAGE FOR ENHANCING PROGRAMMING AND PROBLEM SOLVING SKILLS	209
PHIRAPHONG PROMJAN & NADH DITCHAROEN	

Reciprocal Kit Build Approach for Peer-to-peer Communication: Relationship between Similarities on Knowledge, Transfer of Knowledge, and Affective Responses

Lia SADITA^{a*}, Tsukasa HIRASHIMA^a, Yusuke HAYASHI^a, Pedro Gabriel Fonteles FURTADO^a, Kasiyah JUNUS^b, & Harry Budi SANTOSO^b

^a*Department of Information Engineering, Hiroshima University, Japan*

^b*Faculty of Computer Science, Universitas Indonesia, Indonesia*

*lia@lel.hiroshima-u.ac.jp

Abstract: The present paper describes the Reciprocal Kit Build approach as a designed activity for collaborative concept mapping. We aimed to investigate the effect of differences in individual knowledge (both prior knowledge and knowledge on task) on knowledge transfer, collaborative product, and learners' affective responses during co-construction of a concept map with the Reciprocal Kit Build (RKB) approach. We categorized learners into two groups based on their prior knowledge equivalence and their degree of shared knowledge reflected in the individual maps. The RKB approach allowed learners to create an individual map, to reconstruct a concept map from their partner's map components (nodes and links), and to discuss similarities (or differences) between the initial map with the reconstructed map. The results showed that, following our proposed activity, transfers of individual knowledge regarding the shared and unshared knowledge were considerably high. Although, learners' differences on prior and shared knowledge did not significantly affect the knowledge transfers and the final collaborative products, different group composition influenced the experiences of learners.

Keywords: computer-supported collaborative learning, kit-build, collaborative concept map, knowledge convergence

1. Introduction

In collaborative learning, knowledge is exchanged and converge through social interaction (Weinberger, Stegmann, & Fischer, 2007). Scripts, scenarios, or visualization tools are designed to trigger meaningful interactions within group members. Concept map as a representational tool has been widely employed during collaboration to facilitate ideas generation, communication, and negotiation of meaning. Studies found that creating collaborative concept map have increased students' learning achievements and their positive attitudes, i.e.: motivation and responsibility (Basque & Lavoie, 2006).

Gnesdilow & Bopardikar (2010) suggested that the level of convergence achieved during collaborative concept mapping could have influenced individual performances after collaboration. Mutual understanding of the partner's perspectives and shared interpretations of the problem is an important requirement for collaboration (Jeong & Chi, 2007; Stoyanova & Kommers, 2002). Divergent ideas between group members have a significant impact on collaboration (Gnesdilow & Bopardikar, 2010; Stahl, 2003; Weinberger et al., 2007). Individual prerequisites and diverge in knowledge influence the benefits experienced by group members when learning together (Weinberger et al., 2007). They suggested heterogenous group composition to promote negotiation perspectives towards a shared understanding of classroom collaborative activities.

Another way to attain convergence is by nurturing group members to use the knowledge available to them, both shared and unshared knowledge resources, from their prior knowledge and from learning material (Frank Fischer & Mandl, 2002). Unfortunately, groups more often neglect unshared resources – that is, knowledge and information that only a few numbers of group members possessed or

have access to (Frank Fischer & Mandl, 2002). Providing an open communication environment where individual's shared and unshared knowledge is acknowledged, built, and elaborated is expected to foster knowledge convergence. Hence, improving learning outcomes in tasks and supporting conceptual change through discussion.

We have extended the collaborative concept mapping with Kit-Build, a closed-ended concept mapping approach to assess common understanding between the teacher and his students (Hirashima, Yamasaki, Fukuda, & Funaoi, 2015). In a practical classroom, Kit Build analyzer has been used to find learners' misconceptions and to improve the teacher's lesson plan in the subsequent class (Pailai, Wunnasri, Yoshida, Hayashi, & Hirashima, 2017; Yoshida et al., 2013). In a peer-to-peer context, this approach promotes exploratory talk during group discussion (Wunnasri, Pailai, Hayashi, & Hirashima, 2018a) and aids learners in dyads to share understanding based on individual pre- and post-maps (Wunnasri, Pailai, Hayashi, & Hirashima, 2018b). However, those studies have not identified how individual knowledge has been taken into consideration for constructing a collaborative map, an artifact where group members have to negotiate individual knowledge differences and to reach a common consensus on task. The use of RKB for collaborative concept mapping has shown that most groups produce better high-quality collaborative maps and there was an association between difference map visualization with score gain from individual to collaborative maps (Sadita et al., 2018). The preliminary study has not investigated how similarities of knowledge among each pair may influence the collaborative product as well as knowledge transfer from individual-to-group.

In the present study, we measure the level of convergence prior to collaborative concept mapping in regards to knowledge equivalence and shared knowledge (Weinberger et al., 2007) to gain a deeper understanding on how to form a group and whether our proposed activity let learners to constructively build on individual knowledge. First, we identify the effect of different group composition to learning effectiveness at two dimensions, i.e. as an interaction between group members and as a group achievement (Khamesan & Hammond, 2004; Molinari, 2013; Stoyanova & Kommers, 2002). Second, we survey the learners' affective responses to find out students' experiences after following our proposed activities on different group compositions.

2. Literature Review

2.1 Concept Map to Facilitate Communication in Collaborative Learning

Concept mapping as a representational tool is beneficial for collaborative learning at the individual level and group level (Stoyanova & Kommers, 2002). It makes individual knowledge more explicit and provides a room for reflection and elaboration of cognition (Stoyanova & Kommers, 2002). At a group level, it promotes establishing a common ground as a basis for building a shared understanding within group members (Jeong & Hmelo-Silver, 2016; Roschelle & Teasley, 1995; Stoyanova & Kommers, 2002). Prior studies have also shown that a concept map is an effective tool for elicitation of knowledge and communicating complex ideas (Frank Fischer & Mandl, 2002; Stoyanova & Kommers, 2002; Suthers, 2006; van Boxtel, van der Linden, & Kanselaar, 2000).

2.2 Reciprocal Kit Build

We have employed the Reciprocal Kit-Build (RKB) approach to allow students to generate and exchange ideas with their partner before collaborative concept mapping activities. There are three main parts of the RKB approach, i.e. individual map building, individual map reconstruction by partners, and difference map discussion (Wunnasri et al., 2018a). Prior studies have shown that individual map building in a private space helped students to explain their ideas during the collaborative session (F. Fischer, Bruhn, Gräsel, & Mandl, 2002; Gracia-Moreno, Cerisier, Devauchelle, Gamboa, & Pierrot, 2017).

Through the reconstruction of partner's map and difference map discussion, each group member is feasible to detect partner's comprehension and it leads to the elicitation of knowledge. Awareness of partner's knowledge is beneficial to maintain shared focus during problem-solving, and therefore, students solve the problem faster and more accuracy (Engelmann & Hesse, 2010). Specifically, in collaborative concept mapping with individual preparation, they may have more

sociocognitive conflicts during collaborative concept mapping when they have first overcome a reflective thinking process in a personal workspace (de Weerd, Tan, & Stoyanov, 2017; Gracia-Moreno et al., 2017). It possibly will hinder students to express their unshared knowledge or resources.

Providing a room to assist students to externalize own ideas or to review different perspectives from their partner in an active manner is expected to enrich interaction between group members. Different from standard individual map interchange, the RKB approach encourages students to produce more exploratory talk, which is valued for advancing critical thinking, reasoning, and problem-solving skills (Wunnasri et al., 2018a). After following the RKB activities, students in dyads are having similar knowledge on their post individual maps which indicate that they have higher knowledge convergence. However, previous researches on RKB have not explored how differences in prior knowledge affect the collaborative product, where they have to resolve individual differences to reach a common consensus, and how the differences influence the experiences of learners on collaborative learning (Sadita et al., 2018; Wunnasri et al., 2018a, 2018b).

3. Methods

3.1 *Experimental Settings*

We run the study in a Linear Algebra class for the first year of Computer Science students in one of a public university in Indonesia. The teacher selected topics that facilitated students to draw conceptual knowledge, i.e. the General Vector Spaces and the Inner Product Spaces. An introductory explanation about these topics and the relevant learning resources were delivered by the teacher before conducting the experiment. The participants consisted of 42 students who work in dyads, where 71% of them are men. They were familiar with concept mapping activities since the teacher usually draws a diagram to show the relationship among concepts or asks the students to create it by themselves after finishing a topic. The teacher determined some essential nodes ($n = 15$) be included in the map to aid students with common references, therefore they could maintain focus during the discussion.

We administered the experiment in a computer laboratory for about two hours, which was divided into two main phases, i.e. the individual and the collaborative phase. First, the students created an individual map in 25 minutes. During the collaborative phase, the students reconstructed a map given a set of unconnected nodes and links (components) from their partners' map (20 minutes), discussed a difference map (10 minutes), and created a group map collaboratively (30 minutes). Each group member worked in a close-proximity to build an individual map with a personal computer or laptop, then they used a single computer to draw a collaborative map.

3.2 *Variables and Measurements*

Knowledge convergence is defined as a process by which two or more people share mutual understanding through social interaction (Jeong & Chi, 2007). Weinberger, Stegmann, & Fischer (2007) have conceptualized knowledge convergence as knowledge equivalence and shared knowledge which can be evaluated prior to, during, or after collaboration. Knowledge equivalence refers to learners in a group possessing a similar degree of knowledge related to a specified subject, regardless of the specific concepts constituting knowledge content (Weinberger et al., 2007). While, shared knowledge alludes to the knowledge of specific concepts that learners within a group have in common (Weinberger et al., 2007).

We first measured knowledge convergence at group level prior to collaboration using two different measures, i.e. prior knowledge equivalence and shared knowledge on task. Afterward, we evaluated the learning effectiveness of collaborative problem solving operationalized in two dimensions as follows (Khamesan & Hammond, 2004; Molinari, 2013):

- at the level of the group as a whole, scored numerically on group concept map
- as an interaction between individual and group achievements, scored numerically on individual map and group concept map

We excluded the effectiveness at the level of the individual since we did not collect individual post-collaboration maps because of some limitations in a practical classroom situation. Those three metrics are originally introduced by Stoyanova & Kommers (2002) to provide a deeper quantitative

understanding of the processes of both learning and collaboration in collaborative concept mapping. Khamesan and Hammond (2004) have computed the reliability of the metrics with three raters and demonstrated high interrater reliability for most of the categories.

3.3 Knowledge Convergence Prior to Collaboration

Prior knowledge equivalence scores were calculated from the results of the mid-term test conducted a few days before the experiment. The questions in the test covered essential introductory materials required to understand the main topic in the concept map, but not included the conceptual knowledge that could be drawn in a map form such relationship among concepts. Measures of dispersion were used to analyze differences in prior knowledge between learners as in the prior study (Weinberger et al., 2007). First, individual mid-term tests were evaluated by the teacher. Second, standard deviations between the individual scores in each group were calculated. Last, the standard deviation was divided by the mean score to measure the coefficient of variation as a prior knowledge equivalence score.

Table 1

Sample of 4 Essential Information Included in The Map and Its Possible Substructure

No	Type of information	Possible nodes included in the substructure
1	An inner product space is a vector space with an additional structure called the inner product function	Inner Product (IP) Space – Vector Space (VS)
2	An inner product function takes each ordered pair in a vector space V to a number in R	IP function – domain: $V \times V$ & codomain: R
3	An inner product function is a function that has to satisfy the following axioms: additivity, homogeneity, positivity, & symmetry	IP function – 4 axioms: additivity, homogeneity, positivity, & symmetry
4	Vector is an element of a vector space V	vector – IPS (if the IPS is connected to VS), or vector – VS

Note: “ – ” represents a link / connection between nodes / concepts

We assessed shared knowledge quantitatively from individual concept maps using the approach proposed by Weinberger et al. (2007). First, the teacher defined what are essential information should be included in the maps, given a set of nodes as initial components to build a map. Then, she listed all possible and common substructures from all students’ generated maps. A substructure may consist of two or more connected nodes (concepts) which convey one information only (see Table 1). The propositions may have a few variations depending on the linking words written by the students. Second, the teacher marked whether a student’s map presented any essential information or not. Seven key substructures were expected to appear in the maps. Third, if a pair of learners share the ability to apply a specific concept, then we added the shared prior knowledge score of 1. Finally, we normalized the score by dividing it with the group mean value. In addition, we also defined unshared knowledge at the individual level to identify the degree of information that only possessed by a single member.

Table 2

Sample of Knowledge Distribution in a Group

No	Substructures	Group 01		
		Student A’s map	Student B’s map	Group map
1	Inner Product (IP) Space – Vector Space (VS)	○	○	○
2	IP function – domain: $V \times V$ & codomain: R	X	X	○
3	IP function – 4 axioms	○	X	○
4	Vector – IP Space or VS	○	○	X

Note:

○ : the substructure was available and correct

X : the substructure was not available or incorrect

Following the above procedures, individual knowledge scores of student A's and B's in Group 01 were 3 and 2 consecutively based on the number of correct substructures. Hence, resulting in a mean of 2.5 (Table 2). Group 01 achieved a shared knowledge value of 2 because both members were able to draw the first and the fourth substructures correctly. Subsequently, the normalized shared knowledge score of this group was $2 / 2.5$ or equal to 0.8.

The normalized prior knowledge equivalence score and shared knowledge score were applied to categorize the group. The groups which have normalized prior knowledge equivalence less than 0.2 were categorized as high knowledge equivalence groups and the groups with normalized shared knowledge score more than 0.7 were included in high shared knowledge groups. The prior knowledge equivalence scores provided the differences of individual performances on prior relevant topics, while the shared knowledge scores were more specific to knowledge on the task itself.

3.4 Learning Effectiveness Measures

We examined the learning effectiveness by using the concept map measures (Figure 1) proposed by Khamesan and Hammond (2004):

1. Individual-to-group transfer of shared knowledge (TSK_{AB}): the number of substructures shared by both individual and transferred to the collaborative map. The score was normalized with the number of shared substructures.
2. Individual-to-group transfer of unshared knowledge ($TUK_{A \text{ (or B)}}$): the number of unshared substructures in each individual and transferred to the collaborative map. The score was normalized with the number of unshared substructures.
3. Individual-to-group transfer (TK_{AB} : $TSK_{AB} \cup TUK_A \cup TUK_B$): the total number of transferred substructures from individual maps to the collaborative map. The score was normalized with the number of shared and unshared substructures.
4. Lost knowledge (LK_{AB} : $(IK_A \cup IK_B) \setminus TK_{AB}$): the number of individual substructures not transferred from individual maps to the collaborative map.
5. Group creativity (NK_{AB}): the number of new substructures in the collaborative map that was not included in both individual maps. The score was normalized with the number of unknown substructures.
6. Group achievement (GK_{AB}): the score of the collaborative map.

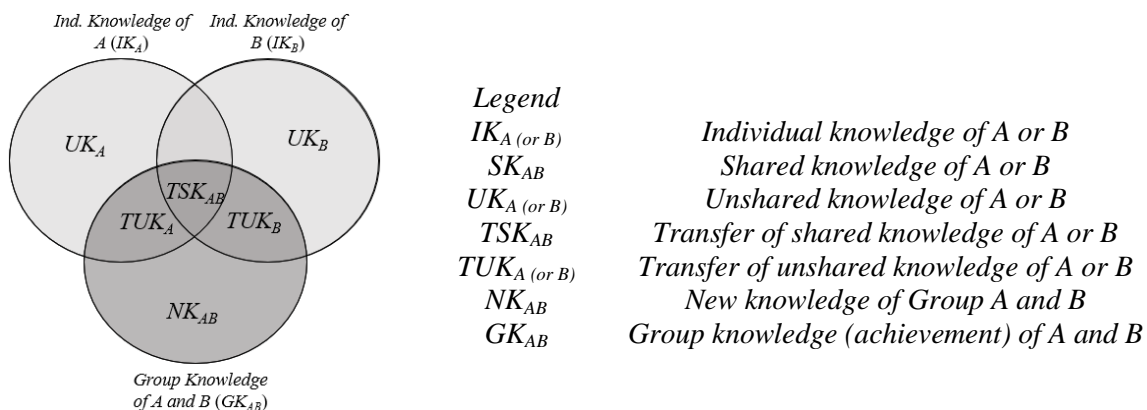


Figure 1. The Illustration of Learning Effectiveness Measures

As an example, from the Table 2, substructures (1) and (4) were the shared knowledge on the task before collaboration, substructure (3) was the unshared knowledge of student B, and substructure (2) was the unknown substructures of Group 01 (ignorance). After the collaboration, the students wrote the substructures (1) to (3) correctly, so we regarded those substructures as Group Knowledge. Specifically, substructure (1) was considered as the individual-to-group transfer of shared knowledge, substructure (2) was categorized as new knowledge, and substructure (3) was referred as the individual-to-group transfer of unshared knowledge. Unfortunately, the Group 01 members did not write substructure (4), thus it became the lost knowledge.

3.5 Affective Responses

We conducted a survey to capture participants' experiences while following RKB activities. The questionnaire consisted of 15 closed-ended items related to attractiveness, stimulation, and perspicuity subscales adapted from an Indonesian language version of User Experience Questionnaire (Santoso, Schrepp, Kartono, Yudha, & Priyogi, 2016). The students were requested to choose a Likert scale from 1 to 7. Six open-ended questions were delivered to capture learners' positive and negatives experiences in each step of collaborative learning activities. All questionnaire items had been face-validated by the teacher before distributed to the students. Cronbach's alpha coefficients were 0.74, 0.84, and 0.77 for attractiveness, stimulation, and perspicuity subscale, respectively, showing good internal consistency.

4. Results

4.1 Learning Effectiveness

Before creating collaborative maps, 82% percent of the important substructures were written in students' individual maps, while the number of the substructures that were not included in the individual maps was as many as 18% ($n = 27$). More than half of those written substructures were shared knowledge (Figure 2). Those shared and unshared knowledge were available in the collaborative maps as many as 91.67%, the remaining became non-transferred (lost) knowledge. Almost all shared knowledge was transferred to the collaborative map, while the percentage of neglected unshared knowledge was 15% of total unshared knowledge. There are 14 groups who were feasible to extend their group map with new information (substructures), that did not exist in their individual maps. From those groups, we found that they were able to draw 8 new important substructures. The number of unknown information (ignorance) in the collaborative maps were also decreasing, from 18% to 13%.

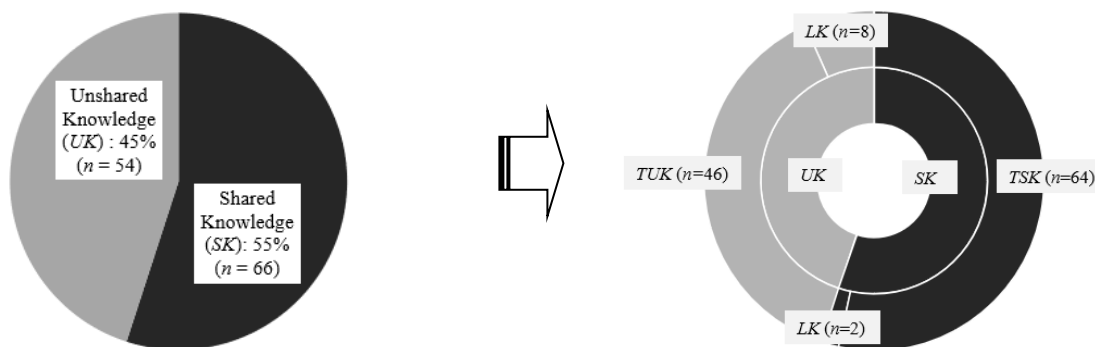


Figure 2. Distribution of The Amount of Shared and Unshared Knowledge in The Individual Maps Prior to Collaboration (left-side) and Distribution of The Individual Knowledge Transferred to Collaborative Maps in All Groups (right-side).

(Note: Please see Section 3.4 to understand the abbreviations in the graph)

Figure 3 and 4 display the distribution of knowledge transfer and group creativity among different conditions regarding their prior knowledge equivalence and shared knowledge on the individual concept maps. Transfer of shared and unshared knowledge in most of the groups in all conditions have similar median values with different score distribution. Two groups of high prior knowledge equivalence and high shared knowledge conditions did not convey their understanding or reach different consensus, i.e. Group 09 and Group 14. There are three out of 11 groups in high shared knowledge condition who did not have unshared knowledge. The remaining groups with the unshared knowledge in low prior knowledge equivalence and low shared knowledge condition have higher agreement to transfer the knowledge. From the 14 groups who had new knowledge, the number of groups in each condition was similarly distributed ($n = 7$). The groups with low shared knowledge had a higher tendency to create new knowledge (Figure 4).

Furthermore, we also investigated whether individual ability affected group tendency to transfer the unshared knowledge by calculating the correlation between individual map score and the normalized score of individual unshared knowledge transfer. Results of the Pearson correlation

indicated that there was no association between individual map score and number of unshared knowledge transfer, ($r(22) = -.06014, p = .7801$). The student with lower individual performance than their partner could possibly to transfer his unshared knowledge, and vice versa, the one with higher individual performance might unable to convey the unshared knowledge.

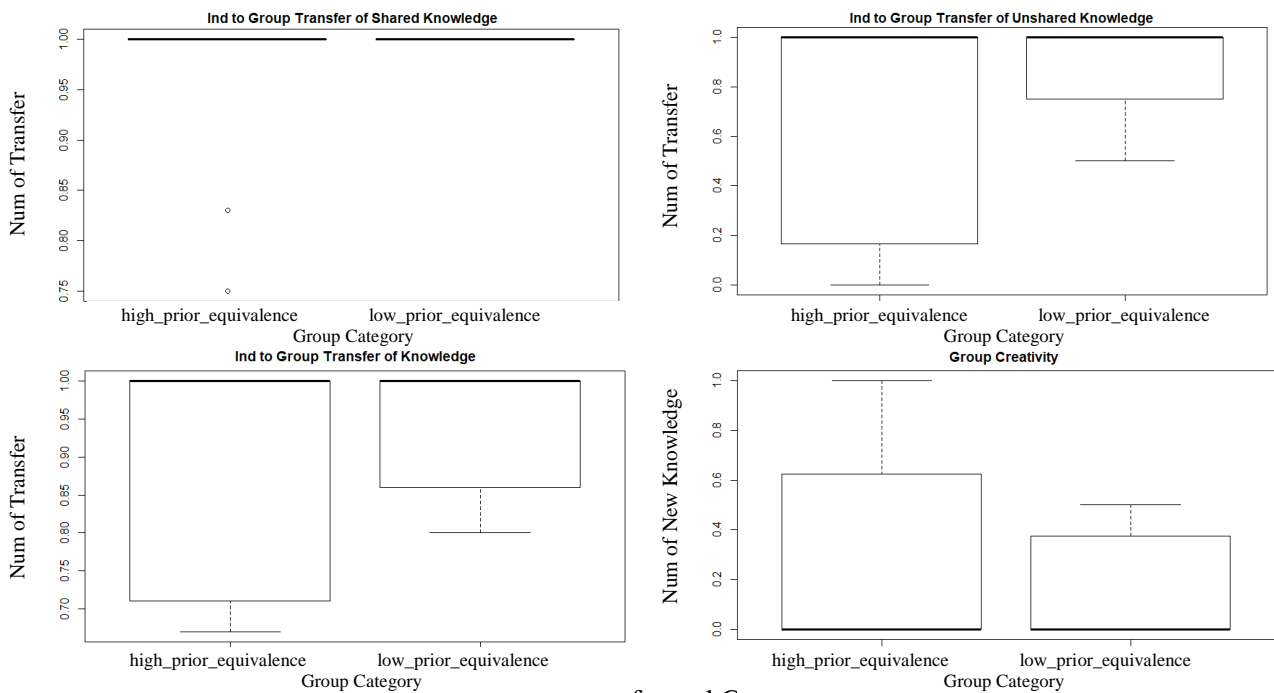


Figure 3. Distribution of Knowledge Transfer and Group Creativity in High- and Low- Prior Knowledge Equivalence Groups ($n = 11$ and $n = 10$, respectively).

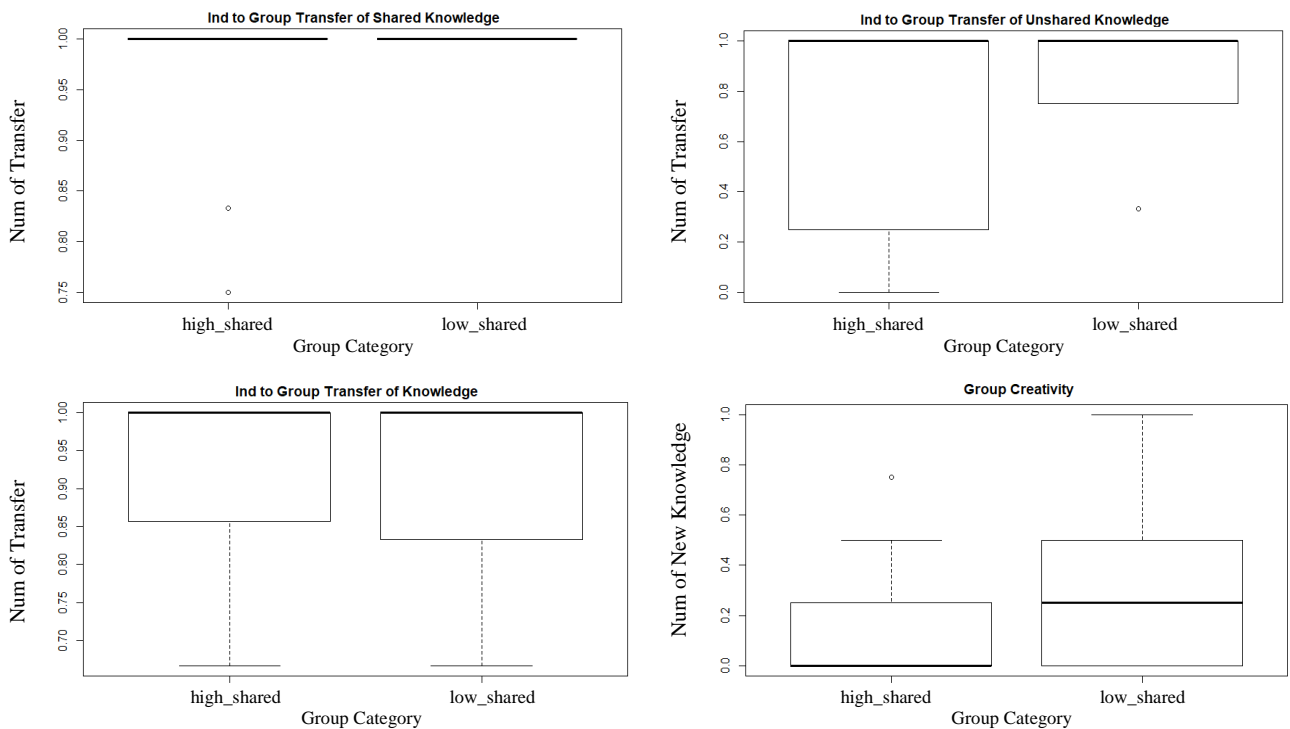


Figure 4. Distribution of Knowledge Transfer and Group Creativity in High- and Low- Shared Knowledge Groups ($n = 10$ and $n = 11$, respectively).

All collaborative map scores were in the range of 75-100 for all conditions ($M = 90, SD = 7.49$). These scores were higher than individual map scores ($M = 72.21, SD = 25.76$). Group achievements did not differ significantly between low- and high-prior knowledge equivalence conditions ($p = 0.47$), or

between low- and high-shared knowledge conditions ($p = 0.302$), though there is dissimilarity of distribution among them (Figure 5).

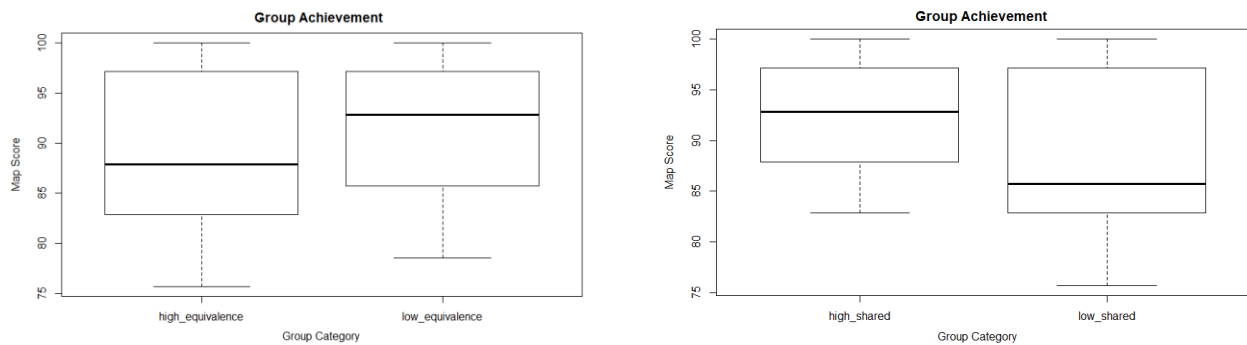


Figure 5. The Collaborative Map Scores Differentiated by Prior Knowledge Equivalence (left-side) and Shared Knowledge on Task (right-side)

4.2 Learners' Affective Responses

Figure 6 shows the distribution of the affective scores among groups with different shared knowledge scores. A Kruskal Wallis rank-sum test indicated that there was a significantly different between the groups in higher similarity scores and lower similarity scores ($H(13) = 56.885, p < .001$). However, the differences were rather small, which demonstrated that the low similarity users were still positive towards the activities, though less positive. Stimulation subscale received the highest rating, followed by attractiveness then perspicuity subscales.

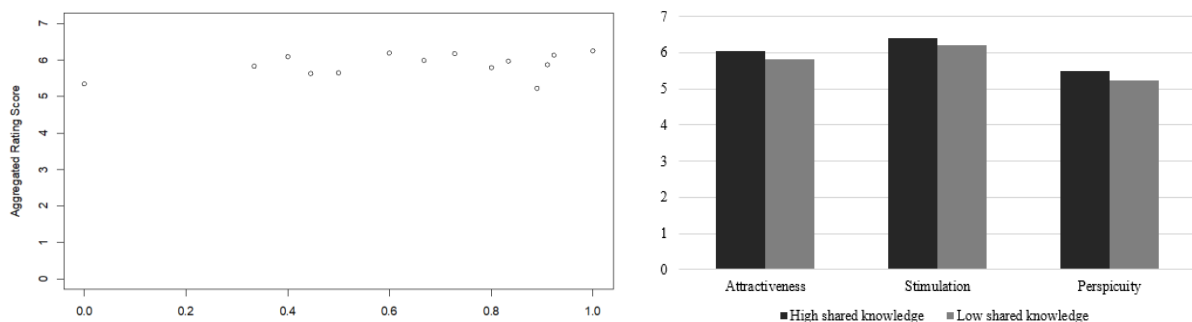


Figure 6. Distribution of Affective Responses Across Different Shared Knowledge Scores

From the open-ended questions, we found that some participants in both conditions mentioned comparable on-task difficulties concerning dissimilarities of ideas or opinions, i.e.: “Difficult to read when the number of visualized differences is too many ($n = 6$).”, “It was hard to read or understand the difference map ($n = 2$).”, “It was difficult to integrate different opinions in order to reach a (group) consensus or determine which one is the correct representation ($n = 5$).”, “The use of ambiguous links makes it hard to select the most suitable relation between two concepts ($n = 1$)”.

5. Discussion and Conclusion

Collaborative concept mapping with Reciprocal Kit Build approach allows learners to represent and manipulate their individual cognitive structures and let their partners provide feedbacks after initial map reconstruction and difference map visualization. This activity provides an active means to review individual maps and to elicit new information. Reviewing members' individual maps as access to distributed cognitive resources positively influence the broadness of group problem solution (Stoyanova & Kommers, 2002). Her study has also suggested that a process of knowledge acquisition and creation through direct interaction have an impact on group learning effectiveness, which consistent with our preliminary findings (Sadita et al., 2018).

The present study displays the learning effectiveness as the interaction between individual to group knowledge, specifically knowledge transfer. The results show that the amount of knowledge transfer is considerably high in all group conditions. Furthermore, the degree of knowledge differences within the group members may not significantly affect the amount of knowledge transfer. While some studies have reported that groups often abandon the unshared knowledge or resources (Frank Fischer & Mandl, 2002; Gracia-Moreno et al., 2017), our study indicates that transfer of both shared and unshared individual knowledge is more than 85% when using RKB. It is interesting to note that a few groups such as Group 09 and Group 14, who were the high prior knowledge convergence condition did not transfer all of their shared knowledge. Further investigation of their behavior is essential to reveal these specific group problems.

Moreover, we found the weak correlation between individual map scores and normalized transfer of unshared knowledge. It is indicating that during collaboration students were able to detect important substructures with less consideration on who are the source of information. The students more often to not merely follow a certain group member. They acknowledge the partner's perspectives and take into account differences in knowledge.

The affective responses of the groups with different shared knowledge scores demonstrated that learners in higher shared knowledge are slightly more positive than the lower shared knowledge conditions. Participants in both conditions show similar patterns, they thought that our activities were more stimulating and attractive, rather than perspicuous. Difficulties appeared when they faced differences in ideas or perspectives and need to resolve those conflicts in order to reach a single group solution. Though pursuing conflict resolution is essential for conceptual change and advancement of knowledge in collaborative learning (Chan, Burtis, & Bereiter, 1997; Roschelle, 1992; van Boxtel, van der Linden, Roelofs, & Erkens, 2002), the learners may feel less positive and it may influence overall learning experiences in collaborative situations. Further research on how computer-based visualization can be utilized to aid learners during conflict-oriented consensus building and integration-oriented consensus building is indispensable.

In summary, our study found that learners in different prior knowledge levels benefit similarly, with respect to the transfer of their individual knowledge, following the proposed activities. Different group composition is not necessarily affected knowledge transfer and collaborative outcomes. However, the amount of joint knowledge between the group members can possibly have more effects on the group outcomes, since the RKB system enables individual knowledge structures more tangible and are ready to be manipulated by their partners. Different opinions or understanding with regard to the collaborative task can affect overall learners' experiences in a collaborative environment.

Our current works have evaluated the learning effectiveness only at two different dimensions (i.e. group and interaction level). It seems that our approach can attain learning effectiveness at the individual level as well, because of high knowledge transfer during collaboration. However, there is a lack of evidence related to individual performances after collaboration. Further studies with a large number of participants from different subjects should be conducted to identify the broadness of our approach. For the future works, it is also interesting to compare the results of groups with reciprocal teaching activity and conventional collaborative concept map without the reciprocal cycle.

References

- Basque, J., & Lavoie, M.-C.-. C. (2006). Collaborative concept mapping in education: Major research trends. *Concept Maps: Theory, Methodology, Technology. Proc. of the Second Int. Conference on Concept Mapping., 1*, 79–86. Retrieved from <http://eprint.ihmc.us/119/>
- Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci1501_1
- de Weerd, J., Tan, E., & Stoyanov, S. (2017). Fostering interdisciplinary knowledge construction in computer-assisted collaborative concept mapping. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10474 LNCS(Ccm), 391–396. https://doi.org/10.1007/978-3-319-66610-5_32
- Engelmann, T., & Hesse, F. W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning*, 5(3), 299–319. <https://doi.org/10.1007/s11412-010-9089-1>
- Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, 12(2), 213–232.

[https://doi.org/10.1016/S0959-4752\(01\)00005-6](https://doi.org/10.1016/S0959-4752(01)00005-6)

- Fischer, Frank, & Mandl, H. (2002). Facilitating knowledge convergence in videoconferencing environments: the role of external representation tools. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 623–624).
- Gnesdilow, D., & Bopardikar, A. (2010). Exploring convergence of science ideas through collaborative concept mapping. *The Journal of the Learning Sciences*, *1*, 698–705. Retrieved from <http://dl.acm.org/citation.cfm?id=1854449>
- Gracia-Moreno, C., Cerisier, J.-F., Devauchelle, B., Gamboa, F., & Pierrot, L. (2017). Collaborative knowledge building through simultaneous private and public workspaces. *Data Driven Approaches in Digital Education*, *10474*, 553–556. <https://doi.org/10.1007/978-3-319-66610-5>
- Hirashima, T., Yamasaki, K., Fukuda, H., & Funaoi, H. (2015). Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, *10*(1), 17. <https://doi.org/10.1186/s41039-015-0018-9>
- Jeong, H., & Chi, M. T. H. (2007). Knowledge convergence and collaborative learning. *Instructional Science*, *35*(4), 287–315. <https://doi.org/10.1007/s11251-006-9008-z>
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*(2), 247–265. <https://doi.org/10.1080/00461520.2016.1158654>
- Khamesan, A., & Hammond, N. (2004). Taxonomy of analysis levels of learning effectiveness in collaborative concept mapping. *Proc. of the First Int. Conference on Concept Mapping*.
- Molinari, G. (2013). From learners' concept maps of their similar or complementary prior knowledge to collaborative concept map: Dual eye-tracking and concept map analyses. *Psychologie Francaise*, *62*(3), 1–19. <https://doi.org/10.1016/j.psfr.2015.11.001>
- Pailai, J., Wunnasri, W., Yoshida, K., Hayashi, Y., & Hirashima, T. (2017). The practical use of Kit-Build concept map on formative assessment. *Research and Practice in Technology Enhanced Learning*, *12*(1), 20. <https://doi.org/10.1186/s41039-017-0060-x>
- Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*. https://doi.org/10.1207/s15327809jls0203_1
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer Supported Collaborative Learning*. https://doi.org/10.1007/978-3-642-85098-1_5
- Sadita, L., Hirashima, T., Hayashi, Y., Wunnasri, W., Pailai, J., Junus, K., & Santoso, H. B. (2018). Preliminary Study on the Use of Reciprocal Kit Build for Collaborative Learning. In *The 26th International Conference on Computers in Education (ICCE 2018)* (pp. 133–142).
- Santoso, H., Schrepp, M., Kartono, Y., Yudha, A., & Priyogi, B. (2016). Measuring user experience of the student-centered e-learning environment. *The Journal of Educators Online*. <https://doi.org/10.9743/JEO.2016.1.5>
- Stahl, G. (2003). Building collaborative knowing: Elements of a social theory of learning. In *What we know about CSCL in higher education*. https://doi.org/10.1007/1-4020-7921-4_3
- Stoyanova, N., & Kommers, P. (2002). Concept mapping as a medium of shared cognition in computer-supported collaborative problem solving. *Journal of Interactive Learning Research*. <https://doi.org/10.1161/CIRCOUTCOMES.110.942318>
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. In *International Journal of Computer-Supported Collaborative Learning* (Vol. 1, pp. 315–337). <https://doi.org/10.1007/s11412-006-9660-y>
- van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330. <https://doi.org/10.1007/s40815-016-0287-5>
- van Boxtel, C., van der Linden, J., Roelofs, E., & Erkens, G. (2002). Collaborative concept mapping: Provoking and supporting meaningful discourse. *Theory Into Practice*, *41*(1), 40–46. https://doi.org/10.1207/s15430421tip4101_7
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2007.03.007>
- Wunnasri, W., Pailai, J., Hayashi, Y., & Hirashima, T. (2018a). Reciprocal Kit-Build concept map: An approach for encouraging pair discussion to share each other's understanding. *IEICE Transactions on Information and Systems*.
- Wunnasri, W., Pailai, J., Hayashi, Y., & Hirashima, T. (2018b). Reciprocal Kit-Building of concept map to share each other's understanding as preparation for collaboration. In *Proceeding of the 19th International Conference on Artificial Intelligence in Education, London, United Kingdom*.
- Yoshida, K., Osada, T., Sugihara, K., Nino, Y., Shida, M., & Hirashima, T. (2013). Instantaneous assessment of learners' comprehension for lecture by using kit-build concept map system. In *International conference on human Interface and the Management of Information* (pp. 175–181).

Flare-Fork: A pedagogy for expanding problem and solution space for design problem solving

Soumya NARAYANAN*, Sahana MURTHY

Indian Institute of Technology Bombay, India

* soumya.n@iitb.ac.in

Abstract: In design, fixation is a recurring problem with novices where they choose the first solution that they think of for a given problem rather than explore a range of possible solutions. Fixation can result in barriers in design process, restricted search and premature commitment to solutions that may be sub-optimal. Among undergraduate engineering students, design fixation is due to limited prior conceptual knowledge, insufficient knowledge of design heuristics, and disconnected and evolving knowledge base, skills and strategies. One mechanism to overcome design fixation is expansion of problem and solution space where learners explore a variety of aspects related to the design problem such as constraints, assumptions, concepts and principles, before deciding on the solution to be followed. To be able to expand problem and solution space, novice designers require support.

We have devised a pedagogical strategy flare-fork which scaffolds students' problem space and solution space expansion in the flare phase and generation of multiple constraint optimized designs in the fork phase. Flare-fork is a systematic process that supports students to build on their opportunistic ideas to expand design boundaries and subsequently identify trade-offs to generate a variety of solutions tailored to fulfill a combination of constraints and assumptions. The pedagogy brings together design strategies, tools and practices such as rapid ideation, shared visual representation, categorization and techniques to restructure thinking patterns in a collaborative environment. In this paper, we first describe the conceptual design of our proposed pedagogy. We then describe a study to examine how the intervention operationalizing the flare phase of the pedagogy supports students in expanding their problem and solution space.

Keywords: Engineering product design, design fixation, expansion of problem-solution space

1. Introduction

Engineering design problems are generally ill structured and complex, making the design activity exploratory in nature. This often leads to designs having emergent characteristics in which the problem and solution co-evolve (Dorst and Cross, 2001). Solutions can vary depending on the initial point of view, the concepts used, the extent of problem explored and the kind of constraints used. While none of the solutions may be wrong, their suitability to the given context may vary.

Students who are novice designers find the unstructured nature of engineering product design quite challenging. They face several design related challenges such as in framing problem, identifying goals, constraints, requirements, making decisions, considering alternatives and switching between different modes of thinking, visualizing and gathering information (Atman et al., 2007). Many novice designers fixate on their first solution idea (depth first approach) and do not make the effort to look for alternate solutions by expanding their problem and solution space (breadth first approach) (Atman et al., 2007). This premature closure on a solution can lead to sub-optimal, pedestrian solutions that can prove intractable at a later stage of design.

One way to prevent premature commitment to a solution is enabling students to expand problem and solution space. The problem space is the task environment of the designer comprising of initial problem state, goals and requirement. The solution space encompasses the possible solutions and the means of evaluating if the solution meets the requirements. During co-evolution, designers iteratively explore the problem requirements and design solutions. In this process new variables emerge in response to changes in each of problem and solution space, which could include new requirements, thereby expanding the initial problem space or new potential solutions which expand the solution space (Hay et al., 2017).

Flare-fork is a pedagogy that we have designed to support and promote expansion of problem and solution space by novice designers while solving an engineering product design problem. The pedagogy combines collaboration, shared visual representation, and strategies to restructure thinking patterns such that students can opportunistically decompose the design problem but still benefit from the subtle structuring introduced by the pedagogy. Flare supports unrestricted search of problem and solution space by facilitating students to think divergently with divergent thinking strategies. Fork supports students in identifying constraints and trade-offs and devising constraint appropriate conceptual designs for the given design problem.

We have developed a collaborative learning environment for the flare phase of the pedagogy. We have used conjecture mapping as a way to articulate how the different components of the intervention viz. the activities, artifacts, resources, scaffolds and the transactions between these components, interact to facilitate problem and solution space expansion. In this paper, we describe an exploratory study as part of the first cycle of design-based research (DBR), where we tested key design and theoretical conjectures that inform the pedagogical support of our collaborative learning environment. The specific research question we are addressing are:

1. How do students explore problem and solution space during design problem solving?
2. In what ways did the different features of the intervention support the expansion of problem and solution space?

2. Theoretical background

2.1 Engineering design

Engineering design is ill structured in nature. During design process, designers structure and frame problem, generate and synthesize solutions, evaluate the alternatives, make decisions towards realizing the project (Atman et al., 2007), (Hay et al., 2017). The elements of design considered include requirements, functionalities, form or structure, behaviour, external environment including stakeholders and constraints (Gero and Kannengiesser, 2004). The complex and ambiguous nature of design can prove difficult for students who then resort to known solutions and pursuing their first idea. This can cause problem of fixation. Fixation has been synonymized with restricted search, satisfaction in search and premature commitment (Crilly and Cardoso, 2017) indicating a restriction of range of ideas that are considered. Fixation occurs when designers unconsciously adhere to prior designs and mimic design features, consciously block generation of new ideas in favour of familiar solution paths (knowledge based fixation), or intentionally resist thinking of new ideas in favour of successful designs (conceptual fixation) (Youmans and Arciszewski, 2014). Fixation can hinder creativity and lead to sub-optimal solutions riddled with problems that require additional fixes. Ways of reducing fixation includes modifying the design environment such that it allows unrestricted consideration of ideas. Group work and rich interactive design environments have been known to lead to original outcomes (Youmans and Arciszewski, 2013).

Designers often undertake wide-ranging exploration of problem and solution space in search of creative solutions. Designers continuously restructure and redefine the problem in their attempt to push the design boundaries. When features and constraints in the solution give rise to new criteria that redefine the problem space there is a response driven modification in each of the design spaces. Expansion of problem and solution space is therefore this incremental extension of problem statements and associated solutions (Maher et al., 1996). Expanding the problem and solution space offers the opportunity to find better designs that have so far not been known to exist. Divergent thinking plays an important role in this idea generation and exploration phase of design (Shah et al., 2012).

2.2 Methods to foster expansion of problem and solution space

Alternative ideas emerge when students think like two different people as in address different perspectives. In this regard, collaboration can prove useful. Collaboration has been known to aid in formation of diverse interpretations. During collaboration, there is a mutual construction of knowledge (Cress et al., 2015) when collaborators grapple with the problem, present alternative perspectives, establish a common frame of reference, negotiate meaning, and restructure ideas. To make the

collaborative problem solving process more effective, instructional support in the form of scaffolding is necessary.

Shared representation facilitates access to parts of knowledge of the collaborating group in the form of distributed resources such as a concept map, sketches, and shared worksheets. Shared representation works as a mediating tool with the purpose of engaging and facilitating cognitive processing. Concept maps (CMAP) for instance, helps students to externalize their internal cognitive structure thereby making individual knowledge more explicit. It presents multiple dimensions or perspectives of the picture at once facilitating creative association between ideas via critical reflection and creative thinking among collaborators (Stoyanova and Kommers, 2002).

Divergent idea generation requires designer to uncover new ways of viewing the problem and solution by intuitive associations and systematic variations. Idea stimulating strategies such as adapting, combining, and rearranging, play an important role in imagination by making the manipulation of information more explicit (Eberle, 1972). Such strategies enable designers to span the design space and contemplate employing non-obvious ideas as a solution concept. Additionally, using such strategies help designers to de-fixate from tried and tested solution ideas (Daly et al., 2012). SCAMPER (Eberle, 1972) and design heuristics (Daly et al., 2012) are ways to stimulate restructuring of ideas. Evocative words also have the potential to convey shades of meaning allowing for different interpretations (Lawson and Loke, 1997). Metaphors, synonyms and antonyms (Linsey et al., 2010) can prove to be effective triggers to open up new lines of thought.

3. Flare-Fork pedagogical strategy

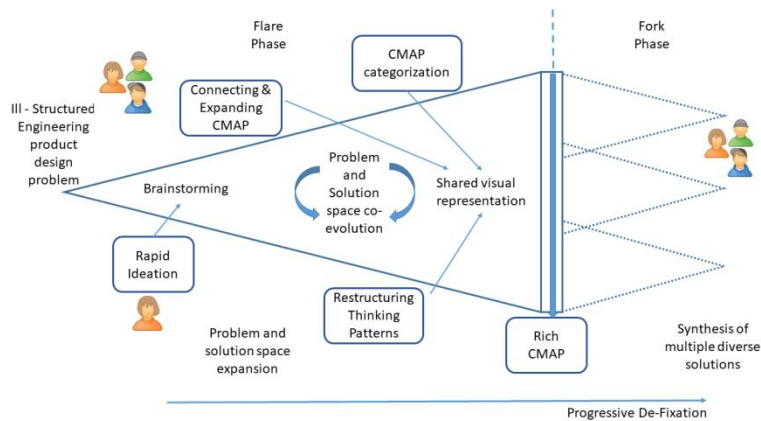


Figure 1. Flare – Fork pedagogical strategy

Flare-fork pedagogy supports novice designers working collaboratively, to build on their opportunistic ideas, expand design boundaries and subsequently identify trade-offs to generate a variety of solutions tailored to fulfill a combination of constraints and assumptions. Flare aspect encourages the designers to progressively expand the problem and solution space by exploring the different aspects of the product design problem at various levels of abstraction in an opportunistic way by brainstorming (non-systematic and multidirectional) (Visser, 2008). A degree of structure is introduced into this opportunistic decomposition (Guindon, 1990) by bringing in the idea of relationship between these sub-systems in the form of concept map (CMAP) creation, followed by categorization. Shared visual representation in the form of collaborative concept map (Stoyanova and Kommers, 2002) brings semantically and conceptually diverse aspects of the design in one place and facilitates simultaneous consideration of all design parts. This enables the designers to make distant connections, which would otherwise have not been apparent, hidden amidst disparate details.

The intervention successively provides students with means of identifying new search cues for exploring the design problem along new lines. While collaboration helps this by bringing in new perspectives from each of the participant, thought transformer strategies (SCAMPER and SynAnt strategy) provide participants with methods to modify existing ideas to get a fresh perspective. At the end of flare, the designers have a rich, connected and categorized concept map covering several aspects of the product design problem.

Fork aspect of the pedagogy encourages the designers to design alternative solutions for the

design problem, considering the conflicting constraints and trade-offs that they can identify. Identifying the conflicts enable designers to view the design problem from multiple levels of abstraction. The conceptual designs so developed is customized for a set of constraints and assumptions. In this paper, we present only the flare intervention design and study its influence in novice design process.

3.1 The flare intervention implementation

The flare intervention begins with a free-wheeling ideation session where the collaborators brainstorm to come up with as many opportunistic ideas regarding the given product design problem, as possible in 5 minutes. This activity is non-collaborative and allows individual designers to produce as many ideas as possible (Stroebe and Diehl, 1994). Idea categories that act as anchor points are used as scaffolds to help trigger thoughts. These anchor points are synthesized from design literature and engineering product specifications. They include functionality, shape/structure (Gero and Kannengiesser, 2004), requirements for working, principle of operation, analogy (Lawson and Loke, 1997), questions (Eris, 2003), and sketches (Lawson and Loke, 1997) as a few exemplar idea categories.

The opportunistic ideas are then connected collaboratively by relating, combining and interlinking them to form a concept map. Addition of new ideas are welcomed so as to make the concept map rich. Scaffolding is provided in the form of SCAMPER (substitute, combine, adjust, modify, put to other uses, eliminate, reverse/rearrange an idea). SCAMPER helps participants transform ideas, change context or perspective thereby triggering imaginative exploration of solution space (Eberle, 1972). Collaborative concept map construction continues until all the ideas from brainstorming are used and no new ideas or links are forthcoming. The collaborative activity is orchestrated by explicit instructions to interact by questioning, explaining, identifying relationship, elaborating and critiquing ideas while constructing the concept map.

SynAnt is a semantic analogy strategy that enables designers to get new search cues for exploring the problem and solution space. Collaborators are prompted to revisit the problem statement, extract key words and search for synonyms and antonyms of these words individually. These synonyms and antonyms (Linsey et al., 2010) are then used to search the World Wide Web for additional inputs informing new ideas. Individual worksheet is provided with task breakdown to orchestrate this activity. Subsequently, participants collaboratively modify the concept map with these new ideas. Towards the end of flare intervention, participants categorize and label the concepts in the concept map. This helps them take stock of the problem and solution space and is an important precursor for consolidation of the ideas in the concept map into conceptual designs. This collaborative activity is orchestrated by explicit instructions to come to a consensus about categorizing concepts in the concept map.

4. Conjecture map

Conjecture mapping is a systematic method to clearly state how the salient features of the learning environment is expected to produce the desired outcomes (Sandoval, 2014). We use conjecture mapping as a means of identifying the pedagogical support for our intervention. Our design and theoretical conjectures were informed from literature on typical processes that guide problem and solution space exploration. Daly et al. (Daly et al., 2014), came up with a rich set of cognitive aspects of creativity that facilitates divergent and convergent thinking towards enhancing creative exploration of the design problem. We used a subset of the cognitive aspects of creativity as the basis for generating design conjectures and theoretical conjectures for our intervention.

Our design conjectures (from Figure 2) are:

- If learners engage in collaborative CMAP creation with CMAP creation tool, then combining and interlinking of far off ideas will occur.
- If learners engage in ideation with semantic analogy thought transformer strategy individually then generating new keywords and ideas will occur.
- If learners engage in collaborative CMAP creation with CMAP creation tool collaboratively, then emergence of partial-solutions from opportunistic ideas will occur.
- If learners engage in transformation and modification of ideas with SCAMPER thought transformer scaffold individually, then reorganization of ideas by changing context or perspective will occur.

Our theoretical conjecture is: If reorganization of ideas by changing context or perspective, combining and interlinking of far off ideas, emergence of partial-solutions from opportunistic ideas and generation of new keywords and ideas occur, then it will lead to learning to expand problem space and solution space.

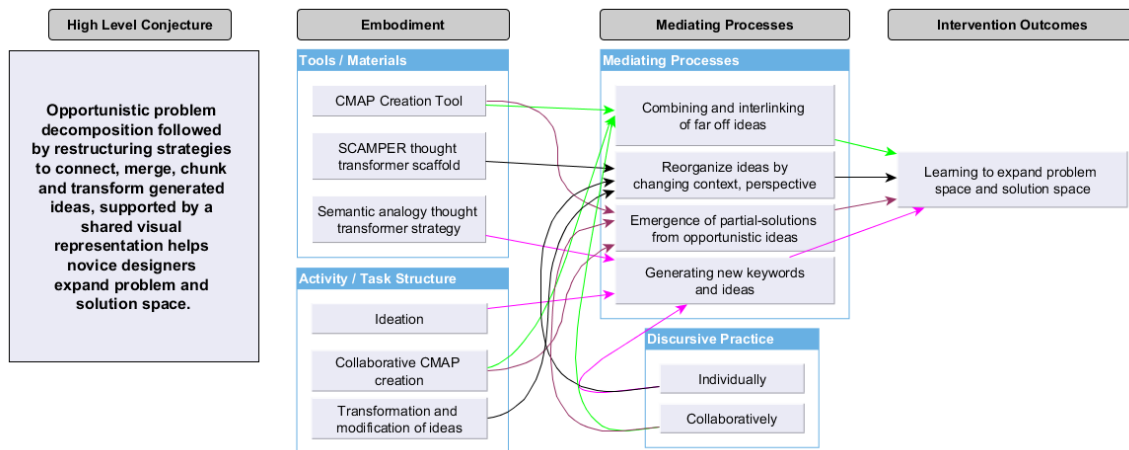


Figure 2. Conjecture map

5. Research Methodology

5.1 Research Questions

The research questions that guide our study are:

1. How do students explore problem and solution space during design problem solving?
2. In what ways did the different features of the intervention support the expansion of problem and solution space?

5.2 Participants

We chose students in their final year undergraduate degree or pursuing their master's degree, as participants for our intervention. At this level, students have already worked on at least one semester long project, but are still not expert designers and are learning new design strategies. Participants were from diverse backgrounds such as computer science, electronics, metallurgical engineering, and chemical engineering. This was based on the idea that teams with diverse background bring in different perspectives to a design problem. We conducted three studies wherein we posed the same engineering design problem to a dyad (group 1) and two triads (group 2 and group 3).

5.3 Procedure

To begin with, we gave the participants the problem of designing a currency cleaner couched as a scenario description. We then outlined the activities, rules and scaffolds in the intervention. We used post-it notes for writing ideas and a chart paper to develop the concept map. We provided a worksheet for SynAnt activity along with access to World Wide Web for focused search.

We video recorded the whole intervention with two video cameras such that one recorder captured the wide-angle view of the actions and behavior of the group while the other camera was a close-up of the map under development. We also audio recorded the whole proceeding. We made regular unstructured observations for recording activities of the participants during the intervention. The audio recording was then transcribed for analysis.

5.4 Data analysis

To investigate how students explore problem and solution space during design problem solving (theoretical conjecture), we used an analytical framework for design process elements and stages framework (Mehalik and Schunn, 2006). The framework is a meta perspective derived from several separate empirical studies on aspects of 'good design'. It identifies 15 elements / stages (summarized in Table 1) of documented observable design process elements that are representative of design activities. The 15 design elements span exploration of problem and solution space. The problem space is the task environment and comprises of initial problem state and problem requirements. The solution space includes the design solutions and foundations for evaluating the requirements (Hay et al., 2016). All possible intermediate states can be considered the bridge (Dorst and Cross, 2001) that identifies problem and solution pair. We used this criteria to categorize the 15 elements into problem space, solution space, intermediate bridge space and others (belonging to stages beyond conceptual design). We chose the unit of analysis to be one conversation turn in the transcription of the three studies. However, to establish context we have referred to sentences immediately before or after the sentence under consideration. We used thematic analysis as a method for identifying, analyzing and reporting patterns corresponding to the codes in our analytical framework, within the data (Braun and Clarke, 2006). Additionally, we analyzed the collaborative concept map generated by students to identify student generated categories and links between and within these categories.

To investigate in what ways the different features of the intervention support expansion of problem and solution space, we used the design conjectures as the analytical framework. The design conjectures are our hypothesis of how our intervention works and by using them as analysis framework, we look for instances that demonstrate the presence of the predicted activity and outcome patterns. We used 5 seconds of video data as the unit of analysis. We localized episodes when students used the particular feature, and scrutinized the video for evidence of productive usage of the feature. We focused on speech, gestures, what was being written and where were the collaborators looking while scrutinizing the video. The above data streams along with the concept map visual artifact provide insights into the view point of each collaborator about the design and ways by which the external visual representation mediated the design process.

Table 1

Common design process elements / stages adapted from (Mehalik and Schunn, 2006)

Category	Definition
Problem representation (EPR)	Framing of a design task including goals, issue, artifact that needs to be analyzed, synthesized, investigated or constructed
Scope of constraints (ESC)	Constraints limit how a design can fulfill goals within problem frame. Designer needs to explore how constraints are affecting the design.
User perspectives (EUP)	Capturing various aspects of requirements, needs of the users.
Evaluate design alternatives (EDA)	Designer's actions to use a framework of performance criteria (goals and constraints) to search and evaluate potential design solutions.
Use functional decomposition (FD)	Breaking down design into several detailed aspects to investigate how the design performs, interacts and contributes to overall functionality
Explore engineering facts (EEF)	Exploring specific knowledge about some property of an aspect of a design. Includes common principles.
Examine existing designs / artifacts (EEDA)	Either look at past designs or existing solution ideas to improve on them in various design dimensions.
Conduct failure analysis (CFA)	Gathering knowledge associated with what produces a failure i.e. when design falls short of goals or performance expectations.
Redefine constraints (RC)	Designer defines a constraint to achieve an original goal.
Validate assumptions and constraints (VAC)	Ensuring that the representation of the user's or other stakeholder's expectations for the design appear to be met.
Graphical representation (EG)	Visualizing details of design to explore overall configuration.
Issues of measurement (EM)	Quantitative information gathering relating to some aspect of design.
Build normative model (BNM)	The ideal, optimal outcomes for the design indicative of designers attempt at formalizing desired outcomes.
Design methodology (FDM)	The documented interactive / recursive / iterative design process.
Reflection on design (ERD)	Designers reflect on their own process that they use to achieve goal.

6. Findings

6.1 Students expansion of problem and solution space

From the graph (Figure 3) depicting group 1 students' design progression with time, we can see that:

- During the design process, student's activities can be mapped to 13 of 15 good design activities.
- Students constantly switch between problem space (green) and solution space (red) with reflection and modification in constraints (brown). This indicates the co-evolution of problem and solution space during design process.

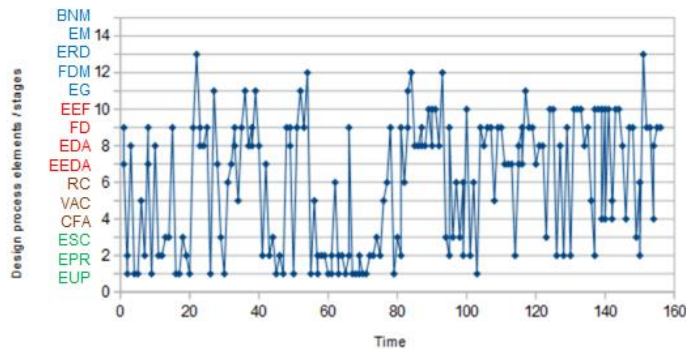


Figure 3. Students design progression depicting exploration of problem and solution space by group 1.



Figure 4. Collaborative concept map generated by group 1 at the end of flare phase

Figure 4 is the expanded problem space and solution space (of group 1) represented as a concept map which is the output of flare of our flare-fork pedagogy. The figure shows student generated categorization and interlinking. The categories that the students identified were usage, good habits, alternate system level solution, function, technology, form design and users. The students connected technology and users to form design thereby forming a relationship between the three disparate categories.

From Figure 5 constructed from an episode in the narrative, we can see how students moved from problem space to solution space, did some analysis on the solution and came up with new constraints and requirements and identified a new problem with a new solution. This episode shows that students expanded problem and solution space during the design process. To summarize, (a) students move back and forth between problem and solution space frequently during the design process, (b) the traversal is via intermediate bridges that link problem and solutions, (c) linking of disparate concepts and categories support the traversal between problem and solution space.

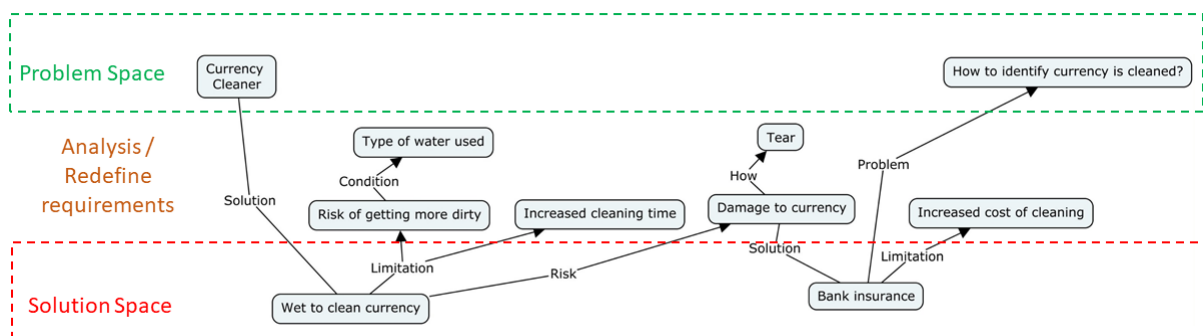


Figure 5. Scenario from narrative depicting expansion of problem and solution space

6.2 Testing design conjectures

6.2.1 Collaborative concept map creation activity promotes interlinking and combination of far off ideas

We found multiple episodes where the collaborators linked different ideas to come up with a new idea. Such links were seen within student identified category as well as across disparate student identified categories. The excerpt extracted from the transcript of group 1, supported by the image sequence in Figure 6 illustrates the linking within and between student generated categories. The hand movements of the collaborators in this excerpt show how the shared external representation mediated the linking of ideas within categories (in this example, under the category of technology solution, linking brush from mechanical domain to UV from electrical domain to form a cleaning mechanism) and linking between categories (in this example category of technology to category of form).

S1 - Or we could combine these two (Points to 'note counter with UV tech' and 'Designing some equipment with brush' as seen from hand movement in frames 1 and 2 of figure 6)

S2 - Ya we can but it would be a bigger equipment (Gesture seen in frame 3 of figure 6)

S1 - But you don't know no, let's see if we come up with something very

S2 - Sleek? (Gesture seen in frame 4 of figure 6)

S1 - It takes care of physical dust particles

S2 - Brush like

S1 - as well as sterilizing it (frame 5 of figure 6)



Figure 6. Image sequence depicting role of shared visual representation (CMAP) during product design

In the above discussion, two diverse aspects of the product viz. the technology to be used and the form factor of the product as identified by the students, is being linked. Observing the gestures during the interaction in the 5 frames of Figure 6, we can see that shared visual representation seems to have facilitated the collaborative linking activity verifying our design conjecture.

6.2.2 Semantic analogy thought transformer usage leads to generating new keywords and ideas

We found episodes where looking at synonyms and antonyms of keywords extracted from the problem, enabled participant to generate new keywords and viewing an idea in a different perspective. An instance of this new perspective is seen in the excerpt below taken from transcript of group 1.

S1: "Currency notes synonyms are cash, roll. So I understood that you know it may not be like my currency is as plain as this. She could be dealing with rolled out or deformed currency also at times so maybe this the form of your input would, just increase the complexity. So again this actually helps in improvising this whole design. What if I just put some money the system could also maybe straighten it up for me and process it? So some adds on straighten up feature".

When the students changed their perspective regarding form of currency, it led to the generation of new requirement - a de-wrinkling system. This new requirement followed by exploration of solutions to satisfy this requirement enabled students to expand problem and solution space. Other instances of productive semantic analogy usage operationalized as SynAnt were (a) 'bleach' as synonym to 'decontaminate' giving rise to bleaching and reprinting clean currency, (b) 'tarnish' as antonym to 'clean' leading to idea of coating currency with protective layer like a 'varnish'. The above instances also indicate flexibility in solution space exploration as a consequence of new keywords triggering reorganization of ideas in new contexts.

6.2.3 Collaborative interlinking of ideas represented in shared concept map facilitates emergence of partial-solutions from opportunistic ideas

Students build on one another's ideas to elaborate the design while engaging in collaborative concept map creation activity. Potential sub-solutions emerge during this elaboration process. In the excerpt below (extracted from transcript of group 1), S1 suggests an opportunistic idea of using a note counting system. We identify this idea as opportunistic because it was not a result of a systematic breakdown of

the problem. Rather, S1 used her prior knowledge and fit it into the present design context. S2 recognizes the relevance of this idea to the design problem and links it with technology idea. The shift in focus triggered by S1's opportunistic idea led identification of a requirement and emergence of a partial-solution in the form of using a card reader - note counter kind of system. The expansion of solution space is due to inclusion of similar equipment in search space.

S1: "Can we do something like this note counting thing". (S1 adds the 'note counter with UV tech' idea to the concept map)

S2: "Or you could just put it here and join these two together" (S2 joins idea of 'note counter with UV tech' and 'Designing some equipment with brush' on the concept map)

S2: "So I had seen this card reader. It takes in a card and inside there is an OCR. It reads it and stacks it on one side. So something similar we can have for a note counter, just have to put the note. It cleans with the brush and the UV as well as counts and puts it on the other side".

We could however not find evidence for the final set of design and theoretical conjectures on influence of SCAMPER usage on reorganizing ideas by changing context or perspective, eventually leading to expansion of problem and solution space.

7. Discussion

Our first research question was 'How do students explore problem and solution space during design problem solving'? We found that students frequently move back and forth between problem and solution space during the design process via intermediate bridges that pair problem and solutions. Links within and between categories in the design space support the traversal between problem and solution space. Of the 15 themes from common design process elements / stages (Mehalik and Schunn, 2006), we found that the students' design process covered 13 process elements / stages pertaining to the exploration of problem and solution space. The two themes not addressed were 'exploring issues of measurement' and 'building normative model'. While the issues of measurement become vital in the detailed design, it may play a relatively smaller role in conceptual design stage and assume a more qualitative form such as speculating if the proposed design alternative would help meet the goal.

The theme related to building normative model implies articulation of ideal outcomes without constraints or limitations by the designers for the given design problem. Since the students followed an opportunistic decomposition path, they predominantly focused on mental simulation of practical scenarios to explore the problem representation and generate solution alternatives. We conjecture that in such context, articulating a normative model would require scaffolding.

The most explored process elements are the exploration of user perspectives, exploration of problem representation and functional decomposition. This is expected as during the initial conceptual design stage, task clarification, requirement understanding, establishing function structures and search for solution concepts and principles sets the solution path and expectations from the design (Pahl and Beitz, 2013). Overall, the intervention seems to support expansion of problem and solution space.

The second research question was 'In what ways did the different features of the intervention support the expansion of problem and solution space'? We found support for three of four design conjectures. This is important because it demonstrates that the intervention in conjunction with orchestration of the whole intervention, support expansion of problem and solution space. There was however a lack of evidence showing that transformation of ideas using SCAMPER thought transformer, influences reorganization of ideas due to changing perspective. We surmise that in the present form, the orchestration of SCAMPER scaffold in the intervention is weak. The usage of SCAMPER thought transformer is kept optional and the mediator occasionally prompts the collaborators to try applying the SCAMPER heuristics to their ideas. We observed that students did not immediately grasp how to implement SCAMPER heuristics to their ideas and asked for examples. This finding aligns with Daly et al (Daly et al., 2012) who commented on the difficulty of applying SCAMPER guidelines. The theoretical conjecture that far off linking can give rise to new constraints and requirements thereby expanding the problem and solution space, could not be verified as the students did not continue investigating this line of thought further.

Going further, we would like to design a more effective orchestration of SCAMPER thought transformer so as to exploit its potential to trigger divergent thinking among students. This is a preliminary study with a small sample size of only three groups and 8 participants, which is a limitation. However the goal of this study is not to generalize, but to understand the role of the various features of

the intervention, in order to refine our learning design for fostering problem space and solution space expansion as a means of avoiding design fixation. One of the objectives of this preliminary study was to gain insights about technology support that will inform the design of a CSCL environment. During the study we found that a digital concept map with provision for simultaneous manipulation by collaborators using tangible technologies, can prove to be a valuable support for engaging in co-construction. Integrating private workspace and joint workspace can help collaborators seamlessly switch between them and have smoother collaboration.

References

- Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *J. Engg. Edu.*, 96(4), 359-379.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Cress, U., Stahl, G., Ludvigsen, S., & Law, N. (2015). The core features of CSCL: Social situation, collaborative knowledge processes and their design. *International Journal of Computer-Supported Collaborative Learning*, 10(2), 109-116.
- Crilly, N., & Cardoso, C. (2017). Where next for research on fixation, inspiration and creativity in design?. *Design Studies*, 50, 1-38.
- Daly, S. R., Yilmaz, S., Christian, J. L., Seifert, C. M., & Gonzalez, R. (2012). Design heuristics in engineering concept generation. *Journal of Engineering Education*, 101(4), 601-629.
- Daly, S. R., Mosyjowski, E. A., & Seifert, C. M. (2014). Teaching creativity in engineering courses. *Journal of Engineering Education*, 103(3), 417-449.
- Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. *Design studies*, 22(5), 425-437.
- Eberle, R. F. (1972). Developing imagination through scamper. *Journal of Creative Behavior*, 6(3), 199-203.
- Eris, Ö. (2003). Asking generative design questions: a fundamental cognitive mechanism in design thinking. In *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm*.
- Gero, J. S., & Kannengiesser, U. (2004). The situated function–behaviour–structure framework. *Design studies*, 25(4), 373-391.
- Guindon, R. (1990). Designing the design process: Exploiting opportunistic thoughts. *Human-Computer Interaction*, 5(2), 305-344.
- Hay, L., McTeague, C., Duffy, A. H., Pidgeon, L. M., Vuletic, T., & Greal, M. (2017). A systematic review of protocol studies on conceptual design cognition. In *Design Computing and Cognition'16* (pp. 135-153). Springer, Cham.
- Lawson, B., & Loke, S. M. (1997). Computers, words and pictures. *Design studies*, 18(2), 171-183.
- Linsey, J. S., Tseng, I., Fu, K., Cagan, J., Wood, K. L., & Schunn, C. (2010). A study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design*, 132(4), 041003.
- Maher, M. L., Poon, J., & Boulanger, S. (1996). Formalising design exploration as co-evolution. In *Advances in formal design methods for CAD* (pp. 3-30). Springer, Boston, MA.
- Mehalik, M., & Schunn, C. (2007). What constitutes good design? A review of empirical studies of design processes. *International Journal of Engineering Education*, 22(3), 519.
- Pahl, G., & Beitz, W. (2013). Engineering design: a systematic approach. *Springer Science & Business Media*.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, 23(1), 18-36.
- Shah, J. J., Millsap, R. E., Woodward, J., & Smith, S. M. (2012). Applied Tests of Design Skills—Part 1: Divergent Thinking. *Journal of Mechanical Design*, 134(2), 021005.
- Stoyanova, N., & Kommers, P. (2002). Concept mapping as a medium of shared cognition in computer-supported collaborative problem solving. *Journal of Interactive Learning Research*, 13(1), 111.
- Stroebe, W., & Diehl, M. (1994). Why groups are less effective than their members: on productivity losses in idea-generating groups. *European review of social psychology*, 5(1), 271-303.
- Visser, W. (2009). Design: one, but in different forms. *Design studies*, 30(3), 187-223.
- Youmans, R. J., & Arciszewski, T. (2014). Design fixation: Classifications and modern methods of prevention. *AI EDAM*, 28(2), 129-137.

Towards an Integrated Framework of Group Awareness Support for Collaborative Learning in Social Media

Lisa OLLESCH^{a*}, Sven HEIMBUCH^a & Daniel BODEMER^a

^a*Media-based Knowledge Construction, University of Duisburg-Essen, Germany*

*lisa.ollesch@uni-due.de

Abstract: In computer-supported collaborative learning, group awareness tools have been shown to be helpful regarding learning processes and outcomes. Previous research has focused on the support via cognitive and behavioural group awareness information, largely neglecting emotional group awareness information and often investigating these three aspects separately. To support large social media groups such as wiki communities, integrating different types of group awareness (GA) information may yield benefits, since these communities encounter several challenges. Although jointly presenting different GA information is assumed to be advantageous for collaborative learning, GA interaction effects including personality traits are still largely unexplored. In order to close this research gap, an integrated framework is proposed, which enables the systematic empirical investigation of (interaction) effects of different types of GA information on behavioural, cognitive, and emotional challenges in computer-supported collaborative learning, with a focus on wikis.

Keywords: Computer-supported collaborative learning, group awareness, social media, wikis

1. Introduction

Humans are inherently social creatures, even in contexts that appear to be primarily related to individual learning. Thus, research has highlighted the central role of social factors in achieving academic success (Wilcox, Winn, & Fyvie-Gauld, 2005). The increasing digital networking and importance of online social media in all areas of life, including individual and collaborative learning, entails new challenges for learners. Over almost 30 years of history, computer-supported collaborative learning (CSCL) has contributed significant research and innovative tools for facilitating learning processes (e.g., Miller & Hadwin, 2015). There seems to be a consensus that CSCL offers wide-ranging potentials for increasing the effectiveness and efficiency of learning and teaching processes, regardless of participants' time zone and location. Results range from positive effects on individual learning, through facilitating dyadic and small group collaboration up to more effective designs of collaborative tasks for large groups in massive open online courses (e.g., Jung & Lee, 2018). In CSCL research, social interaction is not solely observed as a method; Dillenbourg, Järvelä, and Fischer (2009) define it as the essence of cognition and as "the heart of CSCL" (p. 8). One example of widely used CSCL platforms are wikis. These are social media platforms with few social affordances to enable collaboration and therefore social interaction at virtually any point in time and between anyone (Chen, Jang, & Chen, 2015). Although such platforms offer new possibilities, they are also associated with difficulties, as computer-supported collaboration is not inherently advantageous. Learners must cope with further cognitive and social tasks beyond the requirements of individual learning by using digital media (Zheng, Niiya, & Warschauer, 2015), which are caused by the setting itself and the interaction of the learning material with the collaborative setting. To support learners coping with diverse requirements of learning with digital media, tools were designed and experimentally tested that combine established methods of support for individual learning processes. Such *Group Awareness Tools* (GATs) collect, transform, and present information about the learning partners (Bodemer, Janssen, & Schnaubert, 2018). In the following sections we provide an overview of the potentials of state-of-the-art CSCL research, as well as of some important challenges faced by learners in this field. Building upon existing CSCL and GAT frameworks, our objective is to introduce an integrated theoretical framework for GATs, which can serve as a basis for future studies on

the interplay of different types of GA information. Moreover, it intends to guide teachers and instructional designers in the reflective design of formal and informal learning environments that consider behavioural, cognitive and emotional aspects of social learning.

2. Potentials of Collaborative Learning and CSCL

Collaborative learning offers opportunities for elaborated learning processes and critical thinking (Johnson, Johnson, & Smith, 2000), such as a larger and broader knowledge base in groups. Every learner has a different amount of prior knowledge and different perspectives (Bell, 2004) that can benefit collaborative activities. During such activities it is possible to exchange heterogeneous knowledge, opinions and hypotheses, which offers a chance to recognise misconceptions in one's own thinking and to harmonise distributions of diverging knowledge. Furthermore, collaborative learning can also help to recapitulate knowledge through mutual explanation (Webb, 1991). In addition to increasing chances of collaborative learning, CSCL environments can further enhance the effectiveness and efficiency of learning processes. The continuous availability of learning materials and a relatively low threshold for collaborating without the need for face-to-face meetings can be highly advantageous (Walther & Bunz, 2005). As one example for a CSCL environment, wikis enable users to create socially shared artefacts as well as to share their knowledge on two rather distinct levels that relate to each other (e.g., Choy & Ng, 2007), in forms of firstly the article as a collaboratively created product and secondly corresponding talk pages for discussing article-related topics. Compared to other knowledge construction platforms evaluated in educational contexts, wikis enable users to perform in a way that influences the whole environment (e.g., Kimmerle, Moskaliuk, Oeberst, & Cress, 2015), which creates fertile ground for the multi-level occurrence of controversies and socio-cognitive conflicts (Bell, 2004). Controversies can be constructive when based on the exchange of contrasting viewpoints on a specific topic, which provides opportunities to trigger learning processes and foster higher learning outcomes (Johnson et al., 2000). Moreover, they might induce socio-cognitive conflicts within learners as well as between learners and wikis as social systems. This can be beneficial by triggering equilibration processes of accommodation and assimilation of new knowledge artefacts into one's individual cognitive systems (Piaget, 1977). The combination of computer support and collaborative learning not only promotes these potentials but also poses challenges for learners.

3. Challenges in Collaborative Learning and CSCL

Effective learning processes and outcomes are not automatically produced by enabling computer-supported collaboration. Aspects of both computer support and collaboration, and especially in combination, bring their own challenges to learners (Zheng, Niiya, & Warschauer, 2015). Based on the current state of research, we identified three main challenges which serve as an indication of dominant challenges in CSCL and rather refer to large social media learning communities such as wikis. Addressing these challenges is essential to the success of CSCL.

3.1 Behavioural Challenge: Contributing

The lack of behavioural motivation is often considered to be one of the greatest difficulties in online communities. However, the willingness to share knowledge is a prerequisite to CSCL's success. The motivation to contribute or participate is not always present, especially in large social media groups where free riding and social loafing represent more common risks in collaboration (see Kimmerle & Cress, 2008). Following a series of wiki studies conducted at our lab, we found that many undergraduate students participate in joint collaborative knowledge construction and learning activities on wiki talk pages when instructed to do so, offering potentials for elevated wiki quality and improved learning processes. Otherwise, collaborators tend to show cooperative behaviours instead of engaging in valuable social interactions (Heimbuch, Ollesch, & Bodemer, 2018). This could be because individuals' contributions to wiki discussions are not necessarily visible, which highlights the high value of increasing participatory motivation in the social interaction space for successful wiki learning.

3.2 Cognitive Challenge: Dealing with Meaningful Content

Meaningful interactions between wiki collaborators are important in addition to CSCL settings' behavioural requirements, and difficulties achieving meaningful collaboration may occur without required motivation or skills among group members. Such difficulties can be rooted in the lack of understanding others' contributions that can manifest a cognitive challenge to learners. These problems arise when group members fail to pay sufficient attention to individual contributions as well as when such contributions are not sufficiently discussed (Näykki, Järvelä, Kirschner, & Järvenoja, 2014). Although this challenge applies to all CSCL domains, it is obvious that contributions conducive to cognitive learning are especially less simple to identify in larger communities which feature copious content. This can occur due to information overload, an unavoidable reality of larger online discussion forums growing to include hundreds to thousands of contributions (Buder, Schwind, Rudat, & Bodemer, 2015), or wiki talk pages that often lack salience of the aforementioned controversies at first glance (Heimbuch & Bodemer, 2017). Due to the limitations of working memory capacities, this lack means that those meaningful contributions are simply not perceived and therefore not read (see Bagherian & Thorngate, 2000). These processes of collaborative knowledge construction can be difficult and challenging by causing frustration during the learning process (Capdeferro & Romero, 2012), which emphasises the necessity to highlight cognitively relevant contributions and thus facilitate knowledge acquisition.

3.3 Emotional Challenge: Maintaining a Positive Group Climate

The motivation and skillsets to maintain a positive group climate represent another essential aspect that is often neglected in the context of designing and evaluating CSCL settings. Group formation is viewed as a prerequisite to successful collaborative learning (Kirschner & Erkens, 2013) since relational issues can strongly influence interaction, task engagement, and learning (Näykki et al., 2014). If negatively balanced emotions or negatively connotated utterances occur during conflicts, group members become less motivated to solve their assigned tasks and tend to demonstrate inferior performances (Ayoko, Callan, & Härtel, 2008). Therefore, CSCL environments should be designed to be more "sociable" for their users (Kreijns, Kirschner, & Vermeulen, 2013). This emotional challenge is also inherently relevant for wikis, such as on Wikipedia where so-called "edit wars" are likely to occur and difficult to solve when many users with contradicting viewpoints attempt to work on the same knowledge artefacts (Yasseri, Sumi, Rung, Kornai, & Kertész, 2012). This underlines the necessity to support wiki users in solving such socio-emotional issues to help a group of individuals transform into a team.

4. Supporting Learning Processes in CSCL

The many degrees of freedom offered by CSCL lead to users perceiving a high degree of autonomy, which positively influence the individual learning motivation. Nevertheless, this freedom demands a high degree of self-regulation to overcome the aforementioned challenges (see Järvelä & Hadwin, 2013). Moreover, without further support this freedom offers only minimal structure and risks learners experiencing themselves as less competent, which in turn can negatively impact the learning motivation (Rienties et al., 2012). It is thus desirable to structure collaborative learning processes to promote the experience of competence. CSCL research offers different means of support that vary regarding their degree of coercion. For example, collaboration scripts can improve the effectiveness of collaborative learning by providing explicit guidance concerning the manner in which people should form groups, interact with each other and solve group tasks (Kollar, Fischer, & Hesse, 2006). However, this entails an often-discussed risk of overscripting. An alternative regards implicit structuring which provides less coercive guidance and is intended to enable desirable behaviour through visual context stimuli in order to achieve more effective collaboration (Janssen & Bodemer, 2013). Identifying the more effective guidance approach represents a heatedly discussed topic in the CSCL community (Wise & Schwarz, 2017). This controversy will not be discussed in this study due to a lack of universal consensus as well as because this article focuses on GAT support, but instead we define Group Awareness and different types of GATs.

Group Awareness (GA) can be loosely defined as any information about the group possessed by a learner, such as knowledge about activities, skills of group members, or social activities within the group (Janssen, Erkens, & Kirschner, 2011). Prevalent GA is often regarded as an important prerequisite to meaningful collaborations in CSCL but cannot be taken for granted (Bodemer et al., 2018). GATs can be used to complement individuals' GA by providing information about specific properties of group (members), e.g., regarding their participation, knowledge, or feelings. This information can be collected, transformed, and presented in different ways (Bodemer et al., 2018). Although GATs are expected to benefit through stimulating productive interaction activities (Miller & Hadwin, 2015), behavioural, cognitive, and emotional challenges are not addressed by one single GA component due to the significant complexity of social interactions. When thinking about social interaction, it can be differentiated between a content space and a relational space of collaboration (Slof, Erkens, Kirschner, Jaspers, & Janssen, 2010). The content space contains the problem to be solved and interactions in this space refer to the task itself. Learners discuss ideas as well as opinions and thus gain a deeper understanding of the task in order to solve it. The content space contains not only cognitive but also metacognitive activities such as resolution strategies for the task (Janssen & Bodemer, 2013). Interactions in the relational space are activities concerning the social dimension of collaborations (Slof et al., 2010), which are important for the functioning of cognitive activity exchanges in the content space. Here, group members create a collective understanding of the discussed concepts in the content space. To become successful collaborators, both task fulfilment (content space) and team functioning (relational space) are essential in order for randomly assigned group members to become effective team players (Fransen, Weinberger, & Kirschner, 2013). The effects of different types of GA information on the two spaces are further examined by providing examples in the following paragraphs (see Figure 1).

Behavioural GATs (in the community also labelled as activity/(socio-)behavioural GATs) address the behavioural challenge by presenting the collaborators' activities, and they thus serve as a source of motivation for providing contributions in the social interaction space in general (Lin, Tsai, Hsu, & Chang, 2019; Kimmmerle & Cress, 2008; see Figure 1a). Behavioural GA information has promising potential to increase participation rates in terms of motivational processes. This can be achieved by visually juxtaposing individuals' contributions against the group's contributions or average participation (Kimmmerle & Cress, 2008). The possibility of self-presentation is crucial here, however single applications of behavioural GATs do not necessarily lead to increased cognitive performance as measured by means of message and project quality (Lin et al., 2019). Although there are already initial wiki approaches, e.g., to supplement MediaWiki with participation monitoring tools (Popescu, Anca, & Udriştoiu, 2014), they need to be systematically evaluated.

Cognitive GATs (in the community also labelled as knowledge GATs) provide content-related information about group members, such as their knowledge or opinions. These tools are promising for tackling cognitive aspects of learning (Janssen & Bodemer, 2013) and mainly address the cognitive challenge by facilitating the navigation and selection of meaningful content. Moreover, the presentation of partner knowledge facilitates grounding and partner modelling in the content space of social interactions (Bodemer et al., 2018). There is additionally potential to reduce unnecessary extraneous cognitive load (Chandler & Sweller, 1991) induced by the collaborative learning setting. Taking large learning environments as an example, cognitive GA information in the form of visual markers help learners to focus on meaningful content on large wiki talk pages in order to identify relevant controversies (Heimbuch & Bodemer, 2017; see Figure 1b; *blue markers label controversies in general, green markers stand for solved, and red markers for unresolved controversies*) or high-quality contributions in online forums (Buder et al., 2015), which could also be applied to the visualisation of collaborators' expertise/knowledge level in future studies.

Emotional GATs (in the community also labelled as social/(socio-)emotional GATs) are helpful tools to facilitate joint emotion regulation in the relational space of social interactions, to enhance mutual transactivity, and to create a positive group climate by increasing group members' awareness of other members' feelings (Eligio, Ainsworth, & Crook, 2012; see Figure 1c). Educational psychology currently predominantly focuses on cognitive and behavioural support. Socio-emotional issues are treated with a much lower priority in instructional designs, and to our knowledge there is no empirically tested tool that deals with joint emotion regulation in the field of wikis. Such tools could help to identify unfriendly posts on wiki talk pages or highlight self-assessed emotional states of wiki group members to alert the group to emotional grievances. These represent initial design impulses, and a deeper investigation into the effects of emotional GA information on the emotional challenge is necessary.

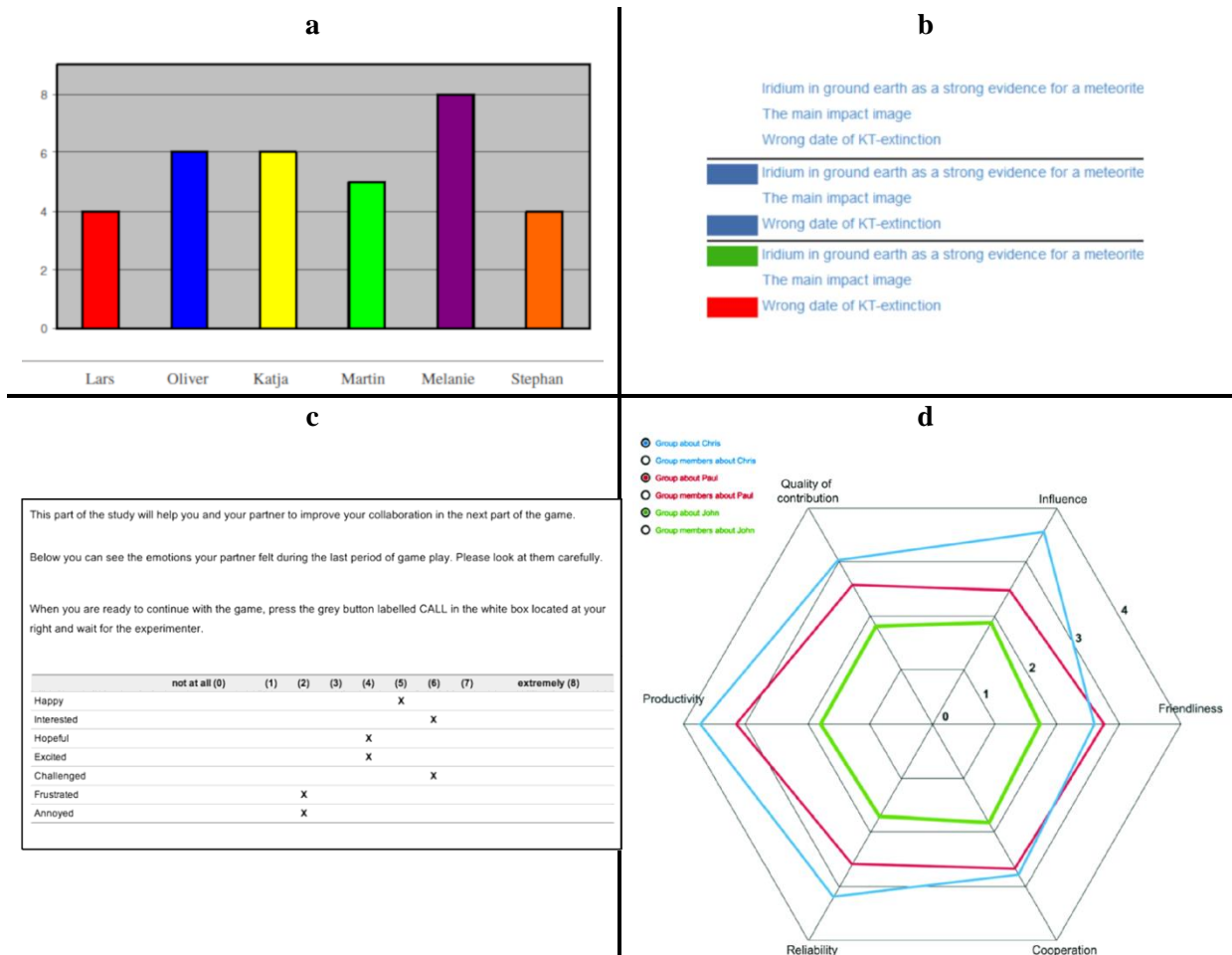


Figure 1. Examples of GAT support: a) Behavioural GAT: Kimmerle & Cress (2008); b) Cognitive GAT: Heimbuch and Bodemer (2017); c) Emotional GAT: Eligio et al. (2012); d) Combination of several tool aspects: Phielix, Prins, Kirschner, Erkens, and Jaspers (2011).

Lin, Mai, and Lai (2015) criticise that only a few studies examine the differences and overlaps between behavioural and emotional GA (social-context awareness) as well as cognitive GA (knowledge-context awareness) information. Their long-term study shows that while social-context awareness stimulates more quantitative peer interactions, knowledge-context awareness unexpectedly does not necessarily increase the quality of messages. They conclude that a combination of different types of GA information may be more effective. We agree that cognitive, behavioural, and emotional GA information may be required in order to achieve effective group performance (e.g., high wiki article quality). Consequently, GATs that provide more than one type of GA information are necessary such as the seldom exception of the RADAR tool (Phielix et al., 2011; see Figure 1d). It is one of the few GATs that reflect different aspects of collaboration and group functioning. This tool presents six self-assessed as well as peer-assessed group dimensions: influence, friendliness, cooperation, reliability, productivity, and quality of contribution. It could be shown that social performance such as group satisfaction is positively affected by communicating this information back to the group, however no effect on cognitive processes and learning outcomes could be observed (Phielix et al., 2011). Analogue to the GA information differentiation that we propose with this framework, influence and quality of contributions can be classified as cognitive GA information, friendliness and reliability as emotional GA information, and cooperation and productivity as behavioural GA information. Although the results of the RADAR tool are promising for GAT research, they have only been examined for smaller groups. In general, our literature review demonstrated that a significant portion of GAT research is not focused on social media communities like wikis and that research such as of Heimbuch and Bodemer (2017) represents a rare exception. It is therefore important to investigate how larger social media communities can be supported with combined GA information as well as what types of group awareness information

are most relevant in such settings. Moreover, the investigations regarding the RADAR tool only allow speculation about the reasons for the non-significant cognitive dependent variables and the different tool functions since the six dimensions have not been examined separately and systematically. Thus, a systematic investigation of different GA information is missing and necessary in GAT research.

5. An Integrated Framework of GAT Support

This section addresses existing conceptual considerations in the field of CSCL and GATs as well as how the proposed framework represents an extension. Kreijns, Kirschner, and Jochems (2003) presented a differentiated view of social and cognitive processes in CSCL, albeit without including GA but regarding the pitfalls in CSCL. The first discussed pitfall is to take social interaction for granted without stimulating it. In addition, the authors criticise the second pitfall or the fact that in many cases, instructors limit their actions to the content space of social interaction. Thus, it has been concluded that collaboration can only be successful if both cognitive and social processes are supported due to their mutual influence. An overview about how cognitive and social processes are stimulated or supported by means of GA information is provided by Bodemer and Dehler (2011). At that time, three types of GA had become distinguished: behavioural GA (e.g., Janssen et al., 2011), cognitive GA (e.g., Sangin, Molinari, Nüssli, & Dillenbourg, 2011), and social GA (e.g., Phielix et al., 2011). Based on an extensive literature review, another framework of GA support was set up in later years (Janssen & Bodemer, 2013). Considering the common use of terms in existing GA-related studies, the authors describe a division into only two GA components: cognitive (e.g., information about knowledge or opinions of group members) and social (e.g., information about participation or perceived friendliness of group members) GATs. Like Kreijns and colleagues (2013), they distinguish between two dimensions of social interaction, which are stimulated by different types of GATs. The framework presented by Janssen and Bodemer (2013) suggests that cognitive and social GA are prerequisites for the effectiveness of social interaction in the two spaces. Recently, Bodemer and colleagues (2018) analysed that a division into two types of GATs is still established at the first level in current research, whereas a more differentiated view on social GATs is supplied at the second level: a differentiation between tools collecting socio-behavioural information (such as information about the participation of group members), socio-emotional information (such as the perceived friendliness within a group), as well as socio-motivational information (such as the commitment of group members).

Regarding social media communities, we have identified a three-way division (see Section 4) based on the three challenges presented in Section 3. Accordingly, there is a need to more closely examine “social” GA information, since especially (socio-)emotional and (socio-)behavioural processes can achieve different effects but are often cumulated. Such a resumption of the three types of GA information is also suggested in a review by Ghadirian, Ayub, Silong, Bakar, and Hosseinzadeh (2016). The following section proposes an integrated framework, which could serve as a basis for new studies in the field of GAT research, especially regarding social media communities. This framework adopts a distinction between three types of GA information (Bodemer & Dehler, 2011) but replaces the term “social” with “emotional” since all types of GATs in the social media area contain a social component. Thus, the framework distinguishes cognitive, behavioural, and emotional GA information (see Figure 2). Despite the presence of promising separate findings on various GATs, this framework contributes by combining findings from different fields of GAT research considering a holistic and differentiated view regarding the effects of cognitive, behavioural, and emotional GA information on different challenges, learning, and social outcomes as well as group performance (see Figure 2). The illustrated framework clarifies that cognitive GA information entails mainly positive effects on interactions in the content space as well as on learning outcomes by addressing the cognitive challenge. Emotional GA information on the other hand mainly affects the emotional challenge of interactions in social interactions’ relational space and, as a result, also entails positive effects on social outcomes. Furthermore, behavioural GA information heightens social interaction motivation in both spaces. The main message of this framework is that the interaction of all GA information may be crucial to consequent group performance. It should be noted that these considerations represent a beginning and need to be expanded or modified in the future, inter alia by highlighting interaction effects of different GA functions, as the current framework only visualises linear effects. To our knowledge, there is no published study which systematically compares different types of GA information and their interaction effects in order to determine which (combination of) GA information is more fundamental than others

in specific contexts (also see the review of Ghadirian et al., 2016). It is thus imperative to establish studies to investigate GA interaction effects in different contexts.

Kirschner and Erkens (2013) have already presented a framework to summarise the most prominent areas of CSCL research including GAT research. Despite parallels between our framework and that of Kirschner and Erkens, especially regarding the appeal to weigh emotional aspects (e.g., support of the well-being and satisfaction of group members) more strongly in CSCL settings, their issue is domain neutral. Our approach on the other hand focuses on the use of group awareness tools to support larger social media communities such as wikis. A further specific aspect of this framework is that it focuses not only on the objective information provided by a GAT but also on the influencing effect of “personal” GA, which is rarely considered in current research as most studies focus on the GA information collected by the tools rather than the actual GA. Few studies consider GA as a dependent variable (Janssen et al., 2011), mediator variable (e.g., Sangin et al., 2011), or as an independent variable regarding a treatment check (Engelmann & Hesse, 2011). Furthermore, no present study considers the subjective importance of GA information, which could help to predict whether specific tool information will be used as intended. A particular feature of the visualised framework is therefore the distinction between the information presented by the GAT and the actual person’s GA, which depends on the individual’s interaction with the GA information (see Figure 2). Especially when several GA information are combined, individuals may use some types of information more heavily than others during the collaboration. As an example, learners may not care about a learning community’s knowledgeability level if there is a friendly group climate. It is also possible for individuals to draw their own conclusions based on presented GA information, and the visualisation of specific actions could thus be associated with much expertise (see Ogata & Yano, 2004), even if this might be a fallacy.

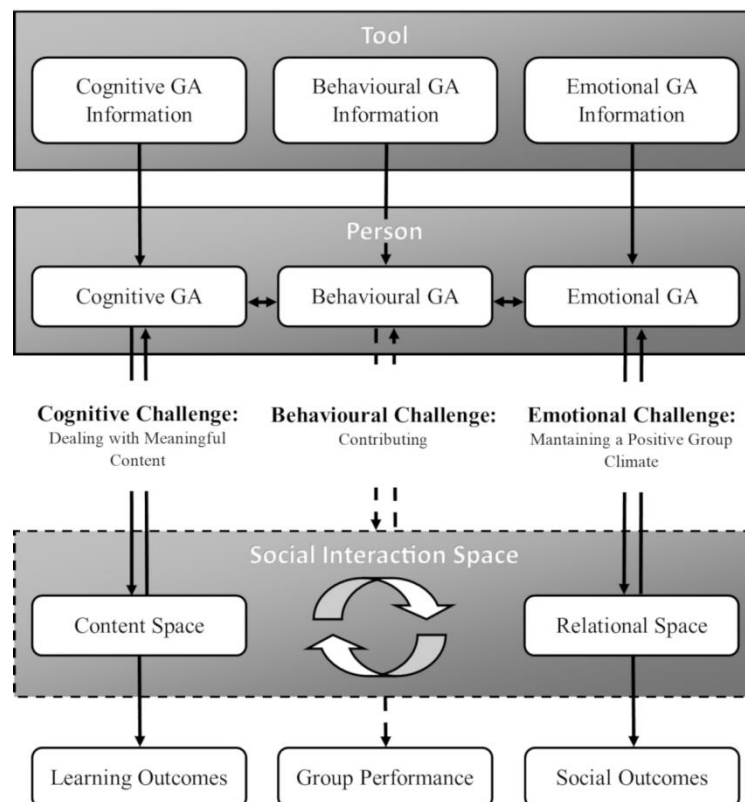


Figure 2. Framework for future studies regarding the interplay of GATs.

The effect of (socio-)motivational GATs (Bodemer et al., 2018) is not explicitly visualised in this framework. Motivational challenges are “related to different goals, priorities, and expectations within the group towards group activities” (Näykki et al., 2014, p. 2). Although different theories of motivation can be very relevant for understanding individual and collaborative learning processes, a complete discussion on this topic exceeds the scope of this article. Rather, we postulate that motivational issues are addressed by each of the three types of GA information (see Sections 3 and 4). Behavioural GA information affects the general *motivation to contribute* (behavioural challenge),

which requires no advanced abilities beyond basic writing and reading. Cognitive GA information addresses, besides the skill (as cognitive states may be difficult to detect in large social media settings) the *motivation to deal with meaningful content* (cognitive challenge). Finally, emotional GA information addresses, besides the skill (as emotional states may also be difficult to detect), the *motivation to maintain a positive group climate* (emotional challenge). In addition, studies have already empirically shown that the simple presentation of motivational states does not need to entail positive effects on outcomes such as increasing one's own motivation or knowledge (e.g., Schoor, Kownatzki, Narciss, & Körndle, 2014). This could be because a simple motivation presentation is highly unspecific, whereas cognitive, behavioural, and emotional GA information concern specific motivational effects.

6. Future Implications

With this framework we want to stress that even though there are already some enlightening and promising results for different types of GATs, it is time to develop a comprehensive full picture regarding their connections. There is a great imbalance in GAT research regarding the types of information provided. The clear focus lies on the support of cognitive GA (Ghadirian et al., 2016) followed by behavioural GATs, whereas the use of emotional GATs remains rather unexplored. To proceed, it is important to examine the positive and negative interaction effects of cognitive, behavioural, and emotional GA information on the respective challenges and outcomes. Although positive effects of cognitive GATs on learning outcomes can already be demonstrated in several contexts (Bodemer et al., 2018), it is likely that behavioural GA information has the potential to intensify these effects. This is based on the finding that explanations help to recapitulate previous knowledge (Webb, 1991). It is also possible that the presentation of cognitive group information leads to information being strategically withheld when learners perceive themselves as experts (Ray, Neugebauer, Sassenberg, Buder, & Hesse, 2013). Here, behavioural GA information could potentially enhance motivation for providing explanations in the social interaction space. However, the presentation of behavioural information could also entail negative effects on emotional challenges or the group climate if the tool visualises unequal participation (Strauß, Rummel, Stoyanova, & Krämer, 2018). There is a need for GATs that present different types of GA information in specific contexts. It is important to examine how this GA information support should look while considering cognitive variables such as mental effort, which could be affected by the interaction with tool information (see Janssen et al., 2011). In this paper, we have exemplarily referred to the area of wikis, however this framework is also transferable to other communities with the challenges being more applicable to larger learning platforms. Since many of the existing studies regarding GATs focus on smaller group collaborations (e.g., Kimmerle & Cress, 2008; Phielix et al., 2011), there is an urgent need in the field of CSCL and GATs to conduct additional research regarding social media platforms, because both students and faculties increasingly use social media in teaching and learning activities (Dabbagh & Kitsantas, 2012). Furthermore, it is important to consider the individual weighting of different GA information as well as the interaction with personality traits. A high tendency towards social comparisons could potentially strengthen behavioural GAT effects (Neugebauer, Ray, & Sassenberg, 2016), whereas need for cognition should influence the cognitive GAT effects. Moreover, a conflict avoidance tendency might affect the interaction with emotional GA information. These represent a few of many conceivable personality interactions that need to be addressed in future GA studies in order to advance this field of research. Laboratory as well as longitudinal field research (Wang, 2011) is needed to examine what kind of processes and outcomes are triggered by the single and combined visualization of GA information, how learners perceive and interact with the different tool information (e.g., by using eye-tracking or qualitative methods) and which role influencing personality variables play. Conducting such studies can help to design and apply adaptive GATs that consider the interplay of different types of GA information as well as support learners according to their specific personalities, which is considered one of the main challenges for future CSCL work (Wise & Schwarz, 2017). Nevertheless, it is not only a question of gaining new insights in the field of GATs, but also of inviting teachers, facilitators, and designers to consider this framework and future research regarding GA interaction effects in order to promote motivation and learning in formal and informal educational settings.

References

- Ayoko, O. B., Callan, V. J., & Härtel, C. E. J. (2008). The influence of team emotional intelligence climate on conflict and team members' reactions to conflict. *Small Group Research, 39*(2), 121–149.
- Bagherian, F., & Thorngate, W. (2000). Horses to water: Student use of course newsgroups. *First Monday, 5*(8).
- Bell, P. (2004). The educational opportunities of contemporary controversies in science. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet Environments for Science Education* (pp. 233–260). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior, 27*(3), 1043–1045.
- Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge.
- Buder, J., Schwind, C., Rudat, A., & Bodemer, D. (2015). Selective reading of large online forum discussions: The impact of rating visualizations on navigation and learning. *Computers in Human Behavior, 44*, 191–201.
- Capdeferro, N., & Romero, M. (2012). Are online learners frustrated with collaborative learning experiences? *The International Review of Research in Open and Distance Learning, 13*(2), 26–44.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.
- Chen, Y.-H., Jang, S.-J., & Chen, P.-J. (2015). Using wikis and collaborative learning for science teachers' professional development. *Journal of Computer Assisted Learning, 31*(4), 330–344.
- Choy, S. O., & Ng, K. C. (2007). Implementing wiki software for supplementing online learning. *Australasian Journal of Educational Technology, 23*(2), 209–226.
- Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education, 15*(1), 3–8.
- Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced Learning: Principles and Products* (pp. 3–19). Dordrecht, Netherlands: Springer.
- Eligio, U. X., Ainsworth, S. E., & Crook, C. K. (2012). Emotion understanding and performance during computer-supported collaboration. *Computers in Human Behavior, 28*(6), 2046–2054.
- Engelmann, T., & Hesse, F. W. (2011). Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. *Computers in Human Behavior, 27*(6), 2078–2087.
- Fransen, J., Weinberger, A., & Kirschner, P. A. (2013). Team effectiveness and team development in CSCL. *Educational Psychologist, 48*(1), 9–24.
- Ghadirian, H., Ayub, A. F. M., Silong, A. D., Abu Bakar, K. B., & Hosseinzadeh, M. (2016). Group awareness in computer-supported collaborative learning environments. *International Education Studies, 9*(2), 120.
- Heimbuch, S., & Bodemer, D. (2017). Controversy awareness on evidence-led discussions as guidance for students in wiki-based learning. *The Internet and Higher Education, 33*, 1–14.
- Heimbuch, S., Ollesch, L., & Bodemer, D. (2018). Comparing effects of two collaboration scripts on learning activities for wiki-based environments. *International Journal of Computer-Supported Collaborative Learning, 13*(3), 331–357.
- Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist, 48*(1), 40–55.
- Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior, 27*(3), 1046–1058.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). Constructive controversy: The educative power of intellectual conflict. *Change: The Magazine of Higher Learning, 32*(1), 28–37.
- Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCs). *Computers & Education, 122*, 9–22.
- Kimmerle, J., & Cress, U. (2008). Group awareness and self-presentation in computer-supported information exchange. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 85–97.
- Kimmerle, J., Moskaliuk, J., Oeberst, A., & Cress, U. (2015). Learning and collective knowledge construction with social media: A process-oriented perspective. *Educational Psychologist, 50*(2), 120–137.
- Kirschner, P. A., & Erkens, G. (2013). Toward a Framework for CSCL Research. *Educational Psychologist, 48*(1), 1–8.
- Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review, 18*(2), 159–185.

- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*(3), 335–353.
- Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social aspects of CSCL environments: A research framework. *Educational Psychologist, 48*(4), 229–242.
- Lin, J.-W., Mai, L.-J., & Lai, Y.-C. (2015). Peer interaction and social network analysis of online communities with the support of awareness of different contexts. *International Journal of Computer-Supported Collaborative Learning, 10*(2), 139–159.
- Lin, J. W., Tsai, C. W., Hsu, C. C., & Chang, L. C. (2019). Peer assessment with group awareness tools and effects on project-based learning. *Interactive Learning Environments, 1-17*.
- Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior, 52*, 573–588.
- Näykki, P., Järvelä, S., Kirschner, P. A., & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning—A process-oriented case study in a higher education context. *International Journal of Educational Research, 68*, 1–14.
- Neugebauer, J., Ray, D. G., & Sassenberg, K. (2016). When being worse helps: The influence of upward social comparisons and knowledge awareness on learner engagement and learning in peer-to-peer knowledge exchange. *Learning and Instruction, 44*, 41–52.
- Ogata, H., & Yano, Y. (2004). Knowledge awareness for a computer-assisted language learning using handhelds. *International Journal of Continuing Engineering Education and Life Long Learning, 14*(4-5), 435-449.
- Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087–1102.
- Piaget, J. (1977). Problems of equilibration. In M. H. Appel & L. S. Goldberg (Eds.), *Topics in Cognitive Development* (pp. 3–13). New York: Plenum.
- Popescu, E., Maria, C., & Udriștoiu, A. L. (2014). Fostering collaborative learning with wikis: Extending MediaWiki with educational features. In E. Popescu, R. W. H. Lau, K. Pata, H. Leung, & M. Laanpere (Eds.), *Advances in Web-Based Learning – ICWL 2014* (Vol. 8613, pp. 22–31). Cham: Springer International Publishing.
- Ray, D. G., Neugebauer, J., Sassenberg, K., Buder, J., & Hesse, F. W. (2013). Motivated shortcomings in explanation: The role of comparative self-evaluation and awareness of explanation recipient’s knowledge. *Journal of Experimental Psychology: General, 142*(2), 445–457.
- Rienties, B., Giesbers, B., Tempelaar, D., Lygo-Baker, S., Segers, M., & Gijsselaers, W. (2012). The role of scaffolding and motivation in CSCL. *Computers & Education, 59*(3), 893–906.
- Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior, 27*(3), 1059–1067.
- Schoor, C., Kownatzki, S., Narciss, S., & Körndle, H. (2014). Effects of feeding back the motivation of a collaboratively learning group. *Electronic Journal of Research in Education Psychology, 12*(1), 191–210.
- Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G. M., & Janssen, J. (2010). Guiding students’ online complex learning-task behavior through representational scripting. *Computers in Human Behavior, 26*(5), 927–939.
- Strauß, S., Rummel, N., Stoyanova, F., & Krämer, N. (2018). Developing a library of typical problems for collaborative learning in online courses. In J. Kay & R. Luckin (Eds.), *Rethinking Learning in the Digital Age: Making the Learning Sciences Count* (Vol. 2, pp. 1045–1048). London, UK: ISLS.
- Walther, J. B., & Bunz, U. (2005). The rules of virtual groups: Trust, liking, and performance in computer-mediated communication. *Journal of Communication, 55*(4), 828–846.
- Wang, T. H. (2011). Developing web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-learning environment. *Computers & Education, 57*(2), 1801–1812.
- Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.
- Wilcox, P., Winn, S., & Fyvie-Gauld, M. (2005). ‘It was nothing to do with the university, it was just the people’: the role of social support in the first-year experience of higher education. *Studies in Higher Education, 30*(6), 707–722.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423–467.
- Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS ONE, 7*(6).
- Zheng, B., Niiya, M., & Warschauer, M. (2015). Wikis and collaborative learning in higher education. *Technology, Pedagogy and Education, 24*(3), 357–374.

Group Awareness Information to Support Academic Help-Seeking

Christian SCHLUSCHE^{a*}, Lenka SCHNAUBERT^a & Daniel BODEMER^a

^a*Media-based Knowledge Construction,
University of Duisburg-Essen, Germany*

*christian.schlusche@uni-due.de

Abstract: Students are expected to deal with complex learning material, that regularly challenges their learning processes especially in the first semester. When students ask peers for explanations, they apply a functional learning strategy, known as academic help-seeking. But finding peers to discuss questions on learning materials is difficult for those that are new to university. Finding adequate peers for help-seeking requires awareness about their knowledge and competencies in and beyond the learners' own study group; i.e. group awareness. This paper investigates how group awareness information supports academic help-seeking among first-semester students on a Moodle-based learning platform that has been adapted with a newly developed extension to collect, transform and visualize group awareness information about peers. In a field study accompanying an introductory statistics lecture ($N = 107$), we examined (1) motivational, attitudinal and social characteristics of helpees that asked for help on the platform, (2) the effect of provided group awareness information on the selection of helpers, and (3) the cognitive characteristics of selected helpers and their difference from the helpees. The participating students were invited to seek help from their peers within Moodle and had access to a sortable list of all course members in order to find an appropriate helper. The control group ($n = 55$) only saw the full names of the peers while the treatment group ($n = 52$) was additionally provided with self-assessed group awareness information (competence and availability of a helper). Unfortunately, participants rarely used the platform and only 22 help-seeking episodes were recorded. Results indicate that helpees did not differ regarding motivational, intentional or social factors from the rest of the course. The overall occurrences of sorting the list was higher for the condition provided with group awareness information, but both groups did not differ with respect to the competence of selected helpers. The results provide useful, practical insights bringing online help-seeking into an educational context.

Keywords: academic help-seeking, learning strategies, group awareness, interactive learning environments, Moodle, undergraduate students

1. Introduction

During the first year at university, students are in a transition from more guided learning at school to more self-regulated learning at university. This transition confronts them with cognitive and social challenges (Credé & Niehorster, 2012). Students are expected to adapt to a plenty of new and complex learning material, which they were not yet confronted with prior to university. Moreover, undergraduate students find themselves within large social groups with mostly unknown peers. When they experience knowledge-related problems they may consult sources of help to acquire missing information, which is called academic help seeking (Nelson-Le Gall, 1981). When seeking help to construct valid understanding, students can choose from various (*im-*)*personal* sources, differing in the adaptivity of their provided help (Makara & Karabenick, 2013; Minna Puustinen & Rouet, 2009). *Impersonal* sources, such as books or the world wide web enable learners to re-read information, but still require them to transfer information from the representation to their problematic task at hand. When this transfer fails the student does not reach a sufficient understanding of the topic. In contrast, *personal* sources of help are individuals that can provide missing information and adaptive assistance for an academic problem; these are usually lecturers, tutors or peer students. When asked for help, personal sources are capable to take the knowledge of their dialogue partner into account and can thus provide explanations that are tailored to the student who asks for help (Webb, 1989). Research has shown that

peers are the preferred source of help among students (Knapp & Karabenick, 1988). This person-centered process of asking for help comes along with the problem of choosing adequate peers as helpers: Before asking a peer, a freshman needs to know which potential helper has mastered the learning material sufficiently to provide help and is additionally willing to do so. Students usually attend lectures in large groups, but regularly work in smaller study groups. As students lack interaction with peers beyond their study groups (Gibney, Moore, Murphy, & O'Sullivan, 2011), they might need more information about peers outside their study group to consider them as sources for help. Information, such as individual competence for a topic, may enable learners to seek help beyond their study group. The provision of information about internal states (e.g. cognitive or motivational) of group members are examined within Group Awareness (GA) research (Bodemer, Janssen, & Schnaubert, 2018; Janssen & Bodemer, 2013). It has been found that such information can guide learners' attention within interactive learning environments (Dehler, Bodemer, Buder, & Hesse, 2011). Such information about peers may be useful to identify adequate personal sources of help that would not have been considered otherwise. Since personal help-seeking is a promising strategy to deal with knowledge-related problems, but requires sufficient GA within the course, the provision of GA information might help identifying potential helpers beyond the learner's study group and may thus improve the resolution of knowledge-related conflicts.

2. Academic Help-Seeking

Learning domains at university are more challenging than in school and thus, students struggle more regularly to establish understanding. Moreover, students are expected to take responsibility for their own learning processes. Students regularly experience difficulties during working with learning materials (e.g., when preparing an assignment), which may cause the learners to realize that their understanding of a concept is insufficient, and help is needed. By seeking help, the learner gets into the role of a *helpee*, who seeks information from a *helper* (or *source of help*). The literature describes help-seeking as a functional learning strategy that is related to study success (Karabenick & Knapp, 1991; Richardson, Abraham, & Bond, 2012). The helpee can mainly pursue two different aims, either seeking *executive* or *instrumental* help (Nelson-Le Gall, 1981). If helpees expect helpers to solve tasks in their stead, the helpee seeks *executive* help; which is neither contributing to understanding nor fostering independence from others. Helpees seeking *instrumental* help are asking for necessary information that enables them to develop a solution on their own. Optimally, this approach improves the helpee's understanding and is a sign of independent behavior (Nelson-Le Gall, 1981).

2.1 Help-Seeking Process

Help-seeking can be understood as a process, consisting of 8 phases (Makara & Karabenick, 2013; Nelson-Le Gall, 1981). At the beginning the learner has to (1) determine that a problem exists, then (2) determine that help is needed and (3) decide to seek help. Next the helpee (4) clarifies the goals of the request, whether seeking executive or instrumental help. The helpee continues to (5) decide whom to ask by choosing an appropriate helper, that may provide the expected type of help. The helpee continues by (6) soliciting help through formulating a question and subsequent communication between the helpee and the chosen helper. In a positive case this may result in (7) obtaining the requested help which leads to a solution for the problem. Nelson Le-Gall appends the subsequent phase (8) evaluation of help-seeking attempts, that may affect future decisions for seeking help considering effort and use of the implemented strategy (Nelson-Le Gall, 1981).

When a student deals with a knowledge-related problem and decides to seek help from a personal source, the next decision he or she has to make is *whom to ask for help* (phase 5). This decision depends on which personal sources can be accessed and if those are likely to provide helpful support. Knapp and colleagues (1988) have differentiated *formal* and *informal* types of personal sources, depending on their obligation to provide help in an educational context (Knapp & Karabenick, 1988). Formal sources usually belong to an academic institution (e.g., lecturers or tutors) and are expected to support learners because of their occupation. Whereas, informal sources (e.g., peers or friends) are not

occupied at an institution and are thus independent in their decision to provide support. *Peer students* usually share a similar understanding, as they were provided with same materials and tasks.

Support from peers in academic help-seeking occurs in various situations during the semester and is difficult to evoke in laboratories. Assessing actual help-seeking behavior is more effortful than asking for attitudes towards help-seeking and inferring to behavior. According to Karabenick & Knapp (1991) help-seeking *intentions* seem to be a good predictor of actual behavior; these are also called attitudes towards help-seeking (Huet, Dupeyrat, & Escribe, 2013; Karabenick & Knapp, 1991). Those are usually assessed through self-report with items formulated to measure the general intention to seek help within a particular class context (Pajares, Cheong, & Oberman, 2004; Sánchez Rosas & Pérez, 2015). However, Huet and colleagues (2013) discuss a gap between assessed intentions and actual help-seeking behavior: According to Huet and colleagues the assessment of the intentions is too general, hence the predictive quality for actual behavior is rather low. Actual help-seeking behavior can be assessed by observation with recorded video material, though transcription of this materials is time-consuming (Nelson-Le Gall & Glor-Scheib, 1986). Another promising approach to assess help-seeking behavior is tracking learners' inputs within interactive learning environments that mediate between helpes and helpers, which provides data that can be processed with little effort.

2.2 Source Selection

When a helpee chooses a helper among peers, relevant qualities as criteria of a good source of help need to be identified. Which qualities of helpers among peers make them likely to provide helpful answers? Makara and colleagues (2013) developed the expectancy-value *model of source selection and utilization (MSSU)* to address this question. Expectancy-value models predict an outcome variable that is expected to be the multiplicative conjunction of an expectancy- and a value-component towards reaching a goal (Eccles & Wigfield, 2002). Makara and colleagues' model predicts the likelihood of soliciting help from a particular source by considering the expectancy that a helper will provide help (e.g., are they available) and the values the helpee ascribes towards the helper (e.g. are they competent to help) (see Fig. 1). High availability of a helper, for instance, leads the helpee to the expectation that a problem can be solved in the near future and waiting time can be reduced to a minimum. When a source is known to be competent and can thus be asked to provide high quality help, the helpee can trust her/his explanations. As both components are multiplied, one component can constrain the whole outcome: whenever one component is low, the predicted likelihood of soliciting help from this source is reduced as well (e.g. if a highly competent source is unavailable). In conclusion, the higher both components are, the higher is the likelihood of soliciting help from this particular source. From this perspective a good potential helper is characterized as highly available and providing high quality answers (i.a., being competent). This source selection model emphasizes the importance of reliable information about potential helpers to identify a good source of help. During their first year, students may lack this information from peers outside their study group. Thus, they either need to invest time to gain information about unknown others on their own, over several help-seeking episodes or they may avoid seeking help from others at all. Finding an appropriate helper is essential for successful help-seeking, thus the provision of supporting information about helpers can improve the help-seeking behavior of undergraduate students.

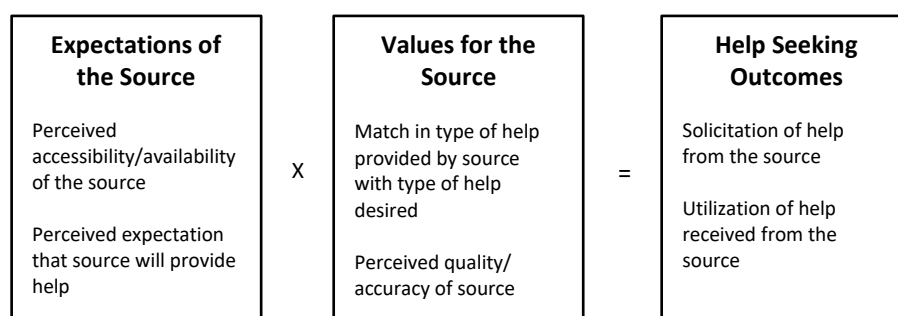


Figure 1. Expectancy-value model of source selection (based on Makara & Karabenick, 2013).

2.3 Group Awareness Tools Supporting Help-Seeking

Asking others for help is an inherently social activity which requires non-trivial communicative behavior: the helpee usually describes a problem and the helper will subsequently try to understand and provide explanations. The coordination of such learning activities, in which at least two learners interact, requires mutual understanding. During communication, each member within a group builds a more or less valid internal representation of socio-cognitive states (e.g., knowledge) or socio-motivational states (e.g., invested time) of other members, referred to as *group awareness* (GA) (e.g., Bodemer et al., 2018; Janssen & Bodemer, 2013). The degree to which GA is developed increases with interaction frequency among the group and varies for individual group members. Research in the field of Computer-Supported Collaborative Learning indicates that GA can be improved by the provision of information about group members within digital learning environments, via so called *GA tools* that collect, transform and visualize learner data (Bodemer et al., 2018). Learners can benefit from GA tools as these tools have been found to support, amongst others, communication and selection processes within collaborative learning settings (e.g., Dehler et al., 2011; Schnaubert & Bodemer, 2019). Ogata & Yano (2004) have used awareness information in form of concept maps to successfully recommend helpers with highest interested in similar topics (Ogata & Yano, 2004). Research suggests to support helpees trying to select an appropriate helper within the process of academic help-seeking, by providing GA information about the potential helpers. Based on help-seeking literature it seems promising to present information as GA information that are in accordance with the MSSU (see Sec. 2.2). We adapted *availability for questions* from the expectancy-component and *competence to give explanations* from the values-component as GA information. Consequently, the question arises of how students consider this information for their selection of a helper?

2.4 Research Questions

The current research addresses the potential of providing group awareness information to guide *whom to ask for help*, supporting students making better decisions on potential helpers on an interactive learning environment. The following research questions shall be examined:

Such an optionally accessible interactive environment demands the investment of free time, which may involve certain attitudes towards learning and motivation to discuss problems. Can students that make use of such an environment be differentiated from those who do not? This brings up the question of who makes use of such an environment and actively seeks help, thus we ask *RQ 1: What are characteristics of helpees asking for help at an ILE?*

Based on the MSSU (see 2.2) the provision of information regarding competence and availability should be useful to identify an adequate source of help, but are these criteria really used by students, when they look for a helper within a larger group of potential helpers? Thus, we want to know *RQ 2: What information do students use when they select a helper out of a larger group?*

Helper with higher levels of both components are predicted by the MSSU (see 2.2) to be selected most frequently. Considering social factors (e.g., zone of proximal development [Vygotsky, 1986]) helpees may also prefer helpers with similar values on competence. What about social factors such as friendship or expertise that may attract helpees towards popular helpers (Nurminen, Heino, & Ihantola, 2017)? Thus, we want to know *RQ 3: How do cognitive characteristics of helper and helpee relate to each other?*

3. Methods and Design

To answer our RQs, we enriched an online course by a group awareness tool which enabled students to seek help from other participants via one-on-one private messages. In this field study the between-subject factor *group awareness information* was systematically varied, information was either available (GA+) or not available (GA-). Within a participant list the control condition was only provided with full names of the participants, whereas the treatment condition was additionally provided with visualizations about self-assessed competence and availability. The online learning platform

accompanied an introductory statistics lecture for four weeks, in the first third of the semester. Unfortunately, students did rarely use the platform.

Participants were attended an introductory statistics lecture (winter term 2017) at the *University of Duisburg-Essen*. 107 participants initially signed up on the online platform and were alternately assigned to either the control condition (GA-; $n = 55$) or treatment condition (GA+; $n = 52$). After four weeks, 96 subjects finished the study (94 were freshman students, 70 females), aged 18 – 49 years ($M = 20.98$ years, $SD = 4.97$). 76 were enrolled in the *Applied Cognition and Media Science* study program, 18 were enrolled in Psychology; 2 participants did not disclose their study program. Participants received credit for participation and had the chance to win a voucher worth 50 €.

3.1 Learning Environment and Procedure

The online platform was realized with Moodle™ Learning Management System (LMS). The interaction of participants was centered around a sortable participant list with all course members. A new Moodle extension was developed to supplement the list with self-reported GA information (GA+ condition). The plugin collects self-reported data from the participants and visualizes the information next to their respective names. Information about helpers' *competence* and *availability* were handled within the plugin and fed back to the helpees (see Sec. 2.2). Designed with the goal to support the decision of *whom to ask for help*, students could sort the list by picking one of the provided information and subsequently choose a helper, by clicking the message icon and writing a private message (see Fig. 2). As the platform was a digital mediator between helpers and helpees, each sorting, each selection and each message was logged. The list emitted a sort-event with the criterion 'lastname' by default, each time it was initially accessed (default setting).

Participants went through the following procedure: After enrolling to the course (assignment to experimental condition), participants were asked to read the instructions on how to use the platform; followed by an introductory survey assessing motivational constructs (see Sec. 3.3). Subsequently, participants were asked to provide group awareness information by filling two self-reports about their competence for the past weeks' lectures and current week's lecture (see Sec. 3.2). These self-reports were assessed each week on the day of the lecture. Moreover, participants shared their availability for requests.

Firstname / Lastname	Competence (agg. past)	Competence	Availability
Zaphod Beeblebrox 			
Fanny Dent 			
Joe Dent 			
Ford Prefect 			

Figure 2. Sortable list of potential helpers with additional information (competence and availability)

3.2 Assessing and Visualizing Group Awareness Information

We provided participants within the GA+ condition with two types of awareness information: competence and availability. GA information regarding competence was assessed once a week after the lecture took place via a short, mandatory self-report questionnaire. Participants were asked whether their knowledge of the contents of the past lectures was sufficient to explain them to a fellow student on a 6-point equidistant response scale (*completely true – not true at all*). Two ratings were collected each week: one assessing all past weeks' topics and one asking about the current week's topic. Both competence-ratings were transformed onto a 3-point scale and were visualized as stars (1 to 3) in the participant list for GA+ (see Fig. 2), if no data was provided no stars were visualized (*left "blank"*).

Additionally, the GA tool provided participants in GA+ with information about the availability of a helper. Data representing availability was collected on a binary scale (*available – not available*). Participants were free to change their availability status at any time during the study, but – apart from the initial instruction – were not explicitly prompted to do so. Information on availability were

visualized within the participant list as a traffic-light-like filled circle, showing green for “available” and red for “not available”.

3.3 Measures

The following measures were used to characterize helpees (see RQ 1): motivational (mastery-goal orientation), intentional (help-seeking intentions) and social (social connectedness). Motivational factors esp. mastery-goal orientation were assessed with the subscales mastery-approach and performance-avoidance of the SELLMO on a 5-point Likert scale (*strongly agree – strongly disagree*). *Help-seeking intentions* were measured in four different types, namely instrumental, executive, avoidance of help-seeking and perceived benefits of help seeking, with three items each on a 7-point Likert scale (*completely disagree to completely agree*) (Pajares et al., 2004). Friendship among course participants was measured by a rating of how good participants know their fellows, on a 4-point scale (ranging from *unknown* to *friends*). *Social connectedness* was operationalized as the number of participants that were given one of the ratings at the upper half of the scale (*known associate* or *friend*) at the beginning of the data acquisition.

Sorting-events and self-reported data were examined to investigate the considered information for selecting a helper (see RQ 2). The tracked *sorting-events* in Moodle indicate which information was considered when choosing a helper. These events usually precede the selection of a helper at the start of a new help-seeking episode. In order to gain an overview of the list presenting all course participants, each represented information could be used as a criterion to sort the entries accordingly. Each sorting-event of the participants’ list was logged including, when it was triggered, which user has sorted and which criterion and order (ascending / descending) was used. It is obviously possible to select a helper independently from the sorting criterion, but the sorting-criterion may be a relevant indicator of how students prioritize available information about helpers.

The characteristics of chosen helpers (targets) (RQ 3) are stored immediately after selection. Various information about the target’s provided group awareness information are stored. The *competence-value* (0-5) represents the mean of the self-assessed competence items. These means get transformed into three *competence-level*, which are visualized by 3 stars to 1 star (higher is better). Furthermore, target’s position in the list and information about the distribution of competence-classes among the course members, at the time the event is triggered, gets stored. Another interesting aspect is the difference between the helpee’s competence level relative to the helper’s, revealing helpees’ preferences for characteristics of helpers.

4. Results

4.1 Characterization of Helpees

In the following, those participants that made use of the platform are characterized (see RQ 1). The subgroup of those participants, in the role of helpees ($n = 14$), shall be described with regard to motivational, attitudinal and social factors, independent of their experimental condition (see Table 1). Results of this subgroup are presented side-by-side with those that did not seek help, which formed the majority. Both groups have relatively high values for mastery-goal-orientation, instrumental help-seeking and perceived benefits of help-seeking. Regarding the constructs, performance-goal-orientation, executive help-seeking intentions, and help-seeking avoidance, both groups have in common that their values are at the lower end of the scale. Regarding social connectedness both groups do not differ. No differences can be found, that particularly characterize helpees who used the platform.

4.2 Used Sorting Criteria

The usage frequencies of each sorting-criterion is depicted in Table 2 (see RQ 2). The table contains data for each week along the columns; each row contains frequencies for a criterion, per condition. In each of the four course weeks (*columns*), the overall usage of the sorting feature is lower in GA- than in

GA+ (*bottom of columns*). When comparing usage over time, between both groups, it gets obvious that they differ at week 2 and 3. Accumulated data over the whole period reveals that GA+ sorted nearly twice as frequent as the control condition (*last column*). The frequencies for the lastname-criterion are highest each week, as it was triggered by default whenever a course participant navigated to the participants' list. For this reason, merely the cautious conclusion can be made that GA+ accessed this page more often; but not necessarily preferred sorting by lastname-attributes. Amongst the additional information only GA+ could leverage, it gets obvious that competence-information was used more often than (albeit still rarely) availability-status, which was not used (at all). Both types of competence-information combined (past weeks and current week) were used in about 22% of the weekly summed sortings of the treatment condition (week 1/2/3/4: 12%, 22%, 22%, 26%).

Table 1
Characteristics of Helpees' Side to Side with Participants who Did Not Use the Platform

	Goal-Orientation † <i>M (SD)</i>		Help-Seeking Intentions †† <i>M (SD)</i>				Connect- edness <i>M (SD)</i>
	Mastery	Performance	Instrumental	Executive	Avoidance	Benefit	# Friends
Helpees*	4.13 (0.51)	2.06 (0.73)	5.77 (1.02)	1.57 (0.59)	2.67 (0.89)	4.57 (1.01)	10.5 (6.92)
Other**	4.15 (0.40)	2.37 (0.70)	5.24 (1.05)	1.83 (0.80)	2.88 (1.17)	4.27 (1.12)	8.19 (5.08)

* Helpees: asking for help: GA-: 6 / GA+: 8

** Other: not asking for help: GA-: 41 / GA+: 41 † [1-5] †† [1-7]

Table 2
Frequencies of the Usage of Sorting-Criteria per Experimental Condition

Criterion	Week 1	Week 2	Week 3	Week 4	Accumulated
	GA -/+	GA -/+	GA -/+	GA -/+	GA -/+
Firstname	3 / 2	12 / 5	5 / 2	0 / 3	20 / 12
Lastname	32 / 34	30 / 56	24 / 48	2 / 11	88 / 149
Availability	-- / 0	-- / 0	-- / 1	-- / 0	-- / 1
Past weeks' competence	-- / 3	-- / 8	-- / 6	-- / 2	-- / 19
Current week's competence	-- / 2	-- / 9	-- / 14	-- / 3	-- / 28
Summed frequencies	35 / 41	42 / 78	29 / 90	2 / 19	108 / 209

4.3 Sortings that Lead to a Selection

The aim of sorting the participants' list was to support identifying an appropriate helper. Over the four weeks of the study there were 22 selections made by 14 different helpees who selected 18 different helpers. These selections were preceded in 68% of the incidents by a lastname-sorting (GA-: 6 / GA+: 9), in 23 % of incidents by a firstname-sorting (GA-: 4 / GA+: 1) and in 9% of the incidents by a competence-sorting (GA-: - / GA+: 2) (of the current week). The lastname-sorting is recorded most often but must be interpreted with caution (see Sec. 3.3). Immediate before a selection the competence-criterion was sparsely used and availability-criterion was not used at all.

4.4 Who Has Been Asked for Help on the Platform?

There were only 22 selections over the four weeks of the course. The frequencies of the selected helpers competence were as follows (split by condition): 3-stars helper (GA-: 4 / GA+: 1), 2-stars helper (GA-: 2 / GA+: 3), 1-star (GA-: 2 / GA+: 0) and no-stars “blank” (GA-: 4 / GA+ : 6). GA+ was expected to select more competent helpers, as they were provided with competence information. The frequencies of selected helper’s competence do not reveal a difference between both conditions: Most often helpees choose students that did not had a visualization for current’s weeks competence (“blank”). 1-star- and 2-star-helpers were selected equally often. The incidence of 3-stars selections is too rare to be interpreted. In conclusion both groups seem not to differ regarding the competence of selected helpers, hence no pattern for preferred helper’s competence could be found.

4.5 Helpee and Helper: Differences Regarding Competence

Helpees are in need for information or explanations, hence it can be expected that they seek helpers who are knowledgeable. But *how* knowledgeable or competent, were the selected helpers relative to the helpees? For the few occurred help-seeking episodes the proportion of the competency-levels of both agents are depicted in Table 3. The recorded data reveals that helpees selected more competent helpers in about 27% of the cases (*in green*), equally competent helpers in 14% of the selections (*in blue*) and similarly selections of less competent helpers made in 14% of the cases (*in orange*), regarding current week’s competence. There were 11 selections of helpers (about 48%) that did not have a visualization for current’s weeks competence (“blank”). Possibly the visualization for “past weeks” might have also been considered before selecting. Overall there are no differences between both conditions regarding the pattern of competence-levels between helpees and helpers.

Table 3
Occurrence of Competence Patterns for Each Help-Seeking Episode, per Condition

		Helper’s competence-level GA-/GA+				
		3	2	1	blank	
Helpee’s competence level	1	0 / 0	2 / 1	0 / 0	2 / 1	* <i>competence visualizations ranged from 3 to 1; blank: no data provided</i>
	2	1 / 2	1 / 0	0 / 1	4 / 2	
	3	0 / 2	0 / 1	0 / 1	0 / 1	
	blank	0 / 0	0 / 0	0 / 0	0 / 1	

5. Discussion

The presented study investigates the effects of group awareness (GA) information on the selection of helpers within an interactive learning environment. Learners in GA+ condition were provided with information on competence and availability of their peers, learners in GA- condition were merely provided with their peers’ names. Unfortunately, the adapted Moodle platform was rarely used for academic help-seeking (HS) which is a major limitation of the study. This was probably caused by the lack of accessibility on mobile devices, which made the communication with helpers more time-consuming. However, the data acquired provides insights into practical challenges when implementing such support into educational settings. Qualitative feedback provides anecdotal evidence that whole groups in messenger services were considered instead of Moodle one-on-one communication for asking questions. One-on-one communication was implemented within this study, with the goal to reduce avoidance tendencies. Addressing whole groups (i.e. one-to-many) could be related to the expectation that one of a group might respond quicker than an individual (i.e., increase of availability, but not necessary competence). Further research should investigate factors (e.g., avoidance tendencies, or group size) under which communication with a group is preferred to a single helper. In group contexts, privacy concerns should be considered (esp. high perceived HS threat) (Kitsantas & Chow, 2007). It seems plausible that students used other channels, because the predictors of HS we found in our study were high in value. We also found unexpectedly high social connectedness among

freshmen (see 4.1). Data was collected early in first semester at university and students might not always be aware of metacognitive inaccuracies, hence future research might consider supporting learners identifying knowledge-related misconceptions (Kruger & Dunning, 1999).

The first research question addresses motivational, intentional and social differences in the characteristics of first-semester students who sought help on the ILE. Those students, asking for help at least once, did not differ from the rest of the course in terms of the goal-orientation, HS intentions and social connectedness (see 4.1). Overall, the sample reported rather high values on constructs that are positively related with HS behavior (e.g., mastery-orientation, instrumental HS intentions). Assessed variables do not indicate a general lack of motivation or high values of perceived threat, surprisingly, HS was still low. As social connectedness was higher than expected for the first weeks at university, participants might have already formed study groups that could provide sufficient help when needed.

The second research question investigates which sorting criterion (information) was considered most useful and was used most often for sorting helpers. The sorting mechanism should enable students in GA+ to identify adequate helpers by using the GA information competence and availability. GA+ compared to GA- produced about twice the number of overall sorting-events (especially in weeks 2 & 3) and sorted by competence in about 22 % of the cases; whereas availability was not used (see 4.2). As these events, were not dependent on selections of helpers, they may be a byproduct of exploring the information of other peers (“lurking”). Besides, GA+ viewed the sortable list more frequently. However, we cannot infer the explicit use of particular information. The immediate use of a sorting-criterion before a selection of a helper could be an indicator for its utility. The last name-criterion preceded a selection most often (68%), followed by first name (23%), with no differences between conditions. GA+ used competence-sortings only two times immediate before selection (see 4.3), that raises the question, if sorting was necessary for helpees to choose a helper. Whether the last name-criterion was used intentionally for sorting or if the recorded data stems from default visits is unclear: For the control condition it is conceivable that this criterion has been helpful and thus has been used intentionally, whereas for the treatment condition it is likely that the frequency resembles “lurking” behavior. There is no evidence that students use the additionally provided information in their HS process, thus the utility value of additional information remains unclear. Future research should consider that a single-criterion sorting functionality is of limited use to measure the utility of multiple information categories, as it does not account for mental references the learner establishes during the use of such information. Availability, for example, has never been used for sorting, but still could have been evaluated, right after sorting. Future studies could use stepwise filtering mechanisms, which let users identify helpers based on multiple decisions each based on a single criterion.

Investigating the third research question provides insight in who has been asked for help and thereby provides indicators which characteristics of a helper are preferred by learners. It was expected that GA+ would consider additional competence information and select highly competent helpers. However, data does not confirm this hypothesis (see 4.4): participants most frequently asked helpers with unknown “blank” competence-levels, independent from condition. 2-stars and 1-star helpers were selected equally often by both conditions. 3-star-selections were so rare that those cannot be interpreted. Both conditions do not differ regarding the patterns of helpees’ and selected helpers’ competence, hence further criteria might be involved in a selection, e.g. friendship or prior acquired knowledge about peers (Nurminen et al., 2017). Future research shall assess reasons why a helper has been chosen, to identify additional factors influencing the decision. Moreover, we also shed a light on the difference regarding cognitive information between both agents that interacted within an episode. Helpers with the same competence as the helpee were selected least often (see 4.5). A high amount of selections were related to helpers who did not provide information for the current week (“blank”) at time of selection. Other sources of information might have been considered, e.g. about past weeks’ competence or previous interactions. The data does not provide any tendency on whether helpees prefer a rather high or low difference between their own and the helper’s competence (Vygotsky, 1978). When examining these decisions, competence levels and motivational factors predicting help-seeking avoidance should be considered (see Ryan & Pintrich, 1997). Future studies are advised to use a more differentiated scale of competence-visualizations (> 3-levels) for the detection of possible differences.

The presented study was a first step into the support of online academic HS and describes a newly developed extension for Moodle. Results have uncovered new, mostly practical issues regarding the implementation of GA information for HS in such environments. In future possible effects of GA on academic HS should be examined assessing data over the whole first semester (esp. closer to exams).

Moreover, research needs to be done on how students actually use social media and messenger apps on mobile devices to seek help in academia and to identify additional ways to support online HS.

References

- Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge/Taylor & Francis.
- Credé, M., & Niehorster, S. (2012). Adjustment to College as Measured by the Student Adaptation to College Questionnaire: A Quantitative Review of its Structure and Relationships with Correlates and Consequences. *Educational Psychology Review, 24*(1), 133–165.
- Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior, 27*(3), 1068–1078.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132.
- Gibney, A., Moore, N., Murphy, F., & O’Sullivan, S. (2011). The first semester of university life; ‘will I be able to manage it at all?’ *Higher Education, 62*(3), 351–366. <https://doi.org/10.1007/s10734-010-9392-9>
- Huet, N., Dupeyrat, C., & Escribe, C. (2013). Help seeking intentions and actual help-seeking behavior in interactive learning environments. In M Puustinen, S. A. Karabenick, & M. Puustinen (Eds.), *Advances in Help-Seeking Research and Applications: The Role of Emerging Technologies* (pp. 37–72). Charlotte, NC: Information Age Publishing.
- Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist, 48*(1), 40–55.
- Karabenick, S. A., & Knapp, J. R. (1991). Relationship of Academic Help Seeking to the Use of Learning Strategies and Other Instrumental Achievement Behavior in College Students. *Journal of Educational Psychology, 83*(2), 221–230.
- Kitsantas, A., & Chow, A. (2007). College students’ perceived threat and preference for seeking help in traditional, distributed, and distance learning environments. *Computers & Education, 48*(3), 383–395.
- Knapp, J. R., & Karabenick, S. A. (1988). Incidence of formal and informal academic help-seeking in higher education. *Journal of College Student Development, 29*(3), 223–227.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134.
- Makara, K. A., & Karabenick, S. A. (2013). Characterizing sources of academic help in the age of expanding educational technology: A new conceptual framework. *Advances in Help Seeking Research and Applications: The Role of Information and Communication Technologies, 37–72*.
- Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review, 1*(3), 224–246.
- Nelson-Le Gall, S., & Glor-Scheib, S. (1986). Academic help-seeking and peer relations in school. *Contemporary Educational Psychology, 11*(2), 187–193.
- Nurminen, M., Heino, P., & Ihantola, P. (2017). Friends and gurus: Do students ask for help from those they know or those who would know. *Proceedings of the 17th Koli Calling International Conference on Computing Education Research, 80–87*.
- Ogata, H., & Yano, Y. (2004). Knowledge awareness map for computer-supported ubiquitous language-learning. *The 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, 19–26*.
- Pajares, F., Cheong, Y. F., & Oberman, P. (2004). Psychometric analysis of computer science help-seeking scales. *Educational and Psychological Measurement, 64*(3), 496–513.
- Puustinen, Minna, & Rouet, J.-F. (2009). Learning with new technologies: Help seeking and information searching revisited. *Computers & Education, 53*(4), 1014–1019.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students’ academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387.
- Ryan, A. M., & Pintrich, P. R. (1997). Should I ask for help? The role of motivation and attitudes in adolescents’ help seeking in math class. *Journal of Educational Psychology, 89*(2), 329–341.
- Sánchez Rosas, J., & Pérez, E. (2015). Measuring Threats, Benefits, Emotional Costs and Avoidance of Academic Help-Seeking in Argentinian University Students. *Pensamiento Psicológico, 13*(2), 49–64.
- Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning*.
- Vygotsky, L. (1986). *Thought and Language—Revised Edition*. The MIT Press.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*(1), 21–39.

Characterization of Different Instantiations of Mathematical Blindness

Errol Matthew C. GARCIA^{a*}, Maria Alva Q. ABERIN^{b*}

^aGraduate Student, Ateneo de Manila University, Philippines

^bFaculty, Ateneo de Manila University, Philippines

*errol.garcia@obf.ateneo.edu *maberin@ateneo.edu

Abstract: Twenty-four (24) Grade-11 senior high school students under the STEM track from a private university located in Quezon City took part in this study. Of the 24 participants who took the pre-test, only 23 proceeded with the eye-tracking test, where only 19 data were deemed treatable due to technical issues. This study examined different instantiations of mathematical blindness thru the lens of the Tripartite theory and with the help of the eye-tracking method that collected empirical data and aided in the investigation on how these instantiations manifested in participants' mathematics problem-solving, particularly problems involving quadratic equations, ratio and proportion, geometry, and concept of speed. Other than the answers, solutions, and interviews recorded, quantifiable data were also extracted from the eye-tracker in addition to gaze movement and heat maps. Three instantiations of mathematical blindness were characterized in this study: (1) The Einstellung effect; (2) spurious correlation; and (3) intuitive rule. Among the three, the predominant instantiation of mathematical blindness observed was Spurious correlation.

Keywords: Psychology of Mathematics Education, problem-solving, reasoning, eye-tracking

1. Introduction

Mathematical blindness is defined as a cognitive impairment caused by lack or surplus of attention given to present stimuli (stimuli referring to words, details, or figures that are present in a problem). The differing amounts of attention given to present stimuli may cause the individual to fail, to some degree, in cognition—to process the necessary concepts, and/or procedures to be able to answer or find the efficient solution to a certain problem. The concept of mathematical blindness is not new as some of its aspects have already been studied both in the field of Psychology and Mathematics Education. This study only aims at characterizing how the said phenomenon manifests in students' solution or reasoning about a given problem. In addition, the difference between the concept of inattentive blindness (in psychology) and mathematical blindness is that the former describes only the instance when an individual lacks in giving attention to stimulus or stimuli while the latter describes the event where a person lacks or gives too much attention to given stimuli leading to misuse or overgeneralization of mathematical tools. Tools that are commonly used by students include formulas (e.g. quadratic formula), algorithms (e.g. setting up a proportion), strategies (e.g. identifying key words), and intuitions (e.g. multiplication makes bigger).

Theoretically, mathematical blindness may manifest in mathematics problem solving. It is a phenomenon that may be observed, but not limited to, when an individual i) uses a wrong formula or concept in solving a mathematics problem; ii) fails to use an efficient method of solving on a particular problem; and iii) fails to answer a mathematics problem. Note that lack of mathematical knowledge was not considered as a manifestation of mathematical blindness as the phenomenon is only referring to already known concepts or constructs that an individual failed to perceive.

The researchers made use of the eye-tracking method for collecting data in conducting empirical studies of human perception, cognition, and behavior. As its name suggests, eye-tracking is a means of determining aspects of participant's sensory perception in the visual modality – where they are looking. In this study, the researchers used the Gazepoint GP3 Eye-tracker as a recording instrument for the participants' eye movement which added to the objective analysis of the participants' reasoning,

in pursue of characterizing the different instantiations of mathematical blindness, which is the main objective of this study.

Specifically, this study aimed to characterize different instantiations of mathematical blindness based on participants' eye-movements and their solution or reasoning about a given mathematical problem, and also to determine predominant instantiations and classify participants' solutions or reasoning.

2. Theoretical and Conceptual Constructs

This study is anchored on different theoretical and conceptual constructs that elicit similar characteristics of mathematical blindness.

The properties of System 1 and System 2 according to the *Dual-Learning Theory*, and the *Tripartite Theory* (Stanovich, 2011) show features of the two processes that contribute as to why mathematical blindness occurs and was used in assessing the participants' cognition in solving mathematics problems.

The *Einstellung* which is described as the mechanized state of mind (Luchins, 1942) refers to an individual's tendency to solve a given problem in a specific manner even though better methods exist.

Ben-Zeev and Star's (2001) concept of *spurious correlation* is an event hypothesized to occur when a student perceives a correlation between an irrelevant feature in a problem and the algorithm used for solving that problem, and then proceeds to execute the algorithm when detecting the feature in a different problem.

Reminiscent of the above-mentioned concept is the concept of *intuitive rules* (Tirosh & Stavy, 1999). For instance, a student who thought that "the heavier the object the faster it falls" is said to rely on the "More A – More B" intuitive rule.

Lastly, this research utilized the eye-tracking method which has been, over the years, is increasingly being used in research in mathematics education. (e.g., Chesney, McNeil, Brockmole, & Kelley, 2013; Andra, et. Al, 2013; Miroslawa & Rosiek, 2016; Shayan, et. Al, 2017; Schindler, M., Haataja, E., Moreno-Esteva, E. G., Shvarts, A., & Lilienthal, A., 2018). This method was used to quantify visual attention such as time of fixation and saccades (movement of the eye from fixation on one point to another), and measure perceptual range when an individual is solving a mathematics problem.

3. Analysis Results

3.1 Quadratic Equations

In this task, the participants were asked to solve 6 quadratic equations using any method they know, 5 of which are quadratic trinomials and the last, a quadratic binomial. The questions were presented as slides on the screen one after another. The AOIs (Areas of Interest), namely AOI 0, 1, and 2 are designated for each of the terms for the first five equations, and AOIs 0 and 1 for the last equation.

For equations 1, 2, 3, and 5, the equations contain quadratic trinomials that are not in the form of any special products. Equation 4 contains a perfect square trinomial and equation 6 is in the form of a difference of two squares. The aim of this test is to characterize how the participants approached different forms of quadratic equations and identify whether certain instantiations could be observed to cause difficulty in solving the problems.

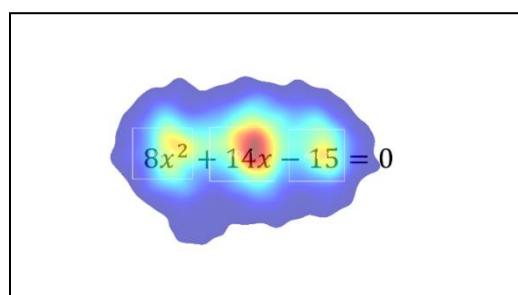


Figure 12. Sample aggregated heat map (equation 1).

Participants' gaze patterns suggest that they give more attention to the middle term when solving quadratic trinomials. The numerical data collected by the eye-tracker also suggests the same as shown below.

Table 1

Recorded data from the eye-tracker for the quadratic equations

AOI Name	Ave Time to 1st View (sec)	Ave Time Viewed (sec)	Ave Fixations (#)	Average Revisits (#)	N
AOI 0 (Leading Coefficient)	4.9642	1.3306	7.2958	6.3758	19
AOI 1 (Middle term)	2.9748	2.6528	13.786	11.1058	19
AOI 2 (Constant term)	6.1278	1.6322	8.5418	7.0908	19

On table 1, AOI 1 has the smallest “average time to 1st view” and the largest “average time viewed”, “average fixation”, and “average revisit”. Further analysis of the gaze patterns, interviews, and written solutions implies that for some of the participants when solving a quadratic trinomial, using factoring as their primary method is more likely. This is evident as the aggregated gaze patterns suggest that the participants used trial-and-error in obtaining the coefficient of the middle term. This means that when students are given a quadratic trinomial, what first comes to mind is to check if the middle term is obtainable thru trial-and-error involved in the factoring method. In this task alone, eight participants used factoring as their primary method in the first four equations.

For this task, the instantiations of mathematical blindness observed are spurious correlation and intuitive rule. First, the spurious correlation for this task was characterized by the participants repeated use of the factoring method or the quadratic formula upon determining whether there is a leading coefficient or not, also upon perceiving whether the coefficients are small or large. Lastly, the intuitive rules that the participants had are the following: (1) since the previous question was solved using the quadratic formula or by factoring, then the succeeding question might also be solved using the same method; and (2) they will use a method that already has become second nature to them when solving quadratic equations.

3.2 Musician Problem

A group of 5 musicians plays a piece of music in 10 minutes. Another group of 35 musicians will play the same piece of music. How long will it take this group to play it?

Figure 13. Slide shown to participants for the Musician problem.

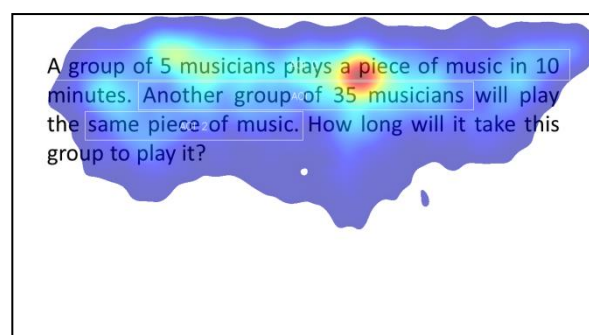


Figure 14. Aggregated heat map for the Musician problem.

AOIs 0, 1, and 2 are designated for the phrases “A group of 5 musicians plays a piece of music in 10”, “Another group of 35 musicians” and “same piece of music”, respectively.

10 out of 18 participants (1 participant’s data was deemed untreatable due to technical issues) used the concept of ratio and proportion in solving the problem where 9 of these participants answered 70, implying that they established a direct proportion between the number of musicians and amount of time to play the music, and 1 participant answered 10/7, who may have established an indirect proportion. These participants who used the same concept twice recognized the feature of the problem:

two units of measurements are identified with a “typical” missing-value proportional question in the end.

Table 2
Data recorded by the eye-tracker for the Musician problem

AOI Name	Ave Time to 1st View (sec)	Ave Time Viewed (sec)	Ave Fixations (#)	Average Revisits (#)	N
AOI 0	1.742	8.006	43.522	29.318	18
AOI 1	1.849	4.441	29.682	24.273	18
AOI 2	3.747	1.832	12	9.762	18

Table 2 shows that the AOI with the least time viewed is AOI 2 which is designated for the phrase “same piece of music”, a plausible cause of why most of the participants did not realize that the problem should not be approached using ratio and proportion. However, some of the participants who gave the correct answer also struggled in realizing that the answer should be 10 minutes. Especially 2 participants, who have shown signs of detachment in reasoning between what is mathematical and realistic. This detachment was elicited when they asked if the problem should be solved “mathematically” or “logically/realistically”.

Again, further analysis of the gaze patterns, interviews, and written solutions tells us that in this case, it is possible that four simultaneous instantiations of mathematical blindness have occurred: the Einstellung effect, spurious correlation, and intuitive rules. First, the Einstellung effect in this problem was characterized by fixation in the two units of measurements that were identified and on the concept of using ratio and proportion as a solution. Second, the spurious correlation was characterized by the overuse of the concept of ratio and proportion upon detecting the two units of measurements. Lastly, the “More A – more B” intuitive rule was evident when the participants established a direct relation between the number of musicians and the amount of time it takes to play the same piece of music.

3.3 Speed Problem

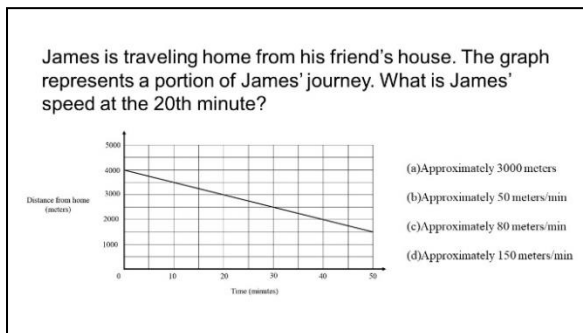


Figure 15. Slide shown to participants for the Speed problem.



Figure 16. Aggregated heat map for the Speed problem.

Note that in this problem, data from only 18 participants were recorded due to technical issues.

The aggregated heat map shows that the participants’ attention was more directed towards the point in the graph where the 3000 meters distance intersects with the 20th minute. Quantitatively, the average number of fixations on that point is 12.409 and with average number of revisits 10.905.

Most of the participants interpreted the problem as a speed problem, where 16 participants read the distance from the graph (3000 meters) that corresponds to the 20th minute, applied the $s=d/t$ formula, and obtained 150 by dividing 3000 by 20. Interestingly, only 2 of the 18 participants selected the correct answer “B” which was based on understanding speed as the ratio of change in distance to change in time.

Since the formula $s=d/t$ is a commonly used formula in solving problems involving speed, it has become a tool that can be inappropriately used in solving mathematical problems most especially when their understanding of the problem or the use of the formula itself is superficial. In this task, the

participants who gave “D” as an answer correlated the use of the formula to the given values or to what was being asked in the problem.

Table 3
Recorded data by the eye-tracker for the Speed problem

AOI Name	Ave Time to 1st View (sec)	Ave Time Viewed (sec)	Ave Fixations (#)	Average Revisits (#)	N
AOI 2	16.102	2.21	12.409	10.905	18
AOI 3	16.672	1.146	7.429	5.8	18
AOI 4	36.186	1.374	6.4	4.833	18
AOI 5	18.555	1.35	7.13	5.818	18
AOI 6	29.954	1.023	7.05	6.053	18

Recall that most of the participants gave “D” as an answer, hence, the AOI designated for choice “D” (AOI 4) should have gained the highest attention. However, the data above show otherwise.

The instantiations of mathematical blindness observed here is spurious correlation. The spurious correlation was characterized by the participants’ overuse of the concept of speed, specifically the use of the formula $s=d/t$ upon detecting that there are two corresponding values of distance and time. Their use of this concept was also deemed superficial because they used the concept, however, in a wrong understanding of the question.

4. Conclusion

Based on the results gathered, this study concludes that mathematical blindness can manifest in students’ mathematics problem-solving and reasoning in different ways. These manifestations can be caused by objects (stimuli) in a problem or by misconceptions. These were observed especially when a participant overuses concepts upon detecting a certain feature in the problem. It is also conclusive that through the use of the eye-tracker, the analysis of their reasoning became more objective with the help of the quantifiable data and visualizations provided. This study led to the realization that in solving mathematical problems, it is not just reasoning alone that influences how a problem is approached, but also by objects that can be perceived in the problem itself.

References

- Andrá, C., et. Al (2015). Reading mathematics representation: an eye-tracking study. *International Journal of Science and Mathematics Education*, 13(suppl 2), 237-259. <https://doi.org/10.1007/s10763-013-9484-y>
- Ben-Zeev, T., & Star, J. (2001). Spurious Correlation in mathematical thinking. *Cognition and Instruction*, 253-275.
- Chesney, D.L., McNeil, N.M., Brockmole, J.R. et al. Mem Cogn (2013). An eye for relations: eye-tracking indicates long-term negative effects for operational thinking on understanding of math equivalence. *Memory and Cognition*, 41(7), 1079-1096. <https://doi.org/10.3758/s13421-013-0315-8>
- Luchins, A. S. (1942). Mechanization in Problem Solving: the effect of Einstellung. *Psychological Monograph*, 248.
- Mirosława Sajka, Roman Rosiek (2016). Solving a problem by students with different mathematical abilities: A comparative study using eye-tracking. Konrad Krainer; Nad’ Vondrová. CERME 9 – Ninth Congress of the European Society for Research in Mathematics Education, Feb 2015, Prague, Czech Republic. pp. 1752-1758, Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education.
- Schindler, M., Haataja, E., Moreno-Esteva, E. G., Shvarts, A., & Lilienthal, A. (2018). Eye-tracking in mathematics education research: A follow-up on opportunities and challenges. Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education, Sweden: Umeå.
- Shayan, S., et. Al (2016). Eye-Tracking the emergence of attentional anchors in a mathematical learning tablet activity in C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 1-370). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-1005-5
- Tirosh, D., & Stavy, S. (1999). Intuitive rules: A way to explain and predict students' reasoning. *Educational Studies in Mathematics*, 51-66.

How to Measure the Collaborative Problem-solving Competency Based on Conversational Agent

Ge Wang ^a, Bian Wu ^{a*}, Yiling Hu ^a

^a *Department of Educational Information Technology, East China Normal University, China*

* *bwu@deit.ecnu.edu.cn*

Abstract: Nowadays, collaborative problem-solving competency has become one of the essential core skills for talents in the 21st century. Limited by the complex and variable factors of collaborative problem-solving activities, the current assessment method was difficult to understand the real situation of students' collaborative activities and problem-solving. Based on the PISA2015 collaborative problem-solving competency assessment framework, this study build an assessment model and integrated the computer conversational agent technology to develop the prototype of the collaborative problem-solving competency evaluation system. We found the performance level of individual and teams' collaborative problem-solving patterns. It provides a reference for the evaluation of collaborative problem-solving competency, and also provides a new direction for human-computer interaction based on conversational agents.

Keywords: Collaborative Problem-solving; Conversational Agent; Assessment Research

1. Introduction

With the development of modern society and the constant change of knowledge, the division of labor becomes more and more refined. Many projects need to be completed by people from different fields and majors through communication and collaboration. Nowadays, collaboration ability has become a necessary competency for talents with social competitiveness.

Collaborative problem-solving (CPS) competency as a composite skill of collaboration ability and problem-solving ability, its cultivation and evaluation have received extensive attention. UNESCO's report on "Learning: The Treasure Within" highlighted the importance of "teamwork competency" (Delors, 1998); The Singapore Ministry of Education (2017) added a "Project work" to the school curriculum, which clearly lists the four skills of knowledge application, communication, collaboration and independent learning as the learning outcomes; The 21st Century Skills Alliance defined problem-solving and collaboration as skills that every American teenager needs to learn at school (Voogt & Roblin, 2012). Scholars from various countries pay more attention to how to set the situation of CPS activities to cultivate students' CPS competency. However, in terms of assessment, it is limited by the variety and complexity of the influencing factors of CPS ability. It is difficult to measure the actual collaboration situation and problem-solving behavior in this environment. In fact, CPS skills are often seen as skills that are difficult to measure (Stecher & Hamilton, 2014). Collaborative activities and problem solving activities are non-linear sequence activities. It becomes a challenge to adopt an effective scheme to model and quantify this complex system.

2. Literature Review

2.1 Concept and Evaluation Framework of Collaborative Problem-solving

Related research on CPS stems from the social attributes of problem solving (Vygotskie, Embong, & Muslim, 1978) . The current accepted concept of CPS competency comes from the OECD's definition in PISA 2015 (OECD, 2017), "the capacity of an individual to effectively engage in a process whereby

two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution.” The existing research on CPS has accumulated a lot of theoretical frameworks and models. Stevens and Campion (1994) proposed a teamwork conceptual framework that included five elements: conflict resolution, collaborative problem solving, communication, goal setting and performance management, planning and task collaboration. Griffin (2011) defined the ATC21S framework, which divided the CPS competency into two dimensions: social skills and cognitive skills. In 2017, the OECD released the PISA2015 Collaborative Problem solving Framework, which is a framework matrix consisting of three core social skills and four traditional personal problem-solving stages (OECD, 2017).The evaluation framework of this paper is based on the PISA2015 framework.

Table 5

Matrix of collaborative problem-solving skills for PISA 2015 (OECD, 2017).

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(A) Exploring and understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve the problem
(B) Representing and formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describe roles and team organisation (communication protocol/rules of engagement)
(C) Planning and executing	(C1) Communicating with team members about the actions to be/being performed	(C2) Enacting plans	(C3) Following rules of engagement, (e.g. prompting other team members to perform their tasks)
(D) Monitoring and reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organisation and roles

2.2 Conversational Agent

In CPS activities, collaborative problem solving competency are susceptible to a variety of factors, such as situational complexity, partner behavior, and so on. The human-computer interaction evaluation model based on conversational agent can be used to reduce the impact of uncertainty factors. The conversational agent combines computer and conversations to collect log data through a broader interaction with the individual or meaning, and is able to assess the individual's ability to acquire knowledge, skills, and language at the cognitive and non-cognitive levels (Jackson & Zapata, 2015). Compared with the traditional evaluation methods, the human-computer interaction evaluation mode makes the performance of individual ability more stable (Graesser, Chipman, Haynes, & Olney, 2005). On the whole, conversational agent technology can provide evidence and help for formative assessment of CPS competency assessment.

To this end, this study aims to answer the following research questions:

Research question 1: how to design conversational agent to simulate real collaborative problem-solving?

Research question 2: how to measure collaborative problem-solving competency based on conversational agent?

3. Model Design

3.1 Evaluation model

PISA2015 provides individuals with complex task situations and further tests their CPS ability, which involves not only the reasoning process of the tested individuals in the problem situation, but also their communication, management and social process in the collaborative environment. Based on the

research of collaborative problem solving, we proposes the following evaluation framework, as shown in Figure 1.

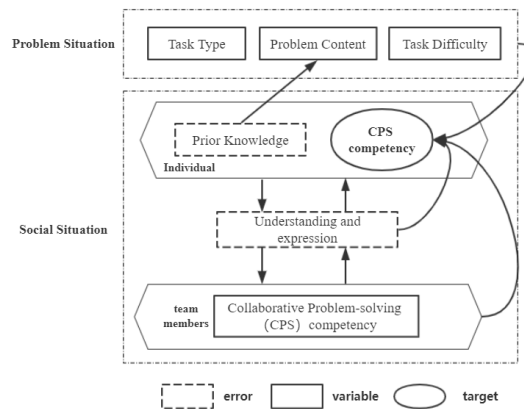


Figure 17. Evaluation Framework of CPS Competency

the CPS evaluation framework should consist of two parts: the problem situation and the social situation. The problem situation includes three variables, task type, problem content, and task difficulty. (1) Task type: It determines the ability tendency of individual in the current problem situation. (2) Problem content: It determines the limitation of the ability to examine the current situation, including content-related problem and content-independent problem. (3) Task difficulty: It is the comprehensive evaluation result of the task. In social situation, Individual CPS skill is influenced by other members of team. we can simply divide them into: external self-expression and individual acceptance to understand the expression of others.

3.2 Logic Design and Architecture Design of Conversational Agent

Figure 2 shows the architectural design of the Collaborative Problem Solving Supported Conversational Agent (CP2SCA). Which aims to simulate the collaborative performance of the tested individuals in the real problem situation with other members of the team by embedding the dialogue manager module.

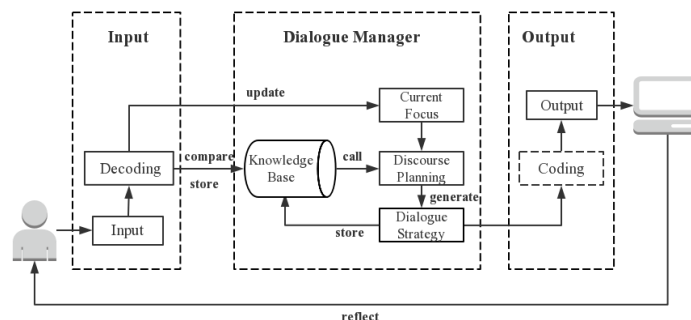


Figure 18. Architecture of Collaborative Problem Solving Supported Conversational Agent (CP2SCA)

The logic of the CP2SCA is: First, when the user inputs the information, the system decodes the input content, stores it into the database, and records the decoded content into the current focus module, updating current problem solving state of user utterance content, implements the tracking of the conversation state. Then, the conversational agent completes the planning of the discourse according to the current focus, the record of the existing information of the database and the input content, and formulates a suitable dialogue strategy coding scheme. Finally, the dialogue strategy is coded as the language of human-computer communication, and the output is the reflection content of the agent to the user.

4. Method

4.1 Experimental Tool Design

In addition to the basic user information collection functions (login interface and task information introduction page), the system focuses on the CPS activities. The main interface and functions are shown below: The system includes three interfaces: experimental control, chat dialogue, and history record. Experimental control is the main operational interface of the system, as shown in Figure 3a. Chat dialogue is a dialogue communication interface between the subject and the conversational agent.(Figure 3b). The history interface (Figure 3c) stores all the experimental and conversation records of the subject and the agent. The subject can view the experimental records on the interface, or forward them to the conversational agent companion to change the current focus status of the team discussion. The problem situation set by the assessment system is “Exploration of New Species”: the scientific laboratory has collected samples of unknown attributes. Participant should cooperate with colleagues (conversational agent), the experimental exploration is carried out, and finally the characteristics of “unknown samples” need to be clarified. During the experiment, the system will collect the number of discourse rounds (number of discourse sets), dialogue rounds (number of statements), step size (number of individual speeches and mouse clicks), performance (number of correct conclusions), etc.

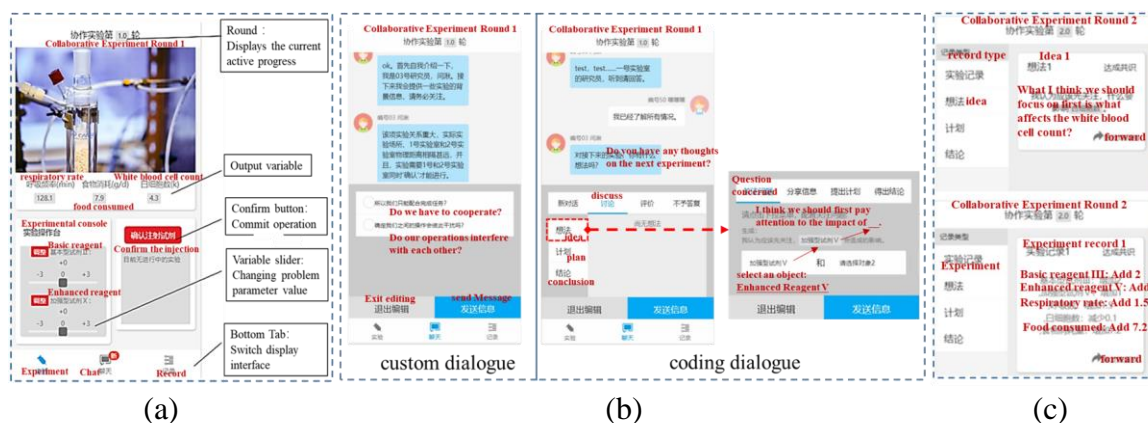


Figure 3. (a) Experimental Control Interface. (b) Chat Dialogue Interface . (c) History Record Interface

4.2 Participant and Process

Since the CPS competency is an indispensable ability for a individual to enter the society from the school, he participants of the experiment are 30 graduating students from a university in East China, aged between 24 and 26. Participants completed CPS tasks by collaborating with the conversational agent in the test environment. Then they filled the survey, which included the technical acceptance scale and open questions.

5. Result

5.1 Validity Test of Prototype Tools

The results show that the technical acceptance is $M = 5.55$, $SD = 0.78$. It can be seen that the technical acceptance of the evaluation system is higher. The answer to the open question indicates that participants generally have a higher degree of recognition of the prototype tool. As a result, prototyping tools can basically simulate real-world CPS activities and provide users with basic interactions for self-expression.

5.2 Analysis of CPS Behavior Characteristics

Behavioral characteristics describe the behavioral sequence preferences exhibited by individuals in the process of collaborative problem-solving activities, while the behavioral characteristics exhibited in individual collaborative problem-solving activities have not been standardized. According to the characteristics of its sequence, We can use the number of rules explored in each stage and the dialogue rounds between stages to construct a directed graph, where the output behavior is nodes and the

feedback behavior is steps. It can integrate nonlinear and unstructured process data into an intuitive structured model.

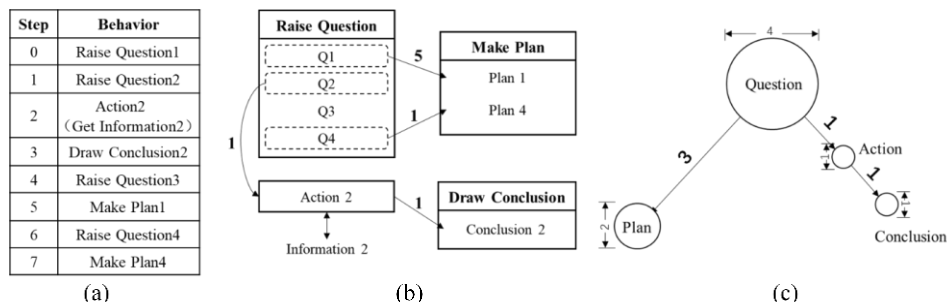


Figure 4. (a) Unstructured Team Behavior Data. (b) Unstructured Data Collation (Note: The number marked between the behaviors is the step size). (c) Directed Graph of Team Behavior Characteristics

For example, the non-structural data of team behavior collected by the system is shown in Figure 4a. The data is classified according to the four stages of problem solving to obtain Figure 4b. The step size between the stages is the interval value of the number of steps involved. Since the relationship is complicated and difficult to express, it can be simplified by the directed graph shown in Figure 4c. The radius of the circle is the total number of consensus reached at this stage; the length of the line is the average step size, and when there are multiple lines of different step length between the two elements, the mean value is taken.

5.2.1 Analysis of Individual Behavior Patterns

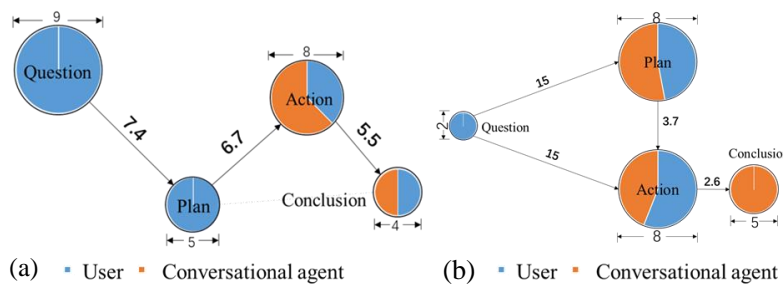


Figure 5. (a)Team09 (b)Team02 behavior characteristics and individual contribution

According to the proportion of individual output behavior in the teamwork CPS behavior, we can visually see the team's CPS behavior characteristics, as well as the individual's contribution in each stage of CPS, and understand the individual's behavior in the group CPS process. Taking sample 09 and sample 02 as an example, a graph of individual output behavior characteristics generated by adding a pie chart to a directed graph is shown below.

From the perspective of individual contribution of sample 09, the contribution output of participant 09 is 100% in the proposed stage and the planning stage. All questions and plans are formulated by the participant. There is no collaboration in these two phases. In the Action phase, five of the eight operations followed the plan, three were unplanned. In the conclusion phase, the participant's contribution rate is 50%. Participant 02 is dominant in the Question stage. During the Plan and Action stage, the contributions of the participant and the conversational agent are more balanced, but the final conclusions are all summarized by the conversational agent.

On the whole, participant 09 had a strong sense of leadership in the CPS process, but his ability to execute was not good, and he was not good at following plans. He need further develop his ability to collaborate with others. Participant 02 is more inclined to the role of leader in the CPS processHis ability to execute is quite good, but his ability to summarize is not strong. His team collaboration ability is strong, but problem-solving ability is slightly weaker.

5.2.2 Analysis of Team Behavior Patterns

Combined with the established CPS characterization model, we found two behavioral patterns that emerged from the team in CPS activities: problem thinking and plan making.

Teams with problem-thinking behavior pattern, their CPS path starts from the question stage, and externalizes the thinking at the Question stage. The team's thinking is active, and it is possible to consider many questions based on the task requirements, and then make plans to draw conclusions. Due to the large number of questions, it is easy to appear in the CPS activities that the plans are difficult to correspond to the questions one by one. As shown in Figure 5a, Team 09 has a much smaller number of plans than the number of questions raised. They are not effective problem-solving behaviors. At the same time, the step length between the question stage and the plan stage is the largest (7.4). It can be found that the team consumes more dialogue communication between the analysis of the task requirements and the development of the corresponding plan. Teams with planned behavior patterns, their CPS path starts from the plan stage, and the thinking externalization is shared with people in the plan stage. In the question stage, the team's thinking is not divergent, but it is better at turning questions into more specific plans, making plans more enforceable. Therefore, the number of plans making based on the questions is much larger than the number of questions. Compared to the problem-thinking pattern, the lines between the “plan-action-conclusion” are shorter and the connections between the three are closer.

6. Conclusion

The study proposes an assessment framework for CPS competency. Based on the existing concept of dialogue management, we developed a CPS competency assessment system, which can simulate the real CPS situation. We design a context-based computer conversational agent to encode unstructured discourse into structured data to achieve "human-machine dialogue". The conversational agent collects procedural data when collaborating with individuals to complete tasks. By visualizing the process data, we can clearly understand the behavioral characteristics of individuals and teams. We can develop evidence-based strategies to improve individual CPS competency. The system simplifies the assessment method, ensures the feasibility of technical application, and the standardized CPS process analysis model has certain promotion value.

References

- Delors, J. (1998). *Learning: The treasure within*. Unesco.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618.
- Griffin, P., & Care, E. (2015). The ATC21S method. In *Assessment and teaching of 21st Century Skills* (pp. 3-33). Springer, Dordrecht.
- Jackson, G. T., & Zapata-Rivera, D. (2015). Conversation-based assessment. *R&D Connections*, 25, 1-8.
- OECD (2017), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/9789264281820-en>.
- Singapore Ministry of Education (MOE). (2017). Retrieved May 12, 2019, from: <https://www.moe.gov.sg/education/programmes/project-work>
- Stecher, B. M., & Hamilton, L. S. (2014). *Measuring Hard-to-Measure Student Competencies: A Research and Development Plan*. Research Report. RAND Corporation. PO Box 2138, Santa Monica, CA 90407-2138.
- Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of management*, 20(2), 503-530.
- Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of curriculum studies*, 44(3), 299-321.
- Vygotskie, L., Embong, A. R., & Muslim, N. (1978). *Mind in society: The development of higher psychological*. info: Cambridge: Harvard University Press, 1978.

Introduction of Educational Technology Engagement Model

Ibrahim Hakki BULUT^{a*} & Omer DELIALIOGLU^b

^a *PhD Student, Istanbul Medeniyet University, Turkey*

^b *Prof. Dr., Middle East Technical University, Turkey*

*ibrahim.bulut@medeniyet.edu.tr

Abstract: This paper proposes an educational technology adoption model considering the limitations of traditional technology adoption models. Based on the model, the authors develop a questionnaire, in which the items are derived from previous technology adoption studies, to identify and indicate the constructs in the proposed model. The paper presents the statistical validity and reliability of the questionnaire. The results based on Exploratory Factor Analysis (EFA) indicate that the items on the questionnaire that are meant to measure the constructs in the model are statistically valid and reliable.

Keywords: Educational technology, technology adoption model, TAM, technology engagement

1. Introduction

With the change of Web 2.0 technologies in the late 90s, Internet dynamics changed dramatically. Users have become more active participants and content creators thanks to Web 2.0 tools made it easier and more accessible. Web 2.0 not only allows individuals to benefit from technologies, systems, and tools on an instrumental level, but also allows them to engage in social and cognitive involvement, and hedonic immersion through various tools.

The development in technology does not only affect the dynamics of everyday life, but also affects the dynamics of education. A recent study indicates that smartphones are used by 87% of US students, 74% of them have at least a computer, and 41% have access to a tablet in which every device has access to Internet, and at least one social media platform is followed regularly by 97.5% of US youth (Villanti et al., 2017). High adoption of Web 2.0 technologies has guided teachers to benefit from them for instructional purposes. Integration of new technologies into schools and classrooms is no longer limited with entrepreneur instructors. E-mail use to communicate between students and instructors, material delivery through system tools, usage of Office programs to create and demonstrate content, and searching Internet to access specific knowledge are regarded as common technology-mediated teaching practices in the last two decades (Chen et al., 2010; Laird & Kuh, 2005). In addition, there are many popular educational applications and systems highly adopted by both academicians and teachers to engage students in learning activities (Elmas & Geban, 2012).

Information and Communication Technologies (ICT) in education necessitate specific focus in terms of acceptance and engagement since they become an integrated component of instruction. The aim of educational technologies, a special form of ICT, is to improve learning, strengthen educational processes and increase students' success. Achieving these objectives requires the systematic implementation of theory to educational practices from the relevant studies especially from motivation and engagement studies. Therefore, the current study finds it valuable to define motivators that anticipate the adoption of technology, and identify and detail the likely relationships among these motivational constructs. Taking into account all of these, the current study's purpose is to offer a new model called Educational Technology Engagement Model (ETEM) by considering the changing dynamics and paradigms over technology, current role of technology over society and anticipated influences on education. Another aim of the study is to test a questionnaire composed of several scales derived from previous studies in the technology adoption and acceptance literature and to decide if the questionnaire is statistically powerful to be able to be utilized in prospective studies on which ETEM is to be based.

2. Literature Review

It has been a challenge for technology acceptance model practitioners for 3 decades to illuminate the underlying motivators to understand why some of the technologies are excessively embraced by the Internet community (Davis, 1985; Bagozzi, 2007). Technology acceptance models can be defined as contextual scaffolds extracted from motivational theories to portray user adoption of technologies. They try to illuminate user participation based on the demographics, characteristics and perceptions of the intended population. Technology adoption models provide a comprehensive framework by identifying what needs to be satisfied as a prerequisite to adopt technology, and how motivators and technology use behaviors are interrelated.

Technology Acceptance Model (TAM) (Davis, 1985) is the most popular adoption model among several others since it is the most referenced, applied, modified, extended, and criticized one in the technology acceptance literature. TAM suggests that users are exposed to 2 motivational beliefs - Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) - when interacting with a new technology. These beliefs determine their adoption decisions. PU refers to perceived advantages emanating from utilizing of a specific technology. PEOU refers to individual assessment of to what extent the utilized technology is easy to use (Cole, 2009).

The model predicts that external factors as independent factors affect PU and PEOU as motivational beliefs. Both PU and PEOU are the estimators of the intention to use to predict the adoption of technology. The parsimonious structure has risen it to a leading model to predict user motivation to accept technology use. It has been repeatedly replicated, implemented and tested for different systems, tools and technologies (Chutter, 2009). However, the basic structure of the model has been criticized several times, leading to efforts to expand by researchers to make it more descriptive.

Although the accuracy and strength of TAM have been tested and verified in many studies, it has been severely criticized several times in terms of several aspects including parsimonious structure and lack of intrinsic motivators (Bagozzi, 2007; Silva, 2007; Smith et al., 2007; Lee, Kozar, & Larsen, 2003). TAM has long been functioning as a theoretical framework that aims providing a single solution for various technologies to explain user motivation of their adoption. However, the parsimonious structure of TAM by regressing to only 2 factors as PU and PEOU to explain adoption causes to dismiss the genuine latent variables that are presumably to refer to technology acceptance problems (Lee et al., 2003). Moreover, ignoring intrinsic factors, which are likely to explain intrinsic motivation via addressing cognitive, social and hedonic needs of humans, is another criticized aspect of TAM resulting in lack of useful information to understand technology adoption of 21st users (Bagozzi, 2007). Lastly, TAM and its derivatives disregard the ever-changing direction and progress in the relationship between technology and society. Although technology is seen as an apparatus accepted for its benefits in the late 80s and early 90s, human interplay with technology have increased to a more advanced level of involvement and commitment at the social and cognitive level after 2000s. Considering the increasing use of internet and social media and the availability of smart devices, society is far beyond the initial acceptance of technology. Although they are sufficient to explain what motivators can predict to make technology a part of people lives, they are weak in explaining how individuals are involved in technology and are part of it. In addition, the behaviorist characters of TAM focusing mostly on systematical components and ignoring the systemic and social components of technology (Smith et al., 2007) presents a partial explanation of the acceptance of technology. As a result, traditional technology acceptance models might benefit from being replaced and updated with new motivational constructs and determinants to overcome aforementioned limitations.

3. Methodology

Existing technologies and systems offer distinct characteristics with systemic, social and technical elements making them unique. Considering the complexity of technology and a broad array of technologies fulfilling several needs, attempts to identify user acceptance grounded on a theoretical framework is oversimplification. Therefore, Smith and his colleagues (2007) recommend proposing contextual models through process-based analysis. Based on the recommendation, analyzing the underlying motivational process toward adopting and using a specific technology through the lens of one or more theory of motivation is more appropriate rather than directly suggesting a more extended and excessive model without a theoretical justification.

As the first step, the appropriate motivational theory was determined by considering specific technology's idiosyncratic structure and the circumstances in which it is being implemented. Both intrinsic and extrinsic motivation play an essential role on the adoption of educational technologies therefore Self-determination Theory (SDT) was decided to be grounded because SDT considers both intrinsic and extrinsic motivators over the engagement of students (Ryan & Deci, 2000). Basically, SDT defines two types of motivation as extrinsic and intrinsic. Extrinsic motivation refers to the motivation of individuals who are arisen and fed from sources outside the scope of the task and independent from the activity. Intrinsic motivation can be defined as self-acceptance and inherent tendency to engage in a specific behavior to achieve the intended results. (Ryan & Deci, 2000). Considering the relevant constructs in SDT, a theoretical model – Educational Technology Engagement Model (ETEM) was drawn as can be seen in Figure 1. Technology-related factors derived from previous adoption studies are mounted into the model to identify relevant motivational constructs of STD.

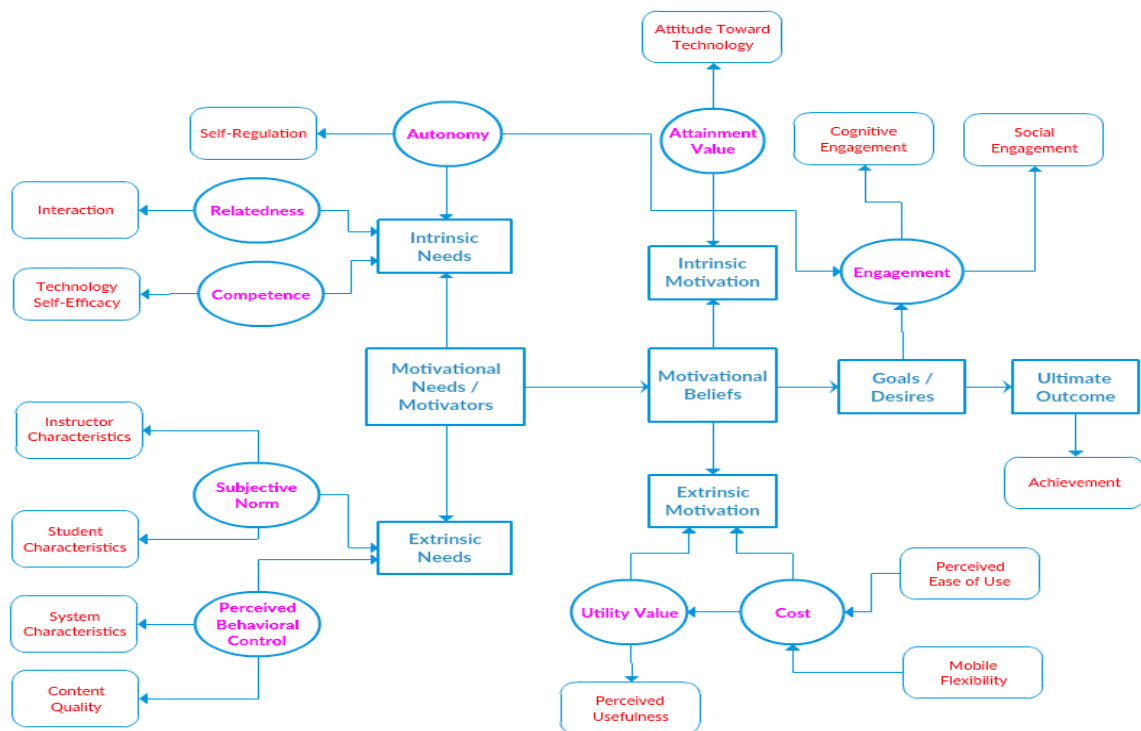


Figure 1. ETEM – A Theoretical Model for Educational Technologies

To test the model being constructed, Edmodo which defines itself as a social learning platform was selected since it is considered an ideal platform as a self-sustained learning environment with its tools and facilities rather than being a complementary technology making instruction and learning easier. Based on the chosen Educational Technology, the population of the study was determined as students at Turkish universities where Edmodo is used actively as the learning environment in Introduction to Information Technologies and Applications (IITA) course taught throughout Turkish universities as a mandatory course. The reason for selecting this course is due to enabling instruction to occur without requiring face-to-face instruction and due to expecting higher engagement and participation of students. Accordingly, 4 instructors who utilize Edmodo intensely for educational purposes by using every tools and utilities of Edmodo to instruct IITA at 4 Turkish universities, namely Amasya, Harran, Ege and Middle East Technical Universities, and their students are selected as samples of the study. Thus, 520 students from 4 universities became conclusive participants of the study.

To be able to base ETEM as a theoretical model in prospective studies, a questionnaire was composed of various scales whose items are extracted from past technology acceptance studies. The scales aim to identify and measure relevant constructs within the motivational concepts in the model,

namely motivational needs, beliefs, goals and ultimate outcome. Some of the sample items, factor names where these items point to, and original sources of the items can be found at Table 1.

Table 1

Sample Items and Sources

Factor	Sample Items	Sources
Content Quality	The contents shared in Edmodo is relevant to the course	Roca et al. (2006)
Technology Attitude	Students learn better in courses where computers are actively used	Arslan (2006)
Social Engagement	Edmodo enabled students to share their feelings and opinions easily	Paechter et al. (2010)
Interaction	My instructor encouraged me to use Edmodo	Soong et al. (2001)
Self-regulation	I am able to learn at my own pace	Lee & Tsai (2011)

The online version of the scales was delivered to the instructors share with participant students at the end of 2016-2017 Fall semester. The instructors shared the survey with students and sent them reminders to complete it. 520 students in total filled out the scales. After data collection, the validity and reliability of scales were checked through Exploratory Factor Analysis and Reliability Analysis.

4. Results

The purpose of the study was to discover latent variables for technology adoption by gathering and analyzing data. Exploratory Factor Analysis (EFA) was used in the analysis of the research data. EFA was conducted to indicate whether expected latent variables are appropriately obtained and the instrument satisfies the required conditions for validity and reliability. 520 students from 4 universities became participant for the study. The survey instrument consists of 84 items. Responses were on a 5-point scale, ranging from “Strongly Disagree” to “Strongly Agree”.

Firstly, outliers and missing data were inspected to proceed for further analysis. Box-plots and z-scores were examined to identify outliers. 42 data points with standardized scores above 3.29 and positioned outside the fences of the boxplots were identified as outliers thus 42 cases were excluded from further analysis. Consequently, assumptions were analyzed based on the data collected from 478 samples. 84 scale items were firstly examined for the factorial structure. Correlation matrix indicated that all of the items are correlated ($> .30$) with at least one other item therefore a reasonable factorial structure can be proposed (Hair et al., 2010). As can be seen in Table 2, the KMO value was .96, above the recommended value of .60, and Bartlett’s test of sphericity was significant ($p < .05$). Univariate normality assumption was controlled by looking at descriptive statistics and through statistical kurtosis and skewness values of the variables. The results met the required interval (-3/3) indicating univariate normality (Tabachnick et al., 2007).

Table 2

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.96
Bartlett's Test of Sphericity	Approx. Chi-Square	31463.38
	df	3486
	Sig.	.00

The initial eigenvalues showed that 13-factors solution was suitable for adoption because first 13 factors have eigenvalues greater than 1.0 and they explained 67.37% of total variance which is above the recommended value of 60% in social sciences (Hair et al., 2010).

The pattern matrix indicated that all items contributed to one of the 13 factors and there was no significant relationship among factors. In order to determine whether the item is loaded sufficiently on the relevant factor, a minimum criterion value of .30 was adopted for factor loading (Tabachnick & Fidell, 2012). According to the criteria, 5 items with loadings below .30 were removed from the scale.

Reliability Analysis was conducted to determine if every scale in the questionnaire functions accurately and internally consistent. Cronbach alpha score was calculated for every factor and internal consistency of each obtained factor was analyzed using Cronbach's alpha value. The acceptable range for Cronbach alpha test of reliability is .70 or above, and .80 or greater is preferred (Cortina, 1993). Cronbach's Alpha Values for each factor is shown in Table 3. Alpha values are above .80 indicating that internal consistency for every factor are beyond acceptable. It was also investigated that if any of the items were deleted, alpha scores would increase significantly. Since Cronbach's alpha values for the factors did not decrease when 4 items were deleted, these items were found to be incompatible with the related factors. All 4 items were dropped from the scales. As the final result, 9 items were removed from the questionnaire and 75 items remained for further analysis in prospective studies.

Table 3

Cronbach's Alpha Values

Factor	Cronbach's Alpha	Factor	Cronbach's Alpha
Perceived Usefulness	.92	Perceived Achievement	.89
Self-efficacy	.82	Student Characteristics	.92
Cognitive Engagement	.93	Interaction	.92
Content Quality	.88	System Characteristics	.87
Technology Attitude	.90	Mobile Flexibility	.92
Perceived Ease of Use	.93	Self-regulation	.89
Social Engagement	.90		

5. Conclusion

The technology adoption models, especially TAM, have guided researchers over 30 years to explain how motivators have influenced human usage of technology. They grant a broad and inclusive lens for understanding the relationship between the needs and the beliefs of users and the objectives and outcomes of using technology. The appropriateness of TAM is questioned considering the ever-changing direction and progress beneath the motivation to adopt and use 21st century technologies. Taking the limitations into consideration, the study introduces a new approach to develop technology models, requiring a theoretical understanding of the relevant motivational structures and processes involved while adopting a new technology. Based on the approach, Educational Technology Engagement Model (ETEM) is proposed considering that current adoption models are not ideally suited to identify adoption of educational technologies due to their disregard of intrinsic factors. The study also developed a comprehensive questionnaire in which the items are derived from various sources to identify and indicate the constructs in the model. The validity and reliability of the items in the questionnaire are proved through EFA and Reliability Analysis. It is desired that both information systems experts and technologists in general, and especially educational technology practitioners and researchers, take advantage of ETEM and help the authors of current study improve and update the proposed model through their criticism and feedback, and also further statistical analyses including Confirmatory Factor Analysis and Structural Equation Modeling.

References

- Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of management education*, 24(1), 32-54.
- Arslan, A. (2006). Bilgisayar destekli eğitim yapmaya ilişkin tutum ölçeği. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 3(2), 24-33.
- Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the association for information systems*, 8(4), 3.
- Chen, P. S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers & Education*, 54(4), 1222-1232.
- Chuttur, M. Y. (2009). Overview of the technology acceptance model: Origins, developments and future directions. *Working Papers on Information Systems*, 9(37), 9-37.
- Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers & education*, 52(1), 141-146.
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS quarterly*, 189-211.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology).
- Elmas, R., & Geban, Ö. (2012). Web 2.0 tools for 21st century teachers. *International Online Journal of Educational Sciences*, 4(1), 243-254.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. (2010). *L. (2010). Multivariate data analysis*. Pearson.
- Laird, T. F. N., & Kuh, G. D. (2005). Student experiences with information technology and their relationship to other aspects of student engagement. *Research in Higher Education*, 46(2), 211-233.
- Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for information systems*, 12(1), 50.
- Lee, S. W. Y., & Tsai, C. C. (2011). Students' perceptions of collaboration, self-regulated learning, and information seeking in the context of Internet-based learning and traditional learning. *Computers in human behavior*, 27(2), 905-914.
- Liaw, S. S., Huang, H. M., & Chen, G. D. (2007). Surveying instructor and learner attitudes toward e-learning. *Computers & Education*, 49(4), 1066-1080.
- Paechter, M., Maier, B., & Macher, D. (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & education*, 54(1), 222-229.
- Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).
- Pituch, K. A., & Lee, Y. K. (2006). The influence of system characteristics on e-learning use. *Computers & Education*, 47(2), 222-244.
- Roca, J. C., Chiu, C. M., & Martínez, F. J. (2006). Understanding e-learning continuance intention: An extension of the Technology Acceptance Model. *International Journal of human-computer studies*, 64(8), 683-696.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54-67.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68.
- Smith, S., Johnston, R. B., Shanks, G., & Rahim, M. M. (2007). Toward a motivational theory of technology implementation processes. In *European and Mediterranean Conference on Information Systems*. Polytechnic University of Valencia, Spain.
- Soong, M. B., Chan, H. C., Chua, B. C., & Loh, K. F. (2001). Critical success factors for on-line course resources. *Computers & education*, 36(2), 101-120.
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics (Vol. 5)*. Boston, MA: Pearson.
- Villanti, A. C., Johnson, A. L., Ilakkuvan, V., Jacobs, M. A., Graham, A. L., & Rath, J. M. (2017). Social media use and access to digital technology in US young adults in 2016. *Journal of medical Internet research*, 19(6).
- Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of management journal*, 40(6), 1282-1309.

Cognitive Group Awareness Tools: versatile devices to guide learners towards discrepancies

Lenka SCHNAUBERT^{a*}, Sven HEIMBUCH^a, Melanie ERKENS^a & Daniel BODEMER^a

^a*Media-based Knowledge Construction, University of Duisburg-Essen, Germany*

*lenka.schnaubert@uni-due.de

Abstract: Collaborative practices cover a vast variety of contexts and educational goals. Despite these differences, most means of support draw on between- and within-learner discrepancies as driving forces of individual and collaborative learning. These discrepancies are a focus of cognitive Group Awareness Tools, that process knowledge-related learner data and feed it back to the group to ease the interpretation of the learning situation thereby guiding collaborative learning activities. In this paper, we examine the features and different characteristics of these tools. Based on three exemplary cognitive Group Awareness Tools focusing on different types of knowledge discrepancies, we explore how data processing is adjusted to different settings and discuss whether cognitive Group Awareness Tools are suitable devices to be deployed throughout various educational contexts.

Keywords: Cognitive Group Awareness Tools, data processing, discrepancies, guidance

1. Discrepancies as a driving force of collaborative learning

Collaborative practices reach from small group interaction to mass collaboration, cover a vast variety of different contexts, such as schools, universities, or leisure time, and are used for very different educational goals, like acquiring domain knowledge or collaboration skills. However, learners need to overcome various challenges to successfully regulate their learning processes. Particularly, they have to register relevant internal and external conditions (e.g., their state of knowledge or their resources at hand) to validly interpret the learning situation and decide upon appropriate courses of action (Winne & Nesbit, 2009). Within collaborative learning settings, other learners are part of this learning situation and thus, learners need to be aware of each other's relevant cognitive states to interpret the collaborative situation appropriately (Bodemer, Janssen, & Schnaubert, 2018).

A central aspect relevant to regulating learning is the identification of discrepancies between current and desired states. These may be caused by information that is not compatible with the learners' existing cognitive structures (causing what Piaget called "disequilibrium"; Piaget, 1977), but also by learners experiencing underachievement in terms of their own learning goals when metacognitively monitoring their performance against their internal standards (e.g., Winne & Hadwin, 1998). Both types of discrepancies may activate resolution processes and activities to overcome them and while they are located within the individual, they can be provoked by external information in the social environment. From a group-level perspective, both discrepancies (between incompatible content or the extent of knowledge of learners) may also activate resolution processes regardless of a learner experiencing an intra-individual need for resolution, because learners may need to establish a shared knowledge base or common ground to collaborate (Clark & Brennan, 1991). This group perspective can also lead to the situation where learners may not even relate group-level information to themselves but identify discrepancies between other learners' cognitions and may be inclined to support resolution regardless of self-involvement. This is especially the case with controversies, which are incompatible assumptions existing within groups or communities (Johnson, Johnson, & Smith, 2000). Thus, within collaborative learning environments, we can distinguish between purely internal discrepancies, internal-external discrepancies, and external discrepancies, which ties in closely with existing definitions of cognitive conflicts, that include internal-internal, internal-external, but also external-external conflicts (Lee & Kwon, 2001). These discrepancies may be due to differences between incompatible content constituting

some kind of conflict or controversy (a qualitative difference in ideas) but may also be due to differences between the extent of knowledge (a quantitative difference in knowledge) (see Figure 1).

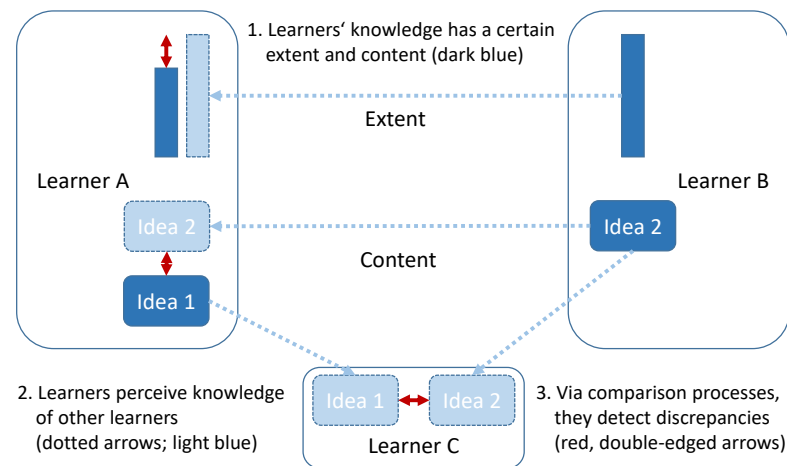


Figure 1. Discrepancies as perceived by learner A (internal-external) and C (external)

Collaborative situations may be beneficial for learning, especially when discrepancies occur. But the occurrence of discrepancies in itself is not sufficient to activate beneficial learning processes, learners also need to become aware of these discrepancies (de Vries, Lund, & Baker, 2002). From a self-regulation perspective, learners have to register and validly interpret the learning situation, before they can conduct beneficial learning processes (Winne & Nesbit, 2009). Supporting these processes is often referred to as “guidance”. Guidance may include activating or supporting favourable behaviours or processes, but also focussing activities on specific content (Bodemer, 2011). Thus, in this paper, we will take a closer look at one group of tools that seem particularly suited for this kind of instructional support: Cognitive Group Awareness Tools (cGATs).

2. Cognitive Group Awareness Tools

A key for learners to benefit from discrepancies within collaborative learning settings is to be aware of their occurrence and to interpret them validly to decide upon appropriate action. This relates to the concept of group awareness, which refers to being informed about relevant characteristics of group members or the group (Bodemer & Dehler, 2011). A group of tools designed to support this are cGATs (see Bodemer et al., 2018). These tools are designed to make learners aware of pre-selected socio-cognitive conditions. By purposefully collecting, transforming and presenting relevant data, they foster the learners' interpretation of the situation to facilitate appropriate action (Buder & Bodemer, 2008). Although the specific goals, content and processing steps used in cGATs differ greatly as well as the settings they are used in (Bodemer et al., 2018), they have specific commonalities: they process knowledge-related information on the extent or content of knowledge to facilitate the learners' identification and interpretation of relevant aspects of the collaborative situation by triggering and easing comparison processes and guiding learners towards discrepancies. By linking such discrepancies to specific content, they further activate content-related resolution processes beneficial for learning but rely on the learners' internal scripts to take it from there. Thus, they are a temporary aid guiding learners' attention and focus in an inherently complex situation. In the following, we will present three exemplary cGATs that have been used to guide collaborative learning in various settings.

Tool 1: Metacognitive and cognitive awareness for dyadic learning. This tool was developed to support co-located dyads of learners exchanging information and discussing learning material presented on a shared screen based on learning tasks designed to activate elaboration of content (see Figure 2). Such a collaborative situation provides the opportunity for learners to facilitate each other's individual learning processes by providing information and explanations to fill in gaps of knowledge or by mutually discussing diverging perspectives activating elaboration processes (e.g., de Vries et al., 2002). The educational goal in this scenario is to maximise individual knowledge gain by exploiting processes triggered by potentially beneficial knowledge distributions, i.e., discrepancies regarding extent or

content of knowledge. The cGAT provides learners with two types of knowledge-related awareness information: cognitive information on the content and metacognitive information on the extent of knowledge (see Schnaubert & Bodemer, 2019). The tool collects the learners' assumptions about the learning material by asking them to answer binary questions (content) and the learners' metacognitive evaluations of their answers by asking them to provide binary confidence ratings (extent). It then feeds the information back to the learners by presenting it during learning using spatial- (content) and colour- (extent) coding to allow for simultaneous and independent comparison processes within and between learners. The tool thus supports the informational guiding function using representational mechanisms. We also developed adapted versions to explore the effects of self-, partner-, and group-level information on individual and collaborative learning of university students (e.g., Schnaubert & Bodemer, 2016, 2019). Empirical findings clearly suggest that (a) awareness information on the extent of knowledge guides learners to focus on content for which there are knowledge discrepancies with regard to a standard (i.e., knowledge gap), (b) awareness information on the content of learners' assumptions guides learners to focus on content for which there are discrepancies between the contents of the learners' knowledge (i.e., conflict), and (c) learners integrate available types of information (content and extent). However, effects on learning outcomes were rather small and inconclusive.



Figure 2. Learning tasks with cGAT: Cropped screenshot (left) and annotated zoom-in (right)

Tool 2: Controversy awareness for collaborative knowledge construction in wikis. We used colour-coded highlights for Controversy Awareness. The aim was to make discussion threads that contain exchanges of controversial arguments more salient to users who want to learn more about a subject matter (see Figure 3). It is common for wikis that the longer an article exists the more neutral the article itself becomes. Controversies about a subject matter are often “hidden” in the article’s background. Thus, potentially relevant and interesting discussions containing meaningful controversies get easily buried in the sheer mass of discussions that are oftentimes not about the original article’s content anymore. Therefore, it can be a frustrating experience for interested users to find the discussion content they are searching for potentially limiting effects on individual learning. For this awareness tool, we built upon research on representational guidance and signalling to implement specific visual highlights as implicit guidance for wikis. We used different colours for coding controversial discussion and their resolution states with the aim of leading students towards a more focused selection of relevant content-related topics covering the subject matter under investigation. As a first step, we assessed existing wiki discussion threads and then manually coded for their content, whether it contained a controversy that was related to the article’s content and if it was led by evidence or dealing with off-topic matters (data transformation). Finally, we assigned colour-coded highlights to the discussion threads for two slightly different tool versions – one simply highlighting controversial exchanges of arguments in the thread (blue) and one further differentiating between resolved (green) and unresolved (red) controversies. We evaluated different codings of Controversy Awareness as implicit guidance in several studies in terms of effects on process variables, learning outcomes and quality of knowledge construction artefacts (Heimbuch & Bodemer, 2017, 2018). Our findings suggest that our signalling approach of highlighting content-related controversies has a direct impact on the navigation behaviour, the intensity of reading and on the likelihood of participating in a controversial topic. Furthermore, we found indirect effects on knowledge-related outcomes mediated through process variables.



Figure 3. Wiki talk page excerpt with and without controversy awareness information

Tool 3: Text mining-based knowledge awareness for knowledge exchange in school. This text mining-based tool was designed for efficiently guiding students' knowledge exchange in school settings which involve the joint solution of tasks in dyads (see Figure 4). In order to solve such tasks, the students are asked to discuss individual differences in prior knowledge or different perspectives on topics. Stimulating such exchange pursues the educational goal that learners uncover and fill their own knowledge gaps, better focus and cognitively elaborate learning content, and achieve better learning outcomes. Therefore, students need to be informed about their cognitive characteristics (Bodemer et al., 2018). Against this background, the discrepancy between one's own current knowledge extent and a higher goal state, revealed by the comparison with a more knowledgeable partner, can initiate attempts to fill one's own knowledge gaps to resolve this discrepancy. The text mining-based Grouping and Representing Tool (GRT; Erkens, Bodemer, & Hoppe, 2016) supports learners by providing them with knowledge-related information about their cognitive characteristics to support the identification of discrepancies. The tool collects learner-generated artefacts as input and transforms it by using text mining. The transformation has three main features (Erkens & Bodemer, 2019): processing the text to identify relevant concepts, pairwise determining the texts' differences for the group formation, and clustering the concepts to identify topics and determine values representing to which extent these topics are represented in each student's text. The presentation of information directly connects to these processes: The GRT lists topics offering information on what learning content to exchange knowledge about (based on the interpretation of concept clusters) and provides information about the learning partners as bar charts (how much both learning partners wrote about a topic within their texts). In a school setting, the usage of the GRT resulted in increased knowledge acquisition and knowledge convergence due to the exchange of knowledge between learning partners (Erkens et al., 2016). In a laboratory study, we found indirect effects on knowledge gains mediated through partner modelling and knowledge integration (Erkens & Bodemer, 2019). We have also found that the provision of GRT-generated information about learning content improves questioning, as it, in combination with the bars on one's own knowledge level, improves the focus on questions about topics with knowledge gaps. Thus, the GRT is suitable to guide students, resulting in better learning processes and outcomes.

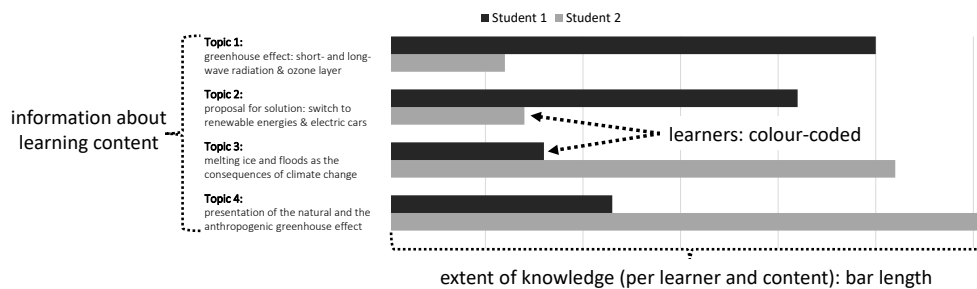


Figure 4. Information provided by the GRT (annotated)

3. Cognitive Group Awareness Tools as versatile devices

Based on theoretical assumptions and our empirical analyses, the preceding chapters illustrate the need for versatile support features of cGATs to support collaborative learning and the benefit that may be gained from making learners aware of knowledge discrepancies.

One of the major interests is the question of whether cGATs can be deployed in different settings towards a variety of goals. The examples above illustrate that cGATs are able to support learners in multiple contexts such as universities, schools, or social media across varying group sizes from dyads to large communities as has been stated previously (Bodemer et al., 2018). Although the empirical examples that we have provided look into different learning outcomes, the main educational

goal throughout is individual knowledge gain. cGATs are seldom deployed for other purposes, for example, to foster the development of higher order skills like collaboration skills. Thus, they seem to be deemed especially suited to support content-specific knowledge exchange in some form to foster individual learning. This seems logical as one function of cGATs is to cue specific information (Bodemer & Scholvien, 2014) and thus, the awareness information selected usually refers to content-specific learning material (Bodemer et al., 2018).

Another key question is whether there are common features that suggest generic functions and make cGATs versatile. Our examples show that cGATs can foster regulatory attempts based on various types of discrepancies. Depending on setting and goal, the tools target information relating to the content (tools 1 & 2) or the extent of knowledge (tools 1 & 3). However, the guiding functions these tools use all rely on processing steps that simplify complex information, either by the way they are collected (tool 1) or by the way they are transformed (tools 2 & 3). Within small groups and internal-external discrepancies, cGATs usually do not break up the relationship between the selected awareness information and the individual learner during transformation. This makes learners easily identifiable fostering partner modelling and communication efforts (e.g., Dehler, Bodemer, Buder, & Hesse, 2011; Erkens & Bodemer, 2019). Tools 1 and 3 deal with such small groups and thus abstain from aggregating data across learners during transformation, which allows learners to differentiate self- and partner-information (e.g., Schnaubert & Bodemer, 2016, 2019), both of which may affect both learners' behaviour (Lin, Tsai, Hsu, & Chang, 2019). On the contrary, to identify controversies within wikis, identifiable individuals are of less importance than the overall state of knowledge within the community. Tool 2 takes this into account when pre-interpreting discussion threads by relating individual comments to transform them into one threefold measure, simplifying this highly complex information (e.g., Heimbuch & Bodemer, 2017). Regarding data presentation, the coding mechanisms used in cGATs support various types of basic cognitive operations, especially comparison processes. For example, tool 1 uses spatial coding to allow to easily detect a discrepancy between the content of the learners' knowledge, while tool 2 uses colour-coding for highlighting controversies. Tool 3 uses bar charts to support between-learner comparisons of the extent of the learners' knowledge on a continuum and additionally colour-codes the graphs to allow to easily relate a graph to a specific learner. Tool 1 on the other hand uses colour-coding (hatching) to represent information on the extent of knowledge and uses location to make the learners easily identifiable. Thus, all use various codes to convey different information. While there seems to be a lot of adequate ways to present the data, it can be argued that colour-coding may reach its limits when comparing continuous information as small differences may become less easy to compare than when using charts. However, with binary or ternary measures, colours may provide a powerful way to draw attention by flagging or highlighting specific information as hue can be easily identified by the low-level visual system (Healey & Enns, 2012). Thus, the colour spectrum may be used to code categorical information (like in tool 3) and brightness, saturation or hatching may even code ordinal information if the number of possible values is low (like in tool 1).

Due to the number of possibilities of tool deployment, there is a need for systematic research. The empirical findings suggest that processing information to allow for easy comparisons between learners (tools 1 & 3) or even pre-interpreting discrepancies (tool 2) guides learners towards tackling those discrepancies (e.g., Erkens & Bodemer, 2019; Heimbuch & Bodemer, 2018; Schnaubert & Bodemer, 2019). How they tackle them seems to depend largely on affordances defined by the setting and it is thus not surprising, that the effects on learning gains are less straightforward and vary between tools. It becomes apparent that – while cGATs seem to be quite versatile and may guide collaborative learning efforts towards potentially beneficial social conditions like various knowledge discrepancies in various contexts – their effective employment largely depends on a number of decisions an educator has to make with regard to their specific setup. Unfortunately, no guidelines exist to support educators who want to use group awareness-support within a specific context. This is largely because research on cGATs has up until now been a rather desultory endeavour with every new study developing effective but essentially new tools with specific features and no overarching framework. Thus, the diversity of the field is its biggest assets, but also one of its biggest issues. Altogether, there seem to be a vast amount of possibilities to process data within cGATs, making them highly adaptable. However, there also seem to be boundaries for specific processing steps and thus, guidelines can be developed that support tool development and employment. To draw on this feature, we thus need systematic research looking more deeply into generic functions of cGATs to develop guidelines for tool processing. These guidelines need to consider the educational goal, but also the intended learning processes in conjunction

with the specific learning situation. This includes ideas about the discrepancies the tool is set to highlight and the processes and activities these discrepancies are meant to trigger (e.g., argumentation, knowledge exchange). We argue that the tools' processing steps (data collection, data transformation, data presentation) may be used as a framework for this kind of systematic approach as they comprise the basic decisions an educator or tool designer has to make and allow to systematically include other research areas such as computer science and research on human information processing.

References

- Bodemer, D. (2011). Tacit guidance for collaborative multimedia learning. *Computers in Human Behavior*, 27(3), 1079–1086. <https://doi.org/10.1016/j.chb.2010.05.016>
- Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, 27(3), 1043–1045. <https://doi.org/10.1016/j.chb.2010.07.014>
- Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge/Taylor & Francis.
- Bodemer, D., & Scholvien, A. (2014). Providing knowledge-related partner information in collaborative multimedia learning: Isolating the core of cognitive group awareness tools. In C.-C. Liu, H. Ogata, S. C. Kong, & A. Kashihara (Eds.), *Proceedings of the 22nd International Conference on Computers in Education ICCE 2014* (pp. 171–179). Nara, Japan: APSCE.
- Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning*, 3(2), 123–139. <https://doi.org/10.1007/s11412-008-9037-5>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA.
- de Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, 11(1), 63–103. https://doi.org/10.1207/S15327809JLS1101_3
- Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, 27(3), 1068–1078. <https://doi.org/10.1016/j.chb.2010.05.018>
- Erkens, M., & Bodemer, D. (2019). Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents. *Computers & Education*, 128, 452–472. <https://doi.org/10.1016/j.compedu.2018.10.009>
- Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning*, 11(4), 387–415. <https://doi.org/10.1007/s11412-016-9243-5>
- Healey, C., & Enns, J. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7), 1170–1188. <https://doi.org/10.1109/TVCG.2011.127>
- Heimbuch, S., & Bodemer, D. (2017). Controversy awareness on evidence-led discussions as guidance for students in wiki-based learning. *The Internet and Higher Education*, 33, 1–14. <https://doi.org/10.1016/j.iheduc.2016.12.001>
- Heimbuch, S., & Bodemer, D. (2018). Interaction of guidance types and the Need for Cognitive Closure in wiki-based learning. *PeerJ*, 6:e5541. <https://doi.org/10.7717/peerj.5541>
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). Constructive controversy: The educative power of intellectual conflict. *Change: The Magazine of Higher Learning*, 32(1), 28–37. <https://doi.org/10.1080/00091380009602706>
- Lee, G., & Kwon, J. (2001). What do we know about students' cognitive conflict in science classroom: A theoretical model of cognitive conflict process. In P. A. Rubba, J. A. Rye, W. J. Di Biase, & B. A. Crawford (Eds.), *Proceedings of the Annual Meeting of the Association for the Education of Teachers in Science* (pp. 309–325). Costa Mesa, CA: Association for the Education of Teachers in Science.
- Lin, J.-W., Tsai, C.-W., Hsu, C.-C., & Chang, L.-C. (2019). Peer assessment with group awareness tools and effects on project-based learning. *Interactive Learning Environments*, 0(0), 1–17. <https://doi.org/10/gf2qm6>
- Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Oxford, United Kingdom: Viking Press.
- Schnaubert, L., & Bodemer, D. (2016). How socio-cognitive information affects individual study decisions. *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, 1*, 274–281. International Society of the Learning Sciences.
- Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 14(1), 7–51. <https://doi.org/10.1007/s11412-018-9293-y>

- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum.
- Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York, NY: Routledge.

Scaffolding computer-supported collaborative lesson design: A spiral model

Wenli CHEN^a, Jesmine S H TAN^{b*}

^{ab} *National Institute of Education, Nanyang Technological University Singapore*

*jesmine.tan@nie.edu.sg

Abstract: This paper presents a computer-supported collaborative lesson design approach supported by a spiral model of collaborative knowledge improvement (SMCKI) to develop pre-service teachers' competency in designing ICT-integrated lesson for the teaching and learning of Chinese language. The computer-supported collaborative lesson design approach served to deepen pre-service teachers' understanding of lesson design using the technology, pedagogy and content knowledge (TPACK) framework. Results showed that the SMCKI is effective in supporting pre-service teachers' collaborative lesson design. Commencing with an individual brainstorming phase, the SMCKI pedagogical model scaffolds the pre-service teachers through an intra-group synergizing phase, an inter-group critique phase and an intra-group refinement phase, to support the advancement of their collective and individual knowledge in the collaborative lesson design. The findings show that the employment of SMCKI on computer-supported collaborative lesson design was able to bring about a collaborative knowledge improvement to designing ICT-integrated lesson in the teacher preparatory course.

Keywords: computer-supported collaborative lesson design, collaborative lesson design, computer-supported collaborative learning

1. Introduction

The integration of Information, Communication and Technology (ICT) to facilitate 21st century Teaching and Learning (T&L) had become pervasive in educational institutes globally (Ghavifekr & Rosdy, 2015; Goh, Lee & Taylor, 2016; Valtonen, Sointu, Mäkitalo-Siegl & Kukkonen, 2015; Wilson, Tete-Mensah & Boateng, 2014; Wu, Yen-Chun, Chia-I & Chih-Hung, 2017). Therefrom, creating a technology-infused learning environment is no longer an alien affair. However, past research findings had shown that educators perceived ICT-integrated lesson design significantly difficult, challenging and demanding, especially when aligning among content, pedagogy and technology knowledge (TPACK) (Chai, Koh, Lim & Tsai, 2014; Koehler, Mishra, Bouck, DeSchryver, Kereluik, Shin & Wolf, 2011; Looi, So, Toh & Chen, 2011; Wong, Chai, Zhang & King, 2015). Hence, in order to prepare these pre-service teachers with the skillset to engage learners with meaningful use of technology, developing them with the relevant competency to design such lessons is exigent (Finger, Romeo, Lloyd, Heck, Sweeney, Albion & Jamieson-Proctor, 2015; Valtonen, Sointu, Mäkitalo-Siegl & Kukkonen, 2015).

In fact, past studies showed that collaboratively design technology-enhanced curriculum materials do support teachers in becoming TPACK competent (as cited in Agyei & Voogt, 2012). Effectively, research had also advocated that the engagement in design teams enabled productive sharing of knowledge, skills, experience and challenges related to technology-enhanced T&L (Kafyulilo, Fisser & Voogt, 2016; Voogt, Laferriere, Breuleux, Itow, Hickey & McKenney, 2015). Even though collaboratively designing a curriculum can be administered with different approaches, the use of technology to support collaborative learning and idea improvement is recommended (Chen, Looi, Wen, 2012; Jeong & Hmelo-Silver, 2016; Resta & Laferrière, 2007; Stahl, Koschmann & Suthers, 2006). Furthermore, not only would computer-support impacts the collaborative lesson design, but it could also improve learning in teacher education. Thus, we perceived that with Computer-supported collaborative lesson design (CSCLD), the learning potential of collaboration in higher education will be enhanced. In view of this, this study aimed to examine the process and outcome of scaffolding CSCLD via the pedagogical support – the Spiral Model of Collaborative Knowledge Improvement (SMCKI) (Chen, Zhang, Wen & Looi, 2019). The research questions (RQ) of the study are as follows:

- 1) How do the pre-service teachers collaboratively design the lesson ideas throughout the different phases of SMCKI?
- 2) How do the inter-group critique help with the collaborative lesson idea design?

2. Literature Review

2.1 Design for technology-enhanced learning (TEL)

Using TPACK as a framework for designing TEL had been long acknowledged as a useful approach to help build teachers' competency in integrating ICT in lessons design and implementation (Chai, Koh, Tsai & Tan, 2011; Koh, 2013; Koh & Chai, 2016). TPACK was defined as the dynamic, transactional relationship between the content, pedagogy and technology, and the combination of these 3 elements was aimed at developing appropriate, context-specific strategies and representations for ICT-infused lessons (Koehler et al., 2007). Offered to teachers as a framework for effective technology integration in T&L (Koehler, Mishra & Yahya, 2007; Mishra & Koehler, 2007). TPACK emphasizes the integrated use of technology, pedagogy and content knowledge for effective technology integrated lesson design (Thompson & Mishra 2007). However, past research studies that employed TPACK as a lesson design knowledge in pre-service education did not seem to yield satisfactory results, although perception and knowledge acquisition was relatively well established (Knoef & Lazonder, 2019).

So and Kim (2009) indicated that since content (CK), pedagogical (PK), and technological (TK) knowledge are all inter-related, teacher education programs should be structured in a holistic manner to allow pre-service teachers to see their connections as studies had found them less able to consider linkages between content and pedagogy when envisioning their lesson agendas as compared with expert teachers (Copeland et al., 1994; Leinhardt, 1989; Sabers et al., 1991). On this ground, Koh (2018) discovered that design scaffolds had positive effects on teachers' TPACK confidence and were useful to help them articulate pedagogical change in their lesson designs. In addition, Tanak (2018) also recommended including authentic experiences to create a practical experience in learning in a TEL environment. For this reason, the learning-technology-by-design approach was adopted in this study.

Learning by Design is an approach whereby an environment is created for teachers to naturally confront them through participating in the TPACK lesson design (Koehler & Mishra, 2005). This differs from a traditional technology workshop or class where teachers were trained to be consumer of tools, with the hope that they can apply them to their practice. By employing Learning by Design approach in this study, not only can pre-service teachers collaboratively worked on designing a TEL environment, they can experience such a learning environment as the capacity of a student as well.

2.2 Collaborative lesson design

Collaborative lesson design approach was deemed as a promising strategy for developing teachers' learning and supporting them in becoming TPCK competent (as cited in Agyei & Voogt, 2012; Voogt, Fisser, Roblin, Tondeur & van Braak, 2013). The benefits of the collaborative designing include: shared cognition, increases the effectiveness of creative and innovation processes, critical understanding of a product through reciprocal reasoning (Kangas, Seitamaa-Hakkarainen & Hakkarainen, 2013; Tschimmel, 2012). Specifically, this approach promotes trying out new tools, expertise sharing, reflection on pedagogy and collaborative work to improve teaching skills and the academic performance of students (Kafyulilo et al., 2015; Lee, Lee & Kuptasthien, 2018) to bring about divergent thinking among the teacher designers (Brown & Wyatt, 2010). In addition, findings had shown that design teams offered collaborative learning opportunities, at the same time, enhance the development of TPCK (Kafyulilo, Fisser & Voogt, 2015). Furthermore, collaborative and reflective practices were reported to increase teachers' understanding and thinking of their professional experiences in various situations, emphasizing that prejudices can be overcome through interactions and collaboration with peers (Aşık, Eroğlu İnce & Şarlanoğlu Vural, 2018).

2.3 Pedagogical model for collaborative lesson design

Although there are several models and approaches to collaborative lesson design, this study adopted the learning technology by design approach. This approach could provide theoretical grounding of including technologies in the pedagogical decisions as well as appreciating the constraints and affordances of using technologies in the working environment (Benning, Linsell & Ingram, 2018). Indeed, this use of technology to support collaborative lesson design is also a form of Computer-Supported Collaborative Learning (CSCL) (Stahl et al., 2006) as its application did “provide communication, coordination and collaboration tools that ease group dynamic regardless of space-time location of group members” (Costaguta, Santana-Mansilla, Lescano & Missio, 2019, p. 159). While CSCL provided the means for CSCLD, SMCKI (Chen et al., 2019) was employed as the pedagogical model in this study. Elaboration on the 5-phase SMCKI will be delineated at the methodology section.

3. Methodology

The participants in this study were 23 pre-service Chinese language teachers who were in their 3rd year of the Bachelor of Arts programme with the National Institute of Education, Nanyang Technological University Singapore. They were enrolled in a course on designing effective ICT-integrated learning environments. Demographics of participants comprises: 78% between 20 – 25 years old and the remaining between 26 – 35 years old. The lecturer is the 2nd author of the paper, who had 2½ years of experiences in educating pre-service teachers on the effective use of technology for T&L of the Chinese language. Given the practical constraints, only a small sample size was recruited.

Among the 23 students, a total of 6 groups were formed by students themselves, with 5 groups of 4 members each and 1 group with 3 members. According to Chapman, Meuter, Toy and Wright (2006), self-selected groups may simulate “real-world” workgroups more closely than randomly assigned groups and evidence had suggested that self-selected groups led to better group dynamics and yield better group collaborative work results” (as cited in Chapman et al., 2006, p. 560).

The CSCLD was carried out in January 2019 with a 1-hour duration. During the face-to-face (F2F) lesson, participants were tasked to collaboratively design an ICT-integrated lesson unit for Chinese language T&L targeting at primary schools students in Singapore. The task requirements comprised: 1) content alignment with the Singapore Chinese language curriculum; 2) at least 2 language skills (listening, speaking, reading or writing); 3) CSCL with appropriate PK and TK. With CK as participants’ prior knowledge, the lecturer taught them the necessary PK and TK. The PK for this task consists of 3 lesson design principles: 1) seven affordances of CSCL (Jeong & Hmelo-Silver, 2016), 2) the seven pedagogical principles for collaborative virtual environment (Rubens, Emans, Leinonen, Skarmeta & Simons, 2005) and 3) the seven principles of designing a student-centered learning environment (Aw, 2018). Among the various TKs taught the specific online collaborative platform Padlet was used as the media for the CSCLD process as well. The selection of Padlet was premised on an affordance analysis based on 3 online collaboration tools (Google doc, Linoit and Padlet) (Mallon & Bernsten, 2015). Padlet ranked highest as it possesses the feature of linking comments with the generated post (idea), which is important for ease of peer-comments visualization in this study.

Prior to the CSCLD, each group leader had to create their group Padlet board and share it with their group members. The hyperlink of the group’s Padlet board was subsequently posted to the virtual class wall, Edmodo. Every participant within the class had an account in Edmodo. Edmodo class page served as a base camp for communication and dissemination of information. Through the hyperlinks, the lecturer could also monitor the online posts for each group at each phase. The following delineates the CSCLD procedure via SMCKI:

Phase 1 Individual brainstorming (10 minutes): Individual lesson ideas posting based on learners’ needs on the Padlet group board.

Phase 2 Intra-group Synergizing (20 minutes): At the individual group level, members view the lesson idea of each other, discuss, consolidate and synthesize a highest quality group lesson idea.

Phase 3 Inter-group critique (15 minutes): Based on round-robin schedule, participants view other group’s lesson idea and provide constructive comments and suggestions for idea improvement.

Phase 4 Intra-group refinement (15 minutes): Participants returned to their group board to read the comments/suggestions given by other groups. Intra-group discussion took place and to refine their lesson idea.

Phase 5 Individual perfection (after class): Individual participant reflects his/her take-away from the collaborative lesson design activity and writes a reflection report.

4. Data collection, Analysis and Results

Lesson ideas generated at phase 1 (P1), phase 2 (P2) and phase (P4), including peer comments at phase 3 (P3) from Padlet were collected to examine the learning process and outcome of the CSCLD activity.

To answer the RQ1 on whether SMCKI is effective in proving students' ideas through the CSCLD, both the quantity and quality of lesson ideas generated were examined via content analysis. The unit of analysis was each lesson idea posted at P1, P2 and P4 in the Padlet group boards. Given the context of the lesson design for ICT-integrated lesson, the coding scheme for evaluating lesson idea on using TPACK for meaningful learning with ICT (Koh, 2013) was adopted in this study. Koh's coding scheme was based on Jonassen, Howland, Moore, and Marra's (2002) and Howland et al.'s (2012) framework of meaningful learning with ICT. The 5 dimensions that characterized how ICT could support "Meaningful Learning" are:

1) active – where students were not passive listeners but actively manipulating objects and information, and observing results; 2) constructive – where students constructed knowledge, reflected, and articulated their personal understandings of the phenomenon; 3) authentic – where students engaged in the solving of real-world problems; 4) intentional – where students set their learning goals and planned their learning pathways; and 5) cooperative – where students worked with peers to learn.

Each post was coded from the 5 dimensions with a scale of 1 to 5 based on the quality of the content. Lesson ideas at P1 were individually scored and the mean score computed to compare against the group synergized idea at P2 and P4. Out of the 6 groups, 5 groups showed collaborative knowledge improvement from P1 to P2. The idea improvement at P4 was evident after P3 as well. Results showed that 5 out of 6 groups had productive intra-group discussions which led to improved lesson idea at P4. Therefore the RQ 1 is answered. Guided by SMCKI, the CSCLD, to a large extent, had supported pre-service teachers' collaborative knowledge improvement in the lesson design process.

To answer the RQ2 on how the inter-group critique helped with the collaborative lesson design, further content analysis was conducted on the inter-group peer comments data at P3 to examine if and how the inter-group peer critique help improve the quality of the lesson ideas. Based on the coded data, there were a total of 56 comments (M=2.4) made during the mere 15-minutes at P3. Of all the comments analyzed, 43% reflected relevant supporting evidence and reasoning using the lesson design principles and task requirement. The coding scheme from Clark and Sampson (2007) and Chen et al. (2019) was adapted to analyze the peer comments. The 4 dimensions in the coding scheme are: 1) Support (with relevant evidence or reasoning); 2) Rebuttal (attack an explanation with relevant shreds of evidence or reasoning); 3) Query (seek clarification with relevant evidence or reasoning), and 4) Emotive appeal (the comment is emotional in nature without relevant evidence or reasoning). Each comment was coded with a scale of 1-5 where 5 refers to perfect and 1 refers to absence or irrelevant for all the 4 dimensions above. Of the 4 dimensions, the rebuttal and query played a pivotal role both as a reminder and a trigger to induce further considerations to the lesson design objectives and procedures during P4. The improvements made to the lesson ideas at P4 can be tracked with the comments they received from other groups in P3. Out of the 6 groups, 4 groups demonstrated improvement to their lesson design at P4. Two cases were presented below to illustrate how each group's lesson ideas were improved by addressing comments from other groups.

Table 1

Translated critiques and lesson idea improvement for Case 1 (G5)

Phase	Lesson idea / Comments (Chinese texts translated to English)
P2	“Lesson objective: Students can list the ways and examples of a healthy lifestyle. Lesson Procedure: Students collaboratively complete a poster using multimedia. 1. Students type or use pictures to populate their idea on the WHITEBOARD on the ways to take good care of our body. 2. Group discussion on individual ideas. 3. Students to complete a group poster after discussion.”
P3	Comments by G4: “What is the relationship between a poster and oral ? ” “Are students able to learn commonly-used oral skills ? ”
P4	“Lesson objective: Students can list the ways and examples of a healthy lifestyle. Students

collaboratively complete a poster using multimedia with **recorded speech**. 1. Students type or use pictures to populate their idea on the WHITEBOARD on the ways to take good care of our body. 2. Group discussion take place referencing individual ideas. 3. Each group to complete a poster after discussion. **4. Inter-group assessment on the poster on the suitability of the content and make corrections where necessary. 5. Group refinement on poster based on comments. 6. Homework: students to record their speech using the group poster.**”

The above Table 1 showed Group 5 (G5)’s synergized lesson ideas at P2, inter-group comments received from Group 4 (G4) at P3, and G5 improved ideas at P4 (bolded words). G5’s original synergized lesson idea (P2) was to engage students with oral ideas collected through the collaborative construction of a poster surrounding the topic. This lesson idea was commented with queries from G4 members (see Table 1). Addressing this at P4, an added lesson objective was included to align the purpose of cultivating oral skill using the poster. Additional activities were also included to help students practice their oral presentation skill. The improvement from P2 to P4 was characterized by the use of ICT tools to support students in investigating real-world phenomena (P4) rather than using ICT tools for mere presentation (P2). According to Koh (2013), “the more the activity facilitated students to make connections between their own experiences and the real-world phenomenon associated with the subject matter, the more it was considered as authentic” (p. 891). In this case, the improved idea served both the practical needs of school-based oral assessment and also as a life skill with poster presentation.

Table 2

Translated critiques and lesson idea improvement for Case 2 (Group 2)

Phase	Lesson idea / Comments (Chinese texts translated to English)
P2	Students brainstorm the letter-writing content at Mindmeister/ Popplet. This is to construct a group planning mind map.
P3	Comments from G1: “I felt that student would probably create different types of mindmap. To avoid that, the teacher could provide a standard template to help students with brainstorming.” “Multi-sensory principle: when working on the mindmap, can let students draw or write in text.” Comment from G4: “Lack of scaffold in the lesson design. The teacher can provide a template, let the students use the template to sprout their ideas and refine their own product using peer comments.”
P4	“At the individual brainstorming, allow each student to search for the information related to the country the group is assigned to. Thereafter, they can post their own ideas to the mindmap. We will provide a letter-writing mindmap template for the students to fill in (eg greeting phrase, what did we do during the last meeting, where do you intend to go during the holidays, what about the season and weather, conclusion./Weather in country N, places of interest ETC)”

Table 2 demonstrated how Group 2 (G2) improved the quality of “constructive” dimension of the idea of the lesson from P2 (quality score = 3) to phase 4 (quality score = 4) by addressing comments from other groups. This improvement was characterized through the use of ICT tools to synthesize information in order to construct verbal, written, visual, conceptual or product-oriented expressions of the subject matter. The inclusion of the “template” with “scaffold” at P4 provided a concrete base where students can brainstorm and construct the letter-writing content more systematically. Although there was no change to the individual brainstorming procedure, including a “mindmap template for the students to fill in (see Table 2) provided the means for information synthesis. This “template” which served as a “scaffold” was suggested by both G1 and G4 at P3. According to Lim and Tay (2003), higher-order thinking would only be possible with appropriate support structures and informative tools. This suggested that the improvement at P4 provided higher levels of the constructive dimension, which was demonstrated by knowledge expressions through the scaffold (Koh, 2013).

From the 2 cases presented, we perceived that the inter-group critique (SMCKI-P3) did contribute to the collaborative lesson design. Not only did the online critiques served as explicit means for the participants to engage in a self-diagnosis and remediation of the learning gaps (Koh, 2013), the same critique was also used as an agent for improving their individual group lesson design at P4.

5. Discussion & Conclusion

Findings from this study suggested that the 5-phase SMCKI model supported pre-service teachers with fruitful CSCLD. While collaboration among designers is often seen as challenging (Leinonen & Durall-Gazulla, 2014), findings from this study revealed that through collaboration in design teams guided by SMCKI, participants were able to learn from each other and from their practices as they designed the ICT-integrated lessons (Kafyulilo et al., 2016). The application of this approach suggested that teacher learning can be effective if it is situated in a meaningful context, with active engagement of their own learning process and collaborating with their peers (Voogt et al., 2011). Critical thinking was promoted throughout the SMCKI process as participants had to continually engage themselves with the design principles during the intra-group discussion and inter-group critique (Kafyulilo et al., 2016). Through the active participation of the activities scaffolded by SMCKI, a knowledge creation culture was well initiated (Scardamalia & Bereiter, 2006). Pre-service teachers could see themselves and their work as “part of the civilization-wide effort to advance knowledge frontiers” (Scardamalia & Bereiter, 2006, p. 98). The culture of knowledge improvement was hence fulfilled in this activity.

Being a small scale case study, there are several limitations. Apart from the small sample size, the current data analysis only focused on the artefacts generated on the online platform. Future study can be conducted to examine both students’ online communication and F2F discussion during the various phases of SMCKI. In addition, individual reflections could also be analyzed to understand the perceptions of collaborative learning in higher-education setting. Finally, considerations can be made to explore the employment of SMCKI as a pedagogical model for other disciplines or tasks in CSCLD.

References

- Aşık, A., Eroğlu İnce, B. H., & Şarlıanoğlu Vural, A. (2018). Investigating learning technology by design approach in pre-service language teacher education: Collaborative and reflective experiences.
- Benning, I., Linsell, C., & Ingram, N. (2018). Using Technology in Mathematics: Professional Development for Teachers. *Mathematics Education Research Group of Australasia*.
- Chapman, K. J., Meuter, M., Toy, D., & Wright, L. (2006). Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. *Journal of Management Education*, 30(4), 557-569.
- Chen, W., Wen, Y., Looi, C.K. (2012). Technology Enhanced Pedagogical Innovation in Second Language Learning. *Global Chinese Journal for Computers in Education*, 8(1/2), 122-133.
- Chen, W., Zhang, S., Wen, Y., Looi, C.K. & Yeo, J. (2019). A Spiral Model of Collaborative Knowledge Improvement to Support Collaborative Argumentation for Science Learning: Technological and Pedagogical Design. In *13th Int'l Conference on CSCL* (pp. 240-247). Lyon, Singapore: École Normal Supérieure de Lyon.
- Costaguta, R., Santana-Mansilla, P., Lescano, G., & Missio, D. (2019). Mining Associations Between Collaborative Skills and Group Roles in Collaborative E-Learning Environments. *Journal of Information Technology Research (JITR)*, 12(2), 159-174.
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?. *Educational Psychologist*, 51(2), 247-265.
- Jonassen, D. H., Howland, J., Moore, J., & Marra, R. M. (2002). Learning to solve problems with technology: A constructivist perspective.
- Kafyulilo, A. C., Fisser, P., & Voogt, J. (2015). Supporting teachers learning through the collaborative design of technology-enhanced science lessons. *Journal of Science Teacher Education*, 26(8), 673-694.
- Kafyulilo, A., Fisser, P., & Voogt, J. (2016). Teacher design in teams as a professional development arrangement for developing technology integration knowledge and skills of science teachers in Tanzania. *Education and Information Technologies*, 21(2), 301-318.
- Koehler, M. J., & Mishra, P. (2005). Teachers learning technology by design. *Journal of computing in teacher education*, 21(3), 94-102.
- Koh, J. H. L., Chai, C. S., Benjamin, W., & Hong, H. Y. (2015). Technological Pedagogical Content Knowledge (TPACK) and design thinking: A framework to support ICT lesson design for 21st century learning. *The Asia-Pacific Education Researcher*, 24(3), 535-543.
- Lim, C. P., & Tay, L. Y. (2003). ICT in an elementary school: Students’ engagement in higher order thinking. *Journal of Educational Multimedia and Hypermedia*, 12(4), 425-451.
- Mallon, M., & Bernsten, S. (2015). Collaborative learning technologies. *USA: Association of College and Research Libraries and American Library Association*.

- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, 108(6), 1017-1054.
- So, H. J., & Kim, B. (2009). Learning about problem based learning: Student teachers integrating technology, pedagogy and content knowledge. *Australasian Journal of educational technology*, 25(1), 101-116.
- Stahl, G., Cress, U., Ludvigsen, S., & Law, N. (2014). Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning*, 9(2), 117-125.
- Thompson, L., & Ku, H. Y. (2010). Degree of online collaboration and team performance: A case study. *Quarterly Review of Distance Education*, 11(2), 127.
- Voogt, J., Laferriere, T., Breuleux, A., Itow, R. C., Hickey, D. T., & McKenney, S. (2015). Collaborative design as a form of professional development. *Instructional science*, 43(2), 259-282.

How shared concept mapping facilitates explanation activities in collaborative learning: An experimental investigation into learning performance in the context of different perspectives

Shigen SHIMOJO* & Yugo HAYASHI

College of Comprehensive Psychology, Ritsumeikan University, Japan

*cp0013kr@ed.ritsumei.ac.jp

Abstract: Studies in collaborative learning have shown that explanation activities drawing on diverse perspectives facilitate deeper understanding and metacognition. However, it is difficult to develop an explicit understanding of others' perspectives and knowledge through communication in a computer-mediated environment. The present study investigated the use of a visually shared concept map interface, expected to facilitate dyadic awareness of different perspectives and thus improve learning performance during explanation activities. In this study, each dyad built a concept map about a key technical term in psychology, and generated explanations of the term and generated explanations of the term on mutually accessible concept maps. We predicted that learners would be able to (1) gain deeper knowledge through the shared explanations; and (2) explain the key term from different perspectives or knowledge sets. Twenty-six university students participated in this experiment, and we assessed their performance through free recall tests before and after they used the concept mapping tool. Our findings showed that learners were able to (1) gain learning performance and (2) explain a concept based on different perspectives. We discuss the implications of our findings and suggest directions for further research on the development of learning support systems.

Keywords: Explanation Activities, Computer-Supported Collaborative Learning (CSCL), Perspective Taking, Concept Map

1. Introduction

Explanations generated and shared among learners in collaborative learning are important strategies to facilitate metacognition and sophisticated interactions (Chi, Leeuw, Chiu, & Lavancher, 1994; Miyake, 1986; Shirouzu, Miyake & Masukawa, 2002; Hayashi, 2019). When developing efficient computer-supported collaborative learning (CSCL) environments employing explanation activities, it is important to understand the type of interactions useful in the facilitation of such activities.

Drawing on Vygotsky's sociocultural learning theories (Vygotsky, 1980; Lave & Wenger, 1991), numerous studies in the cognitive sciences have focused on learner-to-learner collaborative learning (Roschelle, 1992; Schwartz, 1995; Okada & Simon, 1997; Hayashi, 2018a). Throughout these studies, the importance of posing questions and developing explanations to share with peers in the facilitation of metacognition and deeper understanding has been emphasized (Miyake, 1986; Shirouzu et al., 2002). Past research in cognitive psychology has shown that self-explanation facilitates metacognitive processing, further supporting the implication that peer-to-peer explanation benefits the learning process.

What are the mechanisms that facilitate effective peer-to-peer explanation activities? Shirouzu, et al. (2002), through their investigation of a simple origami task, point out that the generation of peer-to-peer explanations provides individuals with an opportunity to externalize their particular perspectives and sharing it with their collaborators. The listener will interpret this externalized thought

based on their own perspective, which could lead to a request for further explanation of the first individual's perspective. In this iterative process, reflective feedback between peers provides an opportunity to reconsider the context that was initially posed to the collaborator. Furthermore, the listener may interpret the content in a different way, providing the speaker with an opportunity to re-consider the content from a different perspective. This indicates that interactions based on different perspectives facilitate metacognition and improved content learning through the exchange of different perspectives.

Studies in cognitive science have explored the process of successful collaboration between participants holding different perspectives (Hayashi, Miwa & Morita, 2007; Hayashi, 2008b) and found that collaborative problem-solving based on different perspectives in a setting with constrained communication channels, including computer-mediated interactions, are more likely to fail (Hayashi & Miwa, 2011). On the other hand, richer environments where learners have access to multiple communication channels, are more likely to succeed in such type of activities. This leads us to consider the types of technology that can be used in CSCL environments to facilitate a successful exchange of knowledge and perspectives among peers.

Previous CSCL studies investigating the use of social awareness tools have explored the ways in which the use of such technology foster learning performance and support learner interaction (Dillenbourg & Fischer, 2007; Bodemer, 2011). For example, studies using shared concept-mapping that allows participants to observe different knowledge sets have demonstrated participants' ability to collaboratively obtain new knowledge (Engelmann & Hesse, 2010; Sangin, Molinari, Nüssli & Dillenbourg, 2011). In the study conducted by Molinari, Sangin, Dillenbourg and Nüssli (2009), dyads were able to access visualizations (individual concept maps) of their own and their partners' prior knowledge through both own and peer-generated maps. They found that the degree to which participants co-manipulate the same objects in the collaborative map was higher when they discussed identical information. This indicates that information provided by their partners helped learners gain deeper levels of understanding. However, the limitations of this study possess two problems. Firstly, the researchers did not analyze learning performance based on a pre-post design, which limits inferences about learners' process of knowledge acquisition. Secondly, the study did not focus on the mechanism through which different perspectives and knowledge were acquired in peer-to-peer interactions. It is therefore necessary to investigate the types of knowledge and perspectives that emerge through the whole process.

Against this backdrop, we conducted an experiment where individuals paired into collaborative learning dyads each built a concept map about a key technical term in psychology and generated explanations to each other while being able to see each other's concept maps. It was predicted that learners would be able to (1) gain deeper knowledge through the explanations (H1), and (2) explain the term from different perspectives or using different knowledge sets (H2).

2. Method

2.1 Research design and participants

For this experiment, 26 university students (male = 11, female = 15) were recruited and separated into 13 pairs. The average age was 20.7 (SD = 1.37) and the experiment utilized a factorial within-subjects design.

2.2 Experimental set-up

In the experiment, two PCs were prepared for use by participants. Two monitors were connected to the PCs; two video-recording devices (Sony, HDR-CX680) were set up; and Cmap software (<https://cmap.ihmc.us/>) for the development and synchronization of concept maps was installed on the PCs. This set-up allowed for the simultaneous production and sharing of concept maps, thereby enabling efficient sharing of knowledge and externalized experiences among participants.

In the experimental task, participants were not able to see each other, although they were able to do so in the face-to-face setting.

2.3 Procedure

The experiment consisted of two phases. Participants first read the material provided on the key psychological term (attribution theory) individually, and then developed their explanations of the term through concept-mapping, using the Cmap software. Before the collaboration, they were instructed to make the concept map about it explaining to each other by themselves. Prior to the explanation activity, participants developed concept maps individually to familiarize themselves with the system, after which they received instruction on concept maps (e.g. the use of concept maps in demonstrating relationships between concepts) and the use of Cmap software.

Next, participants had 15 minutes to study the learning material on attribution theory, followed by the middle-test to evaluate the extent of learners' gained knowledge on attribution theory. It was expected that learners would acquire more information following the middle-test, thus increasing the amount of generated content.

Participants were instructed to "freely write about attribution theory" as a test to assess participants' knowledge. As will be explained in the next section, this was conducted before they read the material(pre-test), after they finished reading(middle-test) and after the task(post-test). It was expected that effective interaction between learners during this task would enhance their post-test performance, compared to the pre- and middle-test conditions.

2.4 Dependent variables

Learning performance and lexical network analysis were set as dependent variables and expected to change between tests. As explained previously, the dependent variables were collected through three phases. The pre-test was conducted prior to starting the task and middle-test after they read the description to determine the change in the first variable. The post-test after collaborative learning using concept maps was conducted to determine changes in the experimental task.

Learning performance was assessed through coding, conducted by two coders. The first coder conducted all coding and the second conducted coding on 20% of the data, selected at random. Krippendorff's alpha coefficient was 0.96 and coding conducted by the first coder was used for analysis (Schneider & Pea, 2014). Coding was standardized as follows:

- naïve correct or not based on demonstrated knowledge: 1 point;
- participants provided an abstract answer based on demonstrated knowledge, without explanation: 2 points;
- participants' answers reflected demonstrated knowledge, but the answer or knowledge were incorrect: 3 points; and
- participants answered correctly based on demonstrated knowledge: 4 points.

In addition to these, one point was added for each of the following: using a concrete example; providing a unique explanation; demonstrating a new discovery through the explanation; and using own words instead of verbatim replication of the text.

The final dependent variable, lexical network analysis, focused on the lexical networks extracted from the responses to the middle- and post-tests. All the textual data from the middle- and post-tests were analyzed using morphological analysis, following the method proposed by Hayashi & Inoue (2015). We developed a lexical network for each phase and compared the results.

3. Result

3.1 First hypothesis: Analysis of the test evaluation scores

To test H1, we analyzed the changes in learners' scores on tests evaluating learning performance.

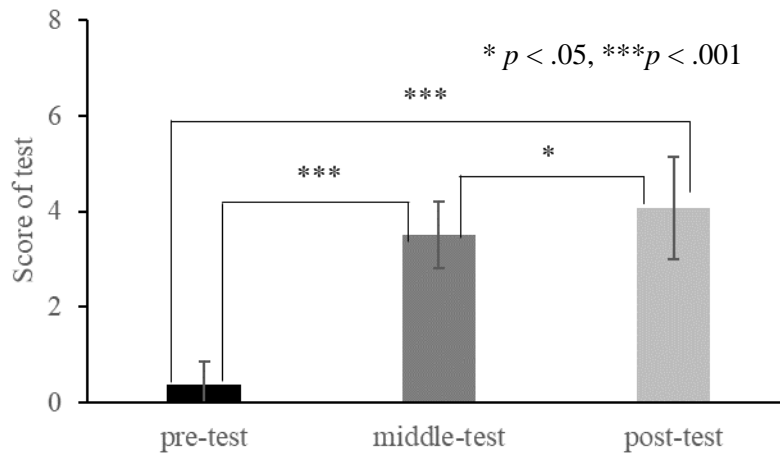


Figure 1. Comparison between mean test scores. The error bar indicates SD and asterisks indicate statistical significance.

Figure 1 depicts the mean test scores derived from coding, showing significant differences among the three test phases ($F(2, 25) = 173.78, p < .001, \text{partial } \eta^2 = 0.87$). Multiple comparisons using the Ryan's method revealed that pre-test scores were significantly lower than other tests ($p = .00, p = .00$), consistent with results obtained through comparison of the average numbers. Furthermore, the average post-test score was higher than that of the middle test ($p = .02$). This result supports our hypothesis, indicating that learners gain better understanding through explanation activities using the concept map.

3.2 Second hypothesis: Analysis of lexical networks of the test descriptions

Learners' performance on the middle and post-test was analyzed to explore the impact of different types of knowledge and perspectives on their performance, in consideration of H2. Pearson's correlation analysis was conducted to compare noun production between middle- and post-test individually, with the average established at -0.39 ($SD = 0.22$). All participants had a negative correlation between the two phases of the lexical network analysis, indicating that learners utilized different types of words between the middle- and post-tests. Figure 2 shows an example of one participant's lexical network, indicating its negative correlations.

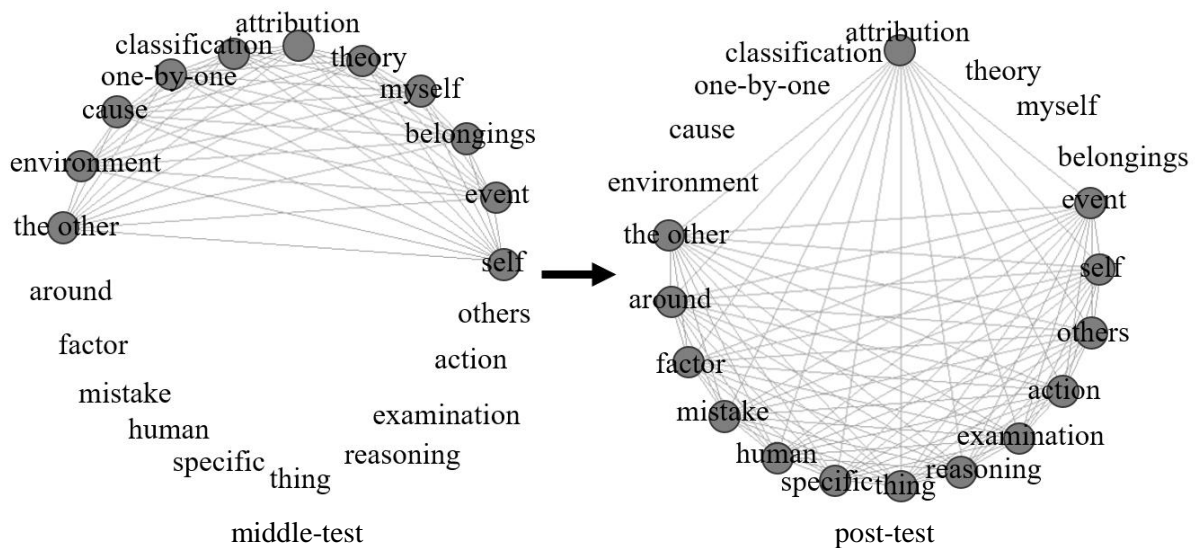


Figure 2. An example of one learner's lexical network. The correlation between lexical networks for this participant was -0.67 .

Taking into consideration our finding that learners used different types of perspective and knowledge sets at post-test, this finding supports our second hypothesis.

4. Discussion and conclusions

The goal of this study was to conduct an experiment on collaborative learning dyads, where each dyad built a concept map on a key technical term in psychology, and generated explanations of the term to each other where they were able to see each other's concept maps. It was predicted that learners would be able to (1) gain deeper knowledge through the explanations (H1), and (2) explain the key term from different perspectives or knowledge sets (H2). Results of the comparisons between the pre-, middle- and post-tests in terms of the amount of descriptions generated and the evaluation scores, suggest that the learners drew on more sophisticated learning strategies to gain knowledge. Moreover, results obtained when we tested H1 suggest that learning performance increased between the middle- and post-test phases, supported our first hypothesis. This indicated that learners actually gained more knowledge through interaction using the concept map. It is important to note that previous studies (Engelmann & Hesse, 2010; Sangin et al., 2011) did not conduct similar comparisons between test phases. Our study therefore expands the empirical evidence base through more sophisticated data collection.

The results obtained in testing H2 demonstrated the variety of perspectives and knowledge sets utilized between the middle- and post-test phases, supporting our second hypothesis. Through this analysis we were able to capture the degree to which learners used different types of knowledge during their explanation activities, using the concept map. Considering the analysis conducted in a previous study (Hayashi & Inoue, 2015), there was a drastic change in the use of different types of perspective and knowledge sets between the two phases. This could be attributed to the use of the concept map. However, as the goal of this study was not the clarification of this phenomenon, further investigation is needed to assess the effect of the concept map in this regard.

One future study will investigate the development of collaborative learning systems that enable the facilitation of increased awareness of other learners' knowledge, and we are considering the use of conversational agent to monitor learners' activities while also providing direct suggestions on the use of different perspectives and knowledge sets during their interactions. In our laboratory, we have conducted a number of studies on the use of these conversational agents (Hayashi, 2019; Hayashi, 2018a; Hayashi, 2018c; Hayashi & Inoue, 2015). However, to date there have been very few attempts to use these technologies to improve awareness of different perspectives and knowledge sets of collaborative partners, based on concept maps. The analysis from the current study provides an initial framework for the use of concept maps in generating different perspectives, suggesting that the use of conversational agent may yield even better results.

Acknowledgements

This work was supported by the Grant-in-Aid for Scientific Research (KAKENHI), No. 16KT0157 & 16K00219.

References

- Bodemer, D. (2011). Tacit guidance for collaborative multimedia learning. *Computer in Human Behavior*, 27(3), 1079-1086.
- Chi, M. T. H., Leeuw, N. D., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.
- Dillenbourg, P., & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 21, 111-130.

- Engelmann, T., & Hesse, F. W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning*, 5(3), 299-319.
- Hayashi, Y. (2019). Multiple pedagogical conversational agents to support learner-learner collaborative learning: Effects of splitting suggestion types, *Cognitive System Research*, 54, 246-257.
- Hayashi, Y. (2018a). Gaze feedback and pedagogical suggestions in collaborative learning: investigation of explanation performance on self's concept in a knowledge integration task. In R. Nkambou, R. Azevedo, J. Vassileva (Eds), *Proceeding of ITS'18* (pp. 78-87), Montreal, QC, Canada: Springer.
- Hayashi, Y. (2018b). The power of a "maverick" in collaborative problem solving: An experimental investigation of individual perspective-taking within a group. *Cognitive Science*, 42(S1), 69-104.
- Hayashi, Y. (2018c). Using embodied pedagogical agents in collaborative learning by knowledge integration: An experimental investigation on instruction and gaze gestures using eye movement analysis. *The Transactions of Human Interface Society*, 20(1), 79-88.
- Hayashi, Y., & Inoue, T. (2015). Designing collaborative learning by multiple pedagogical conversational agents. *IEICE transactions on Fundamentals A*, J98-A(1), 76-84.
- Hayashi, Y., Miwa, K., & Morita, J. (2007). A laboratory study on collaborative problem solving by taking different perspectives. *Cognitive Studies*, 14(4), 604-619.
- Hayashi, Y., & Miwa, K. (2011). Understanding other's perspectives in conflict communication. *Cognitive Studies*, 18(4), 569-584.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10(2), 151-177.
- Molinari, G., Sangin, M., Dillenbourg, P., & Nüssli, M. (2008). Effects of knowledge interdependence with the partner on visual and action transactivity in collaborative concept mapping. In G. Kanselaar, J.J.G. Van Merriënboer, P.A. Kirschner, T. De Jong (Eds), *Proceeding of ICLS'08* (91-98), Utrecht, The Netherlands: ISLS.
- Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, 21(2), 109-146.
- Roschelle, J. (1992). Learning by collaborative: convergent conceptual change. *The Journal of the Learning Sciences*, 2(3), 235-276.
- Sangin, M., Molinari, G., Nüssli, M. A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, 27(3), 1059-1067.
- Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 371-395.
- Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, 4(3), 321-354.
- Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection, *Cognitive Science*, 26(4), 469-501.
- Vygotsky, L. S., & Cole, M. (1981). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University. Press.

A Crowd-Programming Approach for Computational Thinking Education

Christopher KRIZANOVIC, Nguyen-Think LE & Niels PINKWART

Department of Computer Science, Humboldt-University Berlin, Germany

Nguyen-thinh.le@hu-berlin.de

Abstract: This paper proposes a crowd-programming approach for computational thinking education. An evaluation study was conducted with 61 high school students to investigate four hypotheses: 1) Learners using the computer-supported crowd-programming system require less time for problem-solving, 2) they need less attempts for correct solutions, 3) they tend not to request bottom-out hints for programming problems, and 4) they can solve more programming problems within 70 minutes than in other conditions.

Keywords: crowd-programming, peer-review, pair-programming, computational thinking, didactics for computer science education

1. Introduction

Problem solving has been considered one of the key elements of computational thinking (Wing, 2006). The Computer Science Standard K12 framework developed by the Computer Science Teacher Association (CSTA) characterizes computational thinking as a problem solving process that applies selected strategies (K12CS). Dagiene and colleagues (Dagiene et al., 2016) developed a two-dimensional categorization system for the framework BEBRAS that takes account of computational thinking skills as well as concepts in Computer Science. The BEBRAS framework is also based on problem solving. Unlike the frameworks for computational thinking relying on Computer Science education (CSTA K12 and BEBRAS), the PISA framework for Computational Thinking is supposed to be a part of Mathematics (Lorenceau et al., 2019). This framework considers computational thinking skills as part of the problem-solving process.

In order to support learners in solving programming problems, and thus, in enhancing their computational thinking skills, software development strategies could be involved. In this paper, we present a crowd-programming approach that is based on two software development strategies: peer-review and pair-programming. The crowd-programming approach is deployed in a computer-supported learning system, which is intended to support learners in improving their computational thinking skills.

2. The Crowd-Programming Approach

Peer-review is a software development process that is monitored and reviewed by one or several persons. That is, a software solution of a developer is examined by one or several other developers. It is typically applied to assure the quality of software development and is considered as an indispensably useful software development strategy (Rigby et al, 2008). Adopting the criteria for effective software development strategies (Cockburn & William, 2000), we could derive that the peer-review approach contributes only to solve problems that occur during the review process. This strategy does not involve directly in the problem solving process, i.e. a solution must be developed by the developer. In the context of learning, adopting this approach, if a learner is stuck, he/she would not be able to find a solution for a programming problem. Regarding the criterion for detecting errors in the early stage, this approach must suffer, because the review process takes place at a later time point. Regarding the criterion “learning”, reading peer-review is the opportunity for enhancing their programming skills and also, the task of writing a review for peers’ solutions could be another opportunity to learn new solution strategies.

A special variant of peer-review is the in-flow peer-review that requires the review process to take place *during* the software development process (Clarke et al., 2014). Considering the criteria for an effective software development strategy (Cockburn & William, 2000), the traditional peer-review approach and

the in-flow peer review overlap at several points. Of course, the in-flow peer review approach contributes to problem solving at the early stage of the software development process (e.g., design an algorithm, specification of procedures). The in-flow peer-review approach enables the detection of errors at the early stage.

Pair-programming is a software development strategy that requires two programmers to solve a programming problem together. Both persons discuss and solve the same problem constantly and develop a solution for the problem collaboratively. One person is responsible for coding, another one acts as observer, who permanently checks the code and gives feedback. The pair-programming approach has the advantage that both programmers may make exploit knowledge and programming experience from each other. Thus, problems could be solved faster. Cockburn and William (2000) reported that pair-programming contributes to detecting errors at an early stage of software development. As both developers can contribute their knowledge and programming experience in problem solving, both can benefit from each other.

Although the three presented software development strategies are established and widely deployed in the software development branch, they may suffer several disadvantages for learning purposes. The disadvantage of the peer-review approach is that through the late review task, learners do not have the chance to receive hints for problem-solving at early stages. This disadvantage can be solved by the in-flow peer review, because the review tasks take place during the whole period of problem-solving process. The in-flow peer review approach may also have disadvantages. First, learners, who do not have an interest in solving programming tasks by themselves, could misuse the review task in order to copy solutions from peers. A second disadvantage could be caused by imprecise or incorrect reviews, which could mislead the author(s) of a solution. Also the pair-programming approach may not work advantageous as we wish. If a pair consists of an expert and a novice programmer, in this constellation, the expert programmer would have to manage the software development task alone and the novice programmer would not learn anything from the collaborative work.

In order to circumvent the disadvantages of the software development strategies discussed above, this paper proposes the so-called crowd-programming approach. This approach is intended to enable the learner to access hints from peers. Hints are not limited to one single peer, but are created as a collective pool by all learners. Thus, the learner can use the pool of hints and choose the appropriate ones to carry on to solve a specific problem. Learners do not have to wait to the end of a review process (like for the case of peer-review), hints can be provided at real-time. The computer-supported system applying the crowd-programming approach can collect hints (that have been given in the past) regarding a specific problem and generate a review for the learner.

After consulting a review (in form of a list of hints) for the specific problem and once the learner is able to solve that problem, he/she is required to evaluate those hints. The evaluation of hints is referred to as meta-review in the crowd-programming approach. This mechanism helps the system to filter out reviews that have been submitted randomly and are not relevant to the specific problem and to generate only reviews with highest meta-review score. The mechanism of optimizing the quality of review to be provided is intended to help the learner to solve the specific problem without being misled by incorrect/irrelevant reviews. This approach makes use of the collective intelligence from the crowd, thus, the approach being presented in this paper is called crowd-programming approach. The research question to be investigated in this paper is: how effective can the crowd-programming approach support learners as they solve programming problems?

3. SUPELS: A Crowd-programming Based System for Computational Thinking Education

3.1 Requirements

The aim of this system is to apply the crowd-programming approach to support students learn the foundations of programming and to improve their competency in solving programming problems. Thus, the system will have to provide learners with a programming course that consists of lessons. Each lesson is composed of an introduction text for a new programming construct and exercise assignments. The task of the learners is to read the introduction text and to solve those exercise assignments. As a programming language to be learned, JavaScript was suggested, because this programming language is not commonly taught at schools. This is to avoid possible biases in the experiment study to be conducted. If the study subjects had learned the programming language, they would solve the exercises

provided by the system quickly. The topics of the ten lessons are: syntax of JavaScript, comment, variable, variable modification, function, data types and Boolean, logical operators and comparatives (for number), logical operators and comparatives (for Boolean), If-statement, For-statement. The introduction texts and exercise texts should be expressed in an understandable way for high school students. Thus, the composition of ten lessons above was consulted from the books “JavaScript for Kids” (Morgan, 2014) and “JavaScript for Dummies” (Harris, 2012).

3.2 The Problem-solving Scenario

Introducing New Lessons: The first user interface is used to display the introduction text and the corresponding exercise assignment. The introduction text can be folded up after reading in order to focus only on the exercise assignment and can be folded out according to the need of the learner. This adaptive presentation is an important feature required for educational systems to adapt to individual memory capacity (Lestari et al., 2017).

Solving Exercises: The second user interface is provided to solve an exercise assignment. To solve an exercise, a text editor is required and several buttons for different functions (e.g., test code, reset code, show hint, show bottom-out hint) are required. Depending on the lesson and the exercise, the text editor is empty or display with code that requires the learner to extend or to adapt. The function “test code” runs the code against a test bed and compares the submitted code with correct solutions (that have been edited by the exercise author or that have been submitted by the learner and automatically validated by the system as correct). If the submitted code is evaluated as incorrect, the learner is asked to consult peer-review using the function “show hint”. After three unsuccessful attempts, the learner will have the possibility to use the function “show bottom-out hint”, which shows the correct solution for the exercise assignment.

Displaying Hints: The third user interface (Figure 1) is used to display hints. The displayed hint can be meta-reviewed by the learner as positive (thumb up) or negative (thumb down). If one hint is not enough, the learner can request the next hint (because there is a collective pool of hints for a specific problem). In order to select appropriate hints from the collective pool, the programming code submitted by the learner needs to be analyzed. The code analysis is based on the comparison between the learner’s code and the anticipated correct solution and all differences are identified. The differences indicate what the learner has to modify on his/her code in order to reach a correct solution. The coding state, at which the learner asks for hints, is the specific problem state of the learner. The system looks for corresponding reference coding state in the database and choose associated hints.

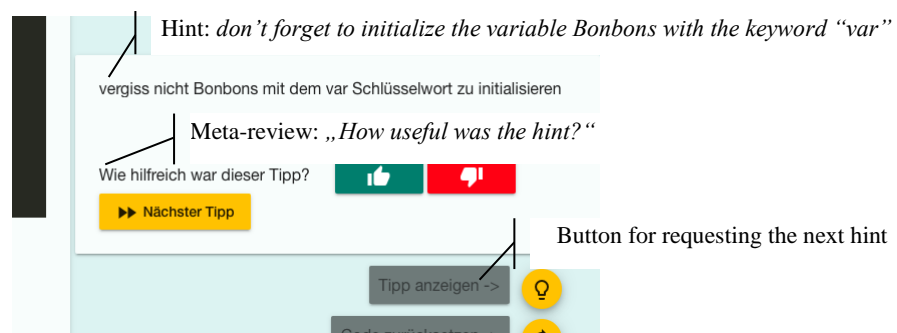


Figure 1. The user interface for displaying requested hints

In order to select the best hints in the collective pool for a specific coding state, the distance between the learner’s code and the reference coding state is computed using the Levenshtein algorithm, because algorithm is commonly used as a metric for string matching (Navarro, 2001). The Levenshtein distance is computed by the number of deletion, addition, and replacement operations that are required to transform the learner’s code to the reference coding state (or vice versa). The algorithm for choosing the best hints follows. The Levenshtein distance between the learner’s code and the referenced code of a feedback is computed. Then, the distance is divided by the length of the referenced code. If the division value is 1, the Levenshtein distance and the length of the referenced code are equal. The best hint is chosen, if it has not been shown and its division value is highest (i.e., closed to value 1). In addition to Levenshtein distance, meta-review values for hints are also considered. If two hints have a difference of the division value within the threshold of 0.2, the hint with higher meta-review value will be chosen.

General feedback (e.g., regarding coding style) that does not have a referenced code will also be selected if it has not been displayed yet and has a highest meta-review value.

Inputting hints: For inputting hints, the fourth user interface (Figure 2) is required. Those learners who have completed an exercise assignment successfully, are asked to input hints and suggestions for helping peers, who will have to go through the lessons later. The procedure for inputting hints is following. When the learner receives the information that his/her solution is correct, he/she is asked to leave hints which may be helpful for peers later. Then, the learner should think about which code line(s) were difficult for him/her and would be problematic for peers. The learner should mark those code lines (which are referred to as reference coding state) and give hint. A learner can input several hints and different reference coding states.

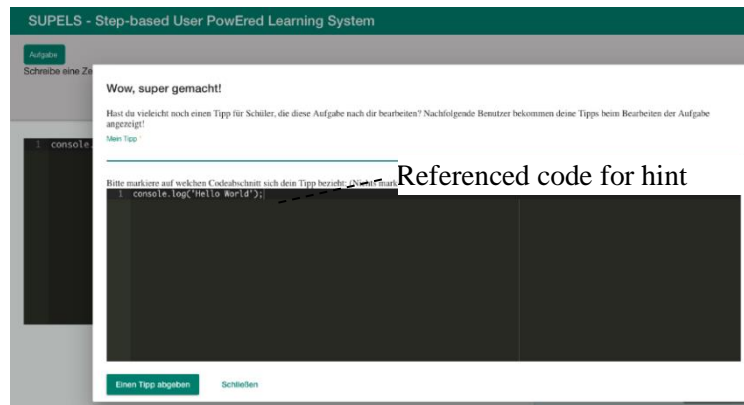


Figure 2. The user interface for inputting hint

4. Evaluation

4.1 Goal

The goal of the study addressed in this paper is to investigate the research question: how effective can the crowd-programming approach support learners to solve programming problems? Since the proposed crowd-programming approach is intended to develop computational thinking, the target group of the developed system includes high school students, who have just started learning Computer Science at schools.

4.2 Design and Participants

The study consists of three phases: 1) a pre-test (10 minutes), 2) the experiment phase (70 minutes), and 3) a post-test (10 minutes). The pre-test is intended to measure knowledge and programming skills of students. The post-test serves to measure the improvement of problem-solving performance after using the learning system. The pre-test and post-test consist of test items for programming constructs that are addressed in the ten lessons provided by the learning system.

A post-test was given to the participants of all conditions after completion of the experiment session. Pre- and post-test were made comparable by a counter-balanced design of the test items. Specifically, two test versions were developed: Test A and Test B. Test A was assigned as pre-test and Test B as post-test to 50% of the participants of each condition, and the rest of the control group had Test B as pre-test and Test A as post-test. The whole process including pre-test, experiment session, post-test, and questionnaire was limited to 90 minutes. Each test can be achieved with a maximum of 21 points. Learning gains are defined as the difference between the score of post-test and pre-test.

Prior to the pre-test, information about the learners were collected. This includes information about the experience of the learners, experience with JavaScript. In case the learner did not have experience in programming, a further question was asked about why he/she has not learned programming. In addition to the post-test, students get a questionnaire for evaluating each learning condition (i.e., using the computer-supported crowd-programming system, pair-programming setting and the control group without help). Each question was rated on the scale between 1 (very bad) and 5 (very good).

During the experiment period, numbers of submission attempts, time required for solving each exercise assignment, requests for bottom-out hints, frequency of requests for feedback were recorded. Student solutions of tests and of experiment exercises were collected to be used for analysis. In addition, students' responses to the questionnaire were used as subjective data.

We contacted high schools in Berlin and could acquire 64 students from grade 8 to 12. Since 3 Pre/Post-tests were missing, these data entries could not be considered. At the end, we had a sample size of 61 participants that were distributed in three conditions whereas more participants were assigned to the experiment condition than in other conditions because the SUPELS system should be evaluated not only quantitatively but also qualitatively: Condition 1: 30 participants in the experiment group using the SUPELS system; Condition 2: 20 participants in the condition with pair-programming using a software development environment; Condition 3: 11 participants in the control group (using introduction texts from course books for solving JavaScript programming exercises).

4.3 Results

Hypothesis 1a: Learners in Condition 1 requires less time for problem-solving than in other conditions.

Hypothesis 1b: Learners Condition 1 needs less solution submission attempts than in other conditions.

Hypothesis 1c: Learners in Condition 1 less request bottom-out hints than other conditions.

Hypothesis 1d: Learners in Condition 1 achieves higher number of lessons in a specific time period (70 minutes) than in other conditions.

In order to test the hypotheses 1a, 1b, and 1d, the one-factorial variance analysis ANOVA can be used to compare the difference between two independent groups, because each group has an independent variable. Considering Hypothesis 1a, the dependent variable is time required for solving exercise assignments. Applying the Shapiro-Wilk-Method, time required for solving exercises with the Gaussian distribution was not normally distributed ($\alpha=0.05$) for all three conditions. However, research studies showed that one-factorial ANOVA is robust for violation of normal distribution (Lix et al., 1996; Salkin, 2010) and the violation of normal distribution is not critical if the data sample size is over 25. The Levene-Test shows that homogeneity of variance is present ($p=0.092$). Thus, the ANOVA analysis is valid. For testing Hypothesis 1a, there were 274 data entries of crowd-programming, 95 of pair-programming, and 109 of control group. Each data entry represents time required for one lesson (reading an introduction text and solve an exercise successfully) by one subject, and thus is independent. Table 1 shows that the crowd-programming group required more time ($m=169.3$; $sd=159$) than other groups, but the difference between the groups is statistically not significant ($F(2, 475)=0.88$; $p=0.42$) at a significance level of 0.05. Thus, Hypothesis 1a can be rejected.

Table 1
Results for Hypotheses 1a and 1b

	Condition 1	Condition 2	Condition 3	F	p
Time for problem-solving (seconds)	169.3 (sd=159)	161.6 (sd=145.2)	146.5 (sd=147)	0.88	0.42
# submission attempts	2.66 (sd=5.34)	3.86 (sd=3.12)	2.87 (sd=2.48)	2.60	0.08
# completed lessons	9.13 (sd=2)	9.5 (sd=1.08)	9.9 (sd=0.3)	2.58	0.10

Regarding Hypothesis 1b, there were also 274 data entries of crowd-programming, 95 of pair-programming, and 109 of control group. Each data entry represents the number of submission attempts required by each subject to achieve the correct solution for each exercise assignment, and thus is independent. Table 1 shows that the pair-programming group ($m=3.86$; $sd=3.12$) needed more submission attempts than other groups, but the difference between the conditions is statistically not significant ($F(2, 475)=2.6$; $p=0.075$) at a significance level of 0.05. Hypothesis 1b can be rejected.

Table 2
Results for Hypothesis 1c

B (coefficient)	SD	Df	p	Exp(B)
-0.63	0.19	1	0.001	0.53

For Hypothesis 1c, 274 data entries of crowd-programming, 95 of pair-programming, 109 of control group were collected. Each data entry represents whether a learner requested the bottom-out hint (i.e., correct solution provided by the system) or developed a correct solution for an exercise assignment by

him-/herself. Thus, the data entries are independent. The probability of requesting a bottom-out hint is 0.49 (sd=0.5) for the crowd-programming group, 0.72 (sd=0.45) for the pair-programming group, and 0.58 (sd=0.5) for the control group, respectively. This statistical result shows that learners in the crowd-programming group requested bottom-out hint least frequently. In order to examine whether the difference between the groups is statistically significant, logistic regression is used for analysis. Since the dependent variable only has the value 0 (learner has not requested the bottom-out hint) and value 1 (learners has requested the bottom-out hint), the assumption for binomial distribution is satisfied. Table 2 shows that the beta coefficient is negative ($B = -0.63$) with a standard deviation of 0.19 and Odds Ratio is less than 1 ($\text{Exp}(B)=0.53$). Thus, the probability that a learner of the crowd-programming group requesting bottom-out hints decreases by 63%. In addition, this probability is significant ($p=0.001$). For testing Hypothesis 1d, the number of completed lessons is calculated, which ranges between 0 and 10. A completed lesson means that a learner has solved the associated exercise assignment successfully. For 10 lessons, there were 70 minutes. Table 1 shows that the control group could ($m=9.9$, $sd=0.3$) complete more lessons than the pair-programming group ($m=9.13$, $sd=2.0$) followed by the crowd-programming group ($m=9.37$, $sd=1.62$). Note, the standard deviation of the pair-programming group is relatively high. The result of WELCH test $F(2, 20.54)=2.58$ shows that the difference between the groups is statistically not significant ($p=0.10$) at a significance level of 0.05. Thus, Hypothesis 1d can be rejected.

5. Conclusions

This paper presents the crowd-programming approach to be deployed in a computer-supported learning system for computational thinking education. The crowd-programming approach exploits the advantageous characteristics of the software development strategies: in-flow peer-review and pair-programming. The current evaluation study investigates the research question whether the crowd-programming approach is an effective approach to improve problem-solving performance. The study shows that the crowd-programming based learning system did not help students reduce time for problem-solving, neither did learners of this condition require less solution submission attempts or achieved more lessons and exercise assignments than in other conditions. However, learners applying the crowd-programming approach requested fewer bottom-out hints than in other conditions. That means, crowd-programming learners tend rather to develop their own solutions than simply ask for correct solutions (bottom-out hints). The behavior of self-developing solutions would more help learners to develop problem-solving skills than merely requesting correct solutions. Authors of this paper are investigating other impacts of the crowd-programming approach (i.e., learning gains) and analyzing subjective attitudes of learners. These results are going to be published in near future.

References

- Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., ... Zotov, A. (2007). The ups and downs of peer review. *American Journal of Physiology - Advances in Physiology Education*, 31(2), 145-152.
- Clarke, D., Clear, T., Fisler, K., Hauswirth, M., Krishnamurthi, S., Politz, J. G., Tirronen, V., & Wrigstad, T. (2014). In-Flow Peer Review. In *Proceedings of the Working Group Reports of the Innovation & Technology in Computer Science Education Conference*, ACM, 59-79.
- Cockburn, A. & Williams, L. (2001). The costs and benefits of pair programming. In *Extreme programming examined*, G. Succi & M. Marchesi (Eds.). Addison-Wesley Longman Publishing Co., 223-243.
- Dagiene, V., Sentence, S. & Stupurienė, G. (2016). Developing a fine-grained two-dimensional categorization system for educational tasks in informatics. *Informatica* 28(1):23-44.
- Harris, A. (2012). *JavaScript für Dummies*. Publisher: Wiley-VCH Verlag.
- Lestari, W., Nurjanah, D., Selviandro, N. (2017). Adaptive Presentation Based on Learning Style and Working Memory Capacity in Adaptive Learning System. In *Proceedings of the 9th International Conference on Computer-Supported Education*.
- K12CS (2016). K-12 Computer Science Framework. Retrieved from <https://k12cs.org/computational-thinking>
- Lorceau, A., Marec, C. & Mostafa, T. (2019). Upgrading the ICT questionnaire items in PISA 2021 OECD, Education Working Paper No. 202.
- Lix, L., Keselman, J., & Keselman, H. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance "F" Test. *Review of Educational Research*, 66(4), 579-619.
- Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*. 33 (1): 31–88.

- Morgan, N. (2014). JavaScript for Kids. Publisher: No Starch Press.
- Rigby, P., German, D., & Storey, M. (2008). Open source software peer review practices. In *Proceedings of the 30th International Conference on Software Engineering*, pp. 541-550. DOI: 10.1145/1368088.1368162
- Salkind, N. J. (Ed.) (2010). Encyclopedia of research design Thousand Oaks, CA: SAGE Publications, Inc. DOI: 10.4135/9781412961288
- Wing, J.M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35.

Using Knowledge Forum to Support the Development of STEAM Literacies

Leanne MA^{a*} & Thelma AKYEA^a

^aOISE/University of Toronto, Canada

*leanne.ma@mail.utoronto.ca

Abstract: Over the last decade, there has been growing interest in Science, Technology, Engineering, Arts, and Math (STEAM) education to promote creativity and design thinking in schools. Knowledge Building, which engages students in sustained, interdisciplinary work with ideas, represents one pedagogical approach to achieve these goals, while Knowledge Forum technology facilitates the implementation of STEAM education in classrooms. This study explores the simultaneous development of 8- and 9-year-old's scientific literacy, graphical literacy, and mathematical literacy in Knowledge Forum. Between fall and winter term, a grade 3 class studied growth and changes in plants. Qualitative analyses on student work in Knowledge Forum reveal increasingly proficient use of expert vocabulary to explain scientific phenomena, as well as productive use of increasingly complex graphical and mathematical representations to explore socio-scientific issues. Study implications are discussed within the context of STEAM education and assessment of 21st Century competencies.

Keywords: Knowledge Forum, STEAM, scientific literacy, graphical literacy, mathematical literacy

1. Introduction

Creativity is becoming widely recognized as one of the most important competencies of the 21st century (e.g., OECD, 2018; World Economic Forum, 2016). Given the growing need to build collective capacity for creativity, educators and policymakers have been advocating to expand Science, Technology, Engineering, and Math (STEM) education into Science, Technology, Engineering, Arts, and Math (STEAM) education (Butler & Luke, 2012; Liao, 2016). STEAM education is defined as:

“The inclusion of the liberal arts and humanities in STEM education; some STEAM conceptions simply use the ‘A’ to indicate a fifth discipline area— namely, arts and humanities, with sub-disciplines as have historically existed for STEM areas; however, an alternative conception is to integrate liberal arts and humanities into STEM education as an expansion of an expanded crossdisciplinary approach being advocated for STEM education.” (Spector, 2015, p. 5)

One of the core aims of STEAM education is to make design thinking an explicit process of knowing and learning to facilitate the integration of the disciplines (Bequette & Bequette, 2012) and the development of students' creativity (Maeda, 2012; Henriksen, 2014). This perspective is in line with Knowledge Building (Scardamalia & Bereiter, 2014), a pedagogical approach that engages students in sustained creative work with ideas. During Knowledge Building, students work together to tackle complex, real-world problems. They generate theories and hypotheses, critique explanations, and examine multiple sources of evidence from different perspectives before refining their theories to advance collective understanding. An online platform called Knowledge Forum (Scardamalia, 2017) is commonly used as the community design space for students to share ideas, build on ideas, take them apart, and reorganize them in more conceptually coherent ways. Past studies provide evidence that the integration of Knowledge Forum into daily classroom practices significantly enhances students' scientific literacy (Sun & Zhang, 2010), mathematical literacy (Moss & Beatty, 2010), and graphical literacy (Gan, Scardamalia, Hong, & Zhang, 2010).

Given the interdisciplinary nature of Knowledge Building and Knowledge Forum, it follows that Knowledge Forum is a promising technology for engaging students in STEAM-related activities. This study explores how students can simultaneously develop scientific literacy, graphical literacy, and

mathematical literacy through sustained collaborative work with ideas in Knowledge Forum. Over the span of eight months, 22 students in a grade 3 class in Toronto, Canada studied growth and changes in plants (Ontario Ministry of Education, 2007). Students wrote over 400 notes across 9 views in Knowledge Forum on topics such as, “Do Plants Grow”, “Plant Roots”, “Worms and Leaves”, “The Bean Video”, “Plant Experiments”, “Photosynthesis”, “Photosynthesis Seen from Space”, and “Travelling Inside a Leaf”. On average, each student wrote 20 notes, read 62 notes, and created 4 drawings, suggesting they were actively engaged in multimedia-rich activities on Knowledge Forum. In the sections below, we traced qualitative shifts in students’ contribution patterns across the three different literacies between the fall term and the winter term.

2. Scientific Literacy

Scientific literacy is “the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen” (OECD, 2016). This includes the ability to explain phenomena, design experiments, and interpret data. In Knowledge Forum, students can contribute their ideas as notes to be further built on by their peers. The note editor includes a textbox, epistemic markers, and keyword tags (see Table 1) to facilitate the learning of new concepts and ideas. In this section, the analysis will focus on shifts in students’ vocabulary use in their notes as they explain the scientific phenomenon of photosynthesis.

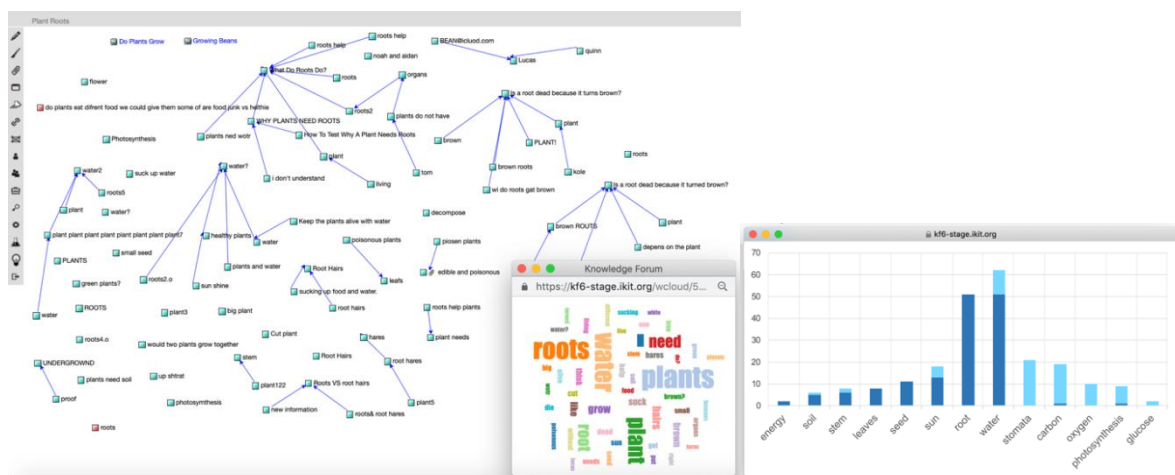


Figure 1. Knowledge Forum view and analytic tools, Word Cloud (left), Lexical Analysis (right).

Figure 1 shows one of the initial views on plants and the word cloud (left) associated with that view. Table 1 (top row) shows examples of students’ notes when they were sharing questions and theories of how plants grow and change. Intuitively, students were able to identify key characteristics of the environment, such as “water”, “sun”, and “soil”, as well as key characteristics of plants, including “stem”, “leaves”, and “roots” that play an important role in photosynthesis. After designing experiments and watching videos about photosynthesis in the fall term, students became more familiar with the process of photosynthesis and started adopting more expert vocabulary in the winter term. Table 1 (bottom row) shows examples of students’ notes when they were discussing the importance of “stomata” in photosynthesis and how it controls the flow of “oxygen” and “carbon dioxide” between plants and their environment. The lexical tool in Figure 1 (right) shows a drastic shift in students’ vocabulary use between fall term (dark blue) and winter term (light blue). Whereas in the fall term, students mainly used “water”, “root”, “sun”, “seeds”, “leaves”, “stem”, and “soil” to describe the process of photosynthesis, in the winter term, students mainly used “stomata”, “carbon”, “oxygen”, “water”, “sun”, and “glucose” to describe the process of photosynthesis. It can be seen that as students gained a deeper understanding of photosynthesis, their vocabulary became more precise and scientific.

Table 1

Examples of Students' Notes from Fall Term and Winter Term

View	Note
Do Plants Grow	<p>[My theory]: All plants grow cause they have a stem and if you break the stem and do nothing with it, it dies.</p> <p>[My theory]: They grow because of water and sun.</p> <p>[I wonder]: how do plants survive with only water and soil.</p> <p>[My theory]: is a plant needs roots to survive.</p> <p>[My theory]: is that not all plants grow because if the plant has no leaves that can cause them not to grow.</p>
Photosynthesis	<p>[A new word I learned]: was that stomata is the thing that lets all water and carbon dioxide in and lets out oxygen.</p> <p>[A new fact I learned]: is that a plant needs water and sunlight.</p> <p>[My improved theory is]: that the sun helps the plant breath.</p> <p>[I still need to understand]: how the stem sucks up water?</p> <p>[My theory]: is that there is lots of stuff in a plant like stomata and glucose and oxygen and alot more.</p>

3. Graphical Literacy

Graphical literacy is “the ability to construct meaning from visual experiences, organize information based on the composition of images, and use symbols to express meaning” (Nelson, 2004). This includes the ability to learn, think, and communicate with non-textual representations. In Knowledge Forum, students can contribute their ideas as drawings to be further built on by their peers. The drawing editor includes colours, lines, shapes, symbols, attachments, and text to facilitate the creation of complex, multi-layered graphical representations. In this section, the analysis will focus on students’ different uses of drawings to advance the community discourse on photosynthesis.

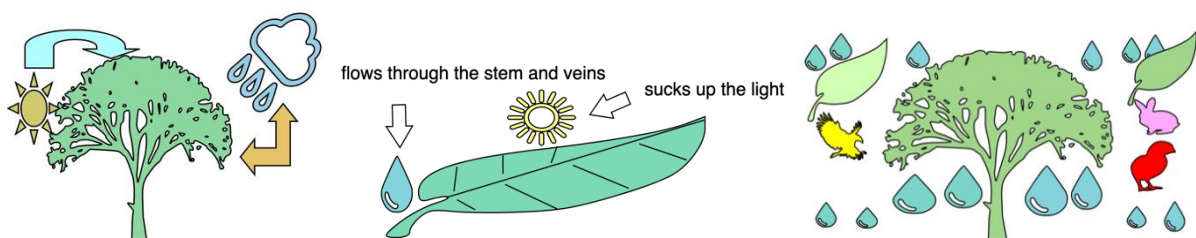


Figure 2. Three different graphical representations related to photosynthesis.

In the winter term, students were introduced to the drawing tool to elaborate on their understanding of photosynthesis. Figure 2 shows three drawings by three different students on the process of photosynthesis that were created a few weeks apart from one another. It can be seen that students used colours, symbols, arrows, and labels as means to graphically represent their ideas. The first drawing shows photosynthesis as a cycle between the sun, water, and trees, thus reinforcing their prior understanding that plants need sun and water to grow. The second drawing shows photosynthesis at a micro-level, highlighting the role of different parts of the leaf (i.e., stem, veins, stomata). The student also added their new understanding that roots and stems act as the transport system in plants – a new fact they learned from watching a video. The third drawing shows photosynthesis at a macro-level, highlighting the relations between water, plants, and animals in ecosystems. After learning about the process of photosynthesis, students came to understand the symbiotic relationship between humans and plants (i.e., humans depend on plants for oxygen and plants depend on humans for carbon dioxide). The last drawing led students to wonder whether photosynthesis is as important for animals as it is for humans. Table 2 shows examples of students’ notes as they debated this controversial issue. It can be

seen that some students became very passionate about the issue, instilling a sense of environmental stewardship in the entire community as the discussion shifted organically toward how they could save living things on the planet from the perils of climate change.

Table 2

Examples of Students' Notes from Winter Term

View	Note
Photosynthesis Seen from Space	Is photosynthesis important for animals too? I think so because humans are animals. Birds, animals, and insects help in the cycle of life!!!!!!! Do you think that animals could live with little carbon? BUILDINGS ARE KILLING PLANTS AND THERE ARE LESS AND LESS ON EARTH... BUILDINGS ARE ALSO KLLING ANIMALS AND WITHOUT PLANTS ANIMALS CAN'T LIVE AND ALSO WITHOUT PLANTS THERE WILL BE NO PHOTOSYNTHESIS AND WE WILL DIE WITH NO AIR.

4. Mathematical Literacy

Mathematical literacy is the ability to “formulate, employ, and interpret mathematics in a variety of contexts... [This] includes making mathematical deductions and applying mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena” (OECD, 2016). Because there is no feature in Knowledge Forum designed specifically to support mathematical literacy, students were not expected to share their understanding through numbers, formulae, and mathematical expressions. Surprisingly, a few students did. In this section, the analysis will focus on students’ uses of numbers and charts to enrich their collective understanding on photosynthesis.

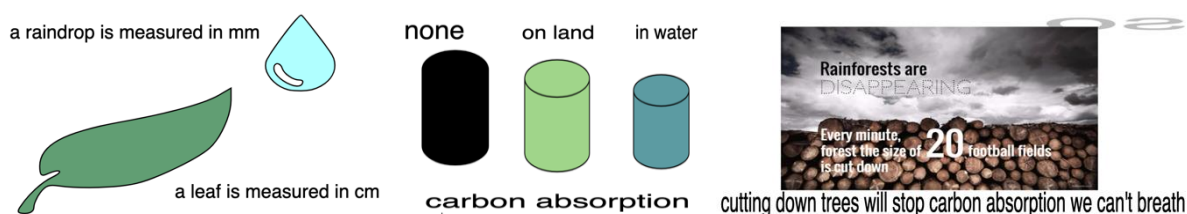


Figure 3. Three different mathematical representations related to photosynthesis.

After studying the process of photosynthesis in the fall, students worked in new Knowledge Forum views in the winter to explore issues at a micro-level (Travelling Inside a Leaf) or at a macro-level (Photosynthesis Seen from Space). In the view, Travelling Inside a Leaf, students watched a video animation on the cellular structures inside leaves. Notes and drawings in this view explored the size of cells and how energy is created in the leaf. Figure 3 (left) shows a comparison between the size of a raindrop and the size of a leaf, as well as their respective unit of measurement. In the view, Photosynthesis Seen from Space, students watched a time-lapse video of carbon absorption in different places in the world over the span of a year. Notes and drawings in this view explored similarities and differences across different countries, as well as changes in the amount of carbon absorption during different seasons of the year. Figure 3 (center) shows a comparison between the proportional amount of carbon absorption on earth’s land water, and places where there was no carbon absorption at all. This drawing led students to hypothesize that places where there was little or no carbon absorption had higher levels of pollution. Table 3 shows examples of students’ notes as they discussed issues surrounding carbon absorption and carbon emissions (i.e., pollution) on earth. Figure 3 (right) shows an image a student found that features a statistic on deforestation. The student further adds their theory that large-scale deforestation will disrupt the naturally occurring carbon absorption patterns on earth. It can

be seen that students used mathematical concepts to effectively describe, explain, and predict scientific phenomena. Furthermore, students were aware of the unit of analysis across the two topics – Travelling Inside a Leaf and Photosynthesis Seen from Space – which spanned from micrometers inside a cell to kilometres around the globe.

Table 3

Examples of Students' Notes from Winter Term

View	Note
Photosynthesis Seen from Space	<p>[My theory]: is that there are a lot of trees and plants in the green and in the dark part there is less trees and plants.</p> <p>I think that the colours represent different environments. Like light green... represents land that has been not in pollution... [and] dark green represents land that has been in pollution.</p> <p>The equator has a lot of photosynthesis. North America has none... - barely any in Antartica, summer or winter... Australia there is less photosynthesis in Australia due to ground pollution.</p> <p>[My theory]: pollution is terrible because of pollution animals and plants are dying. Using plastic is water pollution. Cars make air pollution that's killing the environment.</p>

5. Discussion

Over the span of 8 months, students worked collaboratively on Knowledge Forum to advance their understanding of growth and changes in plants. They shared their initial ideas and theories through notes, reflected on their peers' ideas and new information from authoritative sources (e.g., videos), before building on with new notes and new drawings. They broke down key ideas from videos into notes, drawings, and graphs, which helped deepen collective understanding and spark new cycles of theory development. To summarize, students wrote notes about the process of photosynthesis (e.g., key characteristics of plants, the role of oxygen and carbon dioxide) and climate change (e.g., rates of carbon absorption/emission, the effects of deforestation and pollution). They created drawings about the process of photosynthesis (e.g., the role of sun and water, anatomy of a leaf, plant-animal relations in ecosystems). They even created drawings to show proportional relations (e.g., size of raindrop vs. leaf, carbon absorption on land vs. water). Knowledge Forum provided flexible and dynamic supports for students to engage in these rich STEAM-related activities, thus promoting the simultaneous development of their scientific literacy, graphical literacy, and mathematical literacy. Between fall term and winter term, students demonstrated an increasing proficiency with expert vocabulary to explain photosynthesis, and they created increasingly complex graphical and mathematical representations to explore socioscientific issues, such as air/water pollution, deforestation, and climate change.

It is interesting to note that after the teacher invited students to shift from textual to graphical representation of their ideas in the winter term, students themselves initiated the subsequent shift from graphical to mathematical representation of their ideas. Put differently, as students' ideas evolved and matured, so did their medium of expression regardless of content area. For this reason, we believe that our study reinforces the notion that “graphical literacy and deep understanding are mutually reinforcing, with graphical literacy serving as a powerful thinking tool across content areas” (Gan, Scardamalia, Hong, & Zhang, 2010). Our study also adds that engaging students in STEAM-related activities in Knowledge Forum can serve as a powerful mechanism to sustain ever-deepening, creative, collaborative work with ideas.

5.1 Study Implications

While creativity development may be a top priority in 21st century education, it is equally important to ensure that students develop the foundations of scientific literacy, graphical literacy, and mathematical literacy throughout their schooling experiences. The design challenge for teachers is to create an interdisciplinary space that weaves together Science, Technology, Engineering, Arts, and Math in a

meaningful and authentic way for students. Our study suggests that with appropriate technological supports, students themselves can also be co-designers of STEAM-related activities in the classroom. Students as young as 8 years of age are able to move seamlessly across the disciplines when they are engaged in authentic knowledge work, thus reinforcing the need for interdisciplinary assessments in STEAM education. Liao (2016) claims that STEAM education “affirms the process of creative production, utilizes the creative process to acquire knowledge, and teaches 21st-century skills”. Future studies should explore the development of new features in Knowledge Forum to expand possibilities for creative production, such as a video component (e.g., slideshow, animation, annotation) in the drawing tool, and a mathematical functions component (e.g., algebraic equations, graphing capabilities) in the note editor. For example, students in our class would have enjoyed creating videos to illustrate the life cycle of a plant and/or creating graphs to compare the levels of carbon emission by country. Such rich multimedia-rich activities have the potential to further enhance their Knowledge Building. We believe that to be active citizens thriving in 21st century societies, students must harness their creativity to not only acquire knowledge, but also to create knowledge for public good.

Acknowledgements

We are grateful for the opportunity to work with staff and students at the Dr. Eric Jackman Institute for Child Studies. Their creative ideas and insights have been invaluable to our Knowledge Building.

References

- Bequette, J. W., & Bequette, M. B. (2012). A place for art and design education in the STEM conversation. *Art Education*, 65(2), 40-47.
- Butler, S. & Luke, J. (2012). The art of science learning: Shaping the 21st-Century workforce. USA: Institute for Learning Innovation. Retrieved from http://www.artofsciencelearning.org/wp-content/uploads/2013/09/ASL-1-1.SummativeEvalReport.FINAL_.pdf
- Gan, Y. C., Scardamalia, M., Hong, H. Y., & Zhang, J. (2010). Making thinking visible: Growth in graphical literacy, grades 3 to 4. *Canadian Journal of Learning and Technology*, 36(1). Retrieved from <https://www.cjlt.ca/index.php/cjlt/article/view/26374>
- Henriksen, D. (2014). Full STEAM ahead: Creativity in excellent STEM teaching practices. *The STEAM Journal*, 1(2). Retrieved from <http://scholarship.claremont.edu/steam/vol1/iss2/15>
- Liao, C. (2016) From interdisciplinary to transdisciplinary: An arts-integrated approach to STEAM education, *Art Education*, 69(6), 44-49.
- Maeda, J. (2013). STEM + art = STEAM. *The STEAM Journal*, 1(1). Retrieved from <https://scholarship.claremont.edu/steam/vol1/iss1/34/>
- Moss, J., & Beatty, R. (2010). Knowledge building and mathematics: Shifting the responsibility for knowledge advancement and engagement. *Canadian Journal of Learning and Technology*, 36(1). Retrieved from <https://www.cjlt.ca/index.php/cjlt/article/view/26368>
- OECD (2016), PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, PISA. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264255425-en>
- OECD (2018). Teaching for Global Competence in a Rapidly Changing World. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264289024-en>
- Ontario Ministry of Education. (2007). *The Ontario Curriculum, Grades 1-8: Science and Technology*. Retrieved from: <http://www.edu.gov.on.ca/eng/curriculum/elementary/scientec18currb.pdf>
- Scardamalia, M. (2017). Knowledge Forum. In K. Peppler (Ed.), *The SAGE Encyclopedia of Out-of-School Learning* (pp. 401-403). Thousand Oaks, CA: SAGE.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed.) (pp. 397-417). New York: Cambridge University Press.
- Spector, J. M. (2015). Education, training, competencies, curricula and technology. In X. Ge, D. Ifenthaler, & M. Spector (Eds.), *Emerging Technologies for STEAM Education* (pp. 3-14). Switzerland: Springer.
- Sun, Y., Zhang, J., & Scardamalia, M. (2010). Developing deep understanding and literacy while addressing a gender-based literacy gap. *Canadian Journal of Learning and Technology*, 36(1). Retrieved from <https://www.cjlt.ca/index.php/cjlt/article/view/26369>
- Watson, A. D. (2015). Design thinking for life. *Art Education*, 68(3), 12-18.
- World Economic Forum. (2016, January). The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution. Geneva, Switzerland. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Surveying the Relationships Between Students' Epistemic Curiosity and Their Online Academic Help Seeking Behaviors in Higher Education

Kun-Hung CHENG^{a*}

^a*Graduate Institute of Library and Information Science, National Chung Hsing University, Taiwan*

*khcheng@dragon.nchu.edu.tw

Abstract: This study initiated an attempt to explore the role of students' epistemic curiosity, the desire to obtain intellectual information to fill the gaps in one's knowledge, in their online academic help seeking behaviors in higher education. There were 113 undergraduate and graduate students surveyed in this study. The results showed that the students had more experiences of searching information online and strong epistemic beliefs about seeking help when encountering academic challenges. It was also found that while the beliefs about perceived benefits from online academic help seeking might foster the students' behaviors of information searching and formal query from teachers, their experiences of informal query from peers were significantly predicted by their perceptions of self-efficacy for seeking helps. There were strong significant relationships between the students' I-type epistemic curiosity (stimulating intellectual interest) and their experiences of information searching. On the other hand, the students' D-type epistemic curiosity (eliminating feelings of informational deprivation) was strongly related to their experiences of formal and informal query. Moreover, I-type epistemic curiosity may play a more important role in the students' beliefs about online academic help seeking than D-type epistemic curiosity does. This study is still ongoing with larger sample for verifying the academic evidences.

Keywords: Epistemic curiosity, online academic help seeking, self-efficacy, higher education

1. Introduction

Epistemic curiosity is an emotional-motivational state that individuals desire to acquire intellectual information (Berlyne, 1954); and it has been explored and discussed with other personality traits such as ambiguity tolerance and need for closure previously in the field of psychology (e.g., Litman, 2010). Epistemic curiosity aroused by being aware of gaps in one's knowledge was defined as two types including "interest" (I-type) and "deprivation" (D-type). While I-type epistemic curiosity involves intentions to discover something new for inherent enjoyment, D-type epistemic curiosity involves intentions to seek specific information for reducing undesirable states of uncertainty caused by informational deprivation (Litman, 2008).

Research has identified the significant association between learners' epistemic curiosity and their intellectual development and learning behaviors, particularly for information seeking (Litman et al., 2005). Students' experiences of seeking help online for academic problems has been proposed as the concept of online academic help seeking and explored through three aspects including information searching, formal query (asking question from teachers), and informal query (asking question from peers) (Cheng & Tsai, 2011). Recent studies (Liu, 2017) further probed adult students' beliefs about online help-seeking through three psychological factors including epistemological beliefs (considering help-seeking as knowledge source), perceived benefits (perceiving merits of help-seeking), self-efficacy (perceiving competency to seek help). Following the previous studies, this study attempted to re-examined adult students' experiences of and beliefs about online academic help seeking when facing abundant information from the Internet at the current stage. To obtain more understandings of online academic help seeking behaviors in higher education, what role of adult students' epistemic

curiosity plays in their experiences of and beliefs about seeking help online when encountering academic challenges was initially explored in this study.

2. Method

2.1 Respondents

The respondents of this study were 113 undergraduate and graduate students in Taiwan. All of them were recruited through the Internet. The mean age of the adults was 28.11 years old ($SD=10.32$). Some elder students enrolled in in-service master program (about 12%). Among these respondents, 49 were females (43.4%) and 64 were males (56.6%). While 57 learners' major was science (50.4%), 56 learners' major was social science (49.6%).

2.2 Instruments

To survey the adult students' epistemic curiosity and their experiences and beliefs of online academic help seeking, there were three instruments adopted in this study and presented as a 5-point Likert scale (from 1, "strongly disagree," to 5, "strongly agree"). The *Epistemic Curiosity Scale* developed by Litman (2008) was adopted for examining the students' traits of epistemic curiosity. The students' experiences of help seeking behaviors for academic information were surveyed by the *OAHS (online academic help seeking) questionnaire* (Cheng & Tsai, 2011). To investigate the students' beliefs about online academic help seeking, the *online help-seeking questionnaire (OHSQ)* developed by Liu (2017) was adopted and further adapted to fit the research context of this study. Notably, the Cronbach's alphas for all of the scales were higher than 0.7, indicating a satisfactory reliability of the instruments.

3. Results and discussion

3.1 Experiences and beliefs of online academic help seeking

Based on the collected data by the aforementioned instruments, this study explored the adult students' OAHS behaviors through the two main dimensions: experience and belief. Regarding the students' experiences of OAHS, compared with the experiences of querying formally or informally online, they tended to have more experiences in searching possible solutions via online channels such as search engines, instructional websites, or other databases when they encountered academic challenges ($M=4.46$, $SD=0.63$) ($F=83.38$, $p<.001$). Moreover, these students also showed more inclination to ask their peers or unknown experts for help ($M=3.69$, $SD=0.91$) rather than to seek help from their teachers ($M=3.38$, $SD=1.06$) through the Internet when having academic problems. The findings of this study were partially consistent with the previous studies (Cheng & Tsai, 2011). That is, when the students in higher education level encountered academic issues, searching information online may be their favorite approaches to resolving problems no matter at the previous or present stages. Interestingly, while the past students did not show strong different intention to seek help formally or informally (Cheng & Tsai, 2011), the current students in this study were inclined to ask for help informally via online channels and they might have fair experiences in querying formally.

With regard to the students' beliefs about OAHS, compared with the considerations of perceived benefits ($M=4.15$, $SD=0.76$) and self-efficacy ($M=4.07$, $SD=0.82$) for seeking help online, they possessed strong epistemic beliefs that OAHS behaviors could provide them with vital and useful information for academic problem-solving ($M=4.25$, $SD=0.76$) to a significant level ($F=7.3$, $p<.01$). This study further conducted a series of stepwise regression analysis to examine how the students' beliefs of OAHS predict their experiences of OAHS. The results showed that the students' perceived benefits regarding OAHS may significantly predict their experiences of information search ($\beta=0.43$, $p<.001$) and formal query ($\beta=0.25$, $p<.01$) behaviors. Notably, the students' experiences of informal query were significantly predicted by their perceptions of self-efficacy for OAHS ($\beta=0.35$, $p<.001$). The findings highlight the role of self-efficacy of OAHS in the adult students' behaviors of informal

query. The beliefs about perceived benefits from OAHS might foster the students' behaviors of information searching and formal query when encountering academic problems.

3.2 Relationships between epistemic curiosity and experiences and beliefs of OAHS

According to descriptive data of the adult students' scores on the Epistemic Curiosity Scale, their exhibition of I-type curiosity (M=4.05, SD=0.71) to explore something new and D-type curiosity (M=3.78, SD=0.77) to seek knowledge for reducing perceived information gaps were slightly above average. This study further examined how the students' epistemic curiosity were related to their experiences and beliefs of OAHS. As shown in Table 1, there were strong significant relationships between the students' I-type epistemic curiosity and their experiences of information searching ($r=0.46$, $p<.001$). On the other hand, the students' D-type epistemic curiosity was strongly related to their experiences of formal query ($r=0.34$, $p<.001$) and informal query ($r=0.47$, $p<.001$). Litman et al. (2005) have addressed that D-type curiosity is a stronger motive for seeking knowledge when comparing with I-type curiosity. Taking this a step further, this study identified that, when encountering academic challenges, adult students with stronger I-type curiosity may tend to search information online for help. Moreover, for students in higher education, stronger D-type curiosity may lead to help seeking formally or informally for academic problems.

It was also found that there were significant relationships between the students' epistemic curiosity and their beliefs of OAHS. Table 1 shows that the strength of the correlations between I-type curiosity and beliefs of OAHS was higher than that between D-type curiosity and beliefs of OAHS. Specifically, the students with stronger I-type curiosity may be likely to possess stronger epistemic beliefs ($r=0.38$, $p<.001$), perceive more benefits from OAHS ($r=0.45$, $p<.001$), and exhibit higher level of self-efficacy for OAHS ($r=0.46$, $p<.001$) as compared to D-type curiosity. Since Litman (2008) reported that learners with strong I-type curiosity may tend to hold mastery-oriented learning goals (e.g., considering that success is the results of efforts), this study considered that the students with strong I-type curiosity were inclined to possess positive beliefs about seeking help online for dealing with their academic issues.

Table 1

Relationships between epistemic curiosity and experiences and beliefs of OAHS

	Information searching	Formal query	Informal query	Epistemic beliefs	Perceived benefits	Self-efficacy
I-type EC	0.46***	0.28**	0.30**	0.38***	0.45***	0.46***
D-type EC	0.34***	0.34***	0.47***	0.27**	0.35***	0.44***

EC: epistemic curiosity, ** $p<.01$, *** $p<.001$

References

- Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology*, 45, 180-191.
- Cheng, K. H., & Tsai, C. C. (2011). An investigation of Taiwan university students' perceptions of online academic help seeking, and their web-based learning self-efficacy. *Internet and Higher Education*, 14(3), 150-157.
- Litman, J. A., Hutchins, T. L., & Russon, R. K. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion*, 19, 559-582.
- Litman, J. A. (2008). Interest and deprivation dimensions of epistemic curiosity. *Personality and Individual Differences*, 44, 1585-1595.
- Litman, J. A. (2010). Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences*, 48(4), 397-402.
- Liu, S. H. (2017). Relationship between the factors influencing online help-seeking and self-regulated learning among Taiwanese preservice teachers. *Computers in Human Behavior*, 72, 38-45.

The Role of Technology Identity among Students in Rural Areas using a Web-based Tutoring System

May Marie P. TALANDRON-FELIPE
Ateneo De Manila University, Philippines
Central Mindanao University, Philippines
maymarie.talandron@upou.edu.ph

Abstract: This paper looked at the relationship between the student's technology identity and how it affects their performance in a web-based tutoring system over a 12-week period. Technology identity covers one's beliefs on four areas: opportunities and constraints to use technology (access), technology skills (abilities), role of technology on one's life (importance), and one's own motivation to learn more about technology (opportunities). Results showed that access factors such as frequency of access, owning a computer at home, and owning a personal computer/laptop do have an impact on their ability to use computer applications. A significant difference was found in the number of problems completed (higher) among students who were able to frequently access and use a computer and the Internet for free in school. This could be an indication of the importance of the role of educational institutions to provide computer and Internet access to students who can't afford to have their own at home or rent. Also, 99% of the students, including those who did not solve as many problems, viewed technology as an essential part in education, business, personal development, and their future profession. All (100%) the participants expressed willingness to learn more about technology. However, when it comes to the fourth are (opportunities), almost half (47%) believe that they don't have sufficient opportunities and access to technology to be used in education.

Keywords: Technology Identity, ASSISTments

1. Introduction

The use of Information and Communications Technology (ICT) in education has become an important pedagogical technique in the 21st century (Sarkar, 2012) where students have become more engaged in the meaningful use of computers (Sanchez & Aleman, 2011). It paved way on the enhancement of the quality of education with advanced pedagogical methods, improvement of learning outcomes and reformation for better management of education systems (Sarkar, 2012). Because of these, the adoption of different learning environments has become global and ICT access of developing countries has increased (Nye, 2015). As stressed in Kam et al. (Kam, Mathur, Kumar, & Canny, 2009), educational software can make a positive impact on the learning needs of underserved communities. However, much of the existing educational software are developed in the West (Blanchard, 2012; Ogan et al., 2012) and software, technology tools and other applications that are accustomed in developed countries may cause some challenges in the developing world, so it is not safe to assume that one size can fit all or that a single technology design may be effective for all context and setting (Keengwe & Bhargava, 2014). Since further research is needed in identifying the impact of the technology background of learners in the adoption of educational software (Khan, Hossain, Hasan, & Clement, 2012; Nye, 2015; Rodrigo, Sugay, Agapito, & Reyes, 2014), this study investigates the learning experience of students from rural areas in a developing country. Specifically, this study aims to answer the question: what is the relationship between a student's technology identity and experience in a tutoring system?

2. Technology Identity

Based on the theoretical construction of student's identity of Martin (Martin, 2000), Goode (Goode, 2010) described a student's technology identity as shaped by their beliefs about one's ability to use

technology, the essence of typical computer uses in the context of opportunities, the level of importance attributed to technology and motivation to learn more about computing. Moreover, she argued that exploring the spectrum of beliefs within each of these four categories provides a more nuanced description of individual's identities around technology. Although in today's trend, having technology knowledge is implicitly required for college success and career pathways (Goode, 2010), it is not safe to assume that all students were given equal opportunities to develop their computer skills. In the study of Hargittai (Hargittai, 2010), it was shown that socioeconomic status is an important predictor of how people are incorporating the use of technology in their lives with those from more privileged backgrounds using it in more informed ways for a larger number of activities. For many college students, not having a strong technology identity is a product of an unequal high school education and disparities in home resources, yet the consequences of one's technology identity has a powerful influence on the attitudes and decisions students make regarding their academic and life plans. Examining the technology identity of individuals informs our consideration of how beliefs about oneself and technology are developed, shape daily social interactions and influence future life plans (Goode, 2010).

3. Experiment

Eighty-eight (88) freshmen students from a state university in southern Philippines were asked to use ASSISTments and complete 12 problem sets in 12 weeks. The 12 problem sets were selected based on the lessons designed for College Algebra in the Philippines. The students were divided into 2 groups: Sequential Group (SG) which were assigned 12 problem sets in sequential order (weekly released) based on how they are delivered in the classroom and the Non-Sequential Group (NSG) where all the 12 problem sets were made available to them where they can decide the order of solving the problems. The students were asked to answer a pre-test prior to using ASSISTments and a post-test after the 12-week period. ASSISTments provided a proficiency report for each student.

The participants were asked to answer a questionnaire before they used ASSISTments. This self-assessment method was based on the conceptual framework discussed in Goode (Goode, 2010). The questionnaire was divided into 4 areas: fluency, experience, importance of technology, and feelings toward technology. These areas are based on the theoretical construction of student's identity (Martin, 2000). Fluency is based on their beliefs about one's ability to use technology; experience as the typical computer use in the context of opportunities; importance based on how they see the use of computers to improve lives; and whether they are interested or motivated to learn more about technology (Goode, 2010). A Likert scale was used for the responses in each area and the participants were then classified based on their responses. The results were taken into consideration in examining the relationship of one's technology identity and performance in ASSISTments.

4. Findings

First, access to technology and how it affects the students' computer literacy in terms of using basic applications (e.g. word processor, spreadsheet) and troubleshooting skills was investigated. Factors such as length of experience in using a computer, frequency of access, owning a computer at home, and owning a personal computer/laptop do have an impact on their ability to use computer applications and computer troubleshooting skills. Next, the place of student's computer access and number of problems completed was looked into. Sixty of the 87 participants (70%) have no computers at home (family owned) and 50 (57%) have no personal computer or laptop. However, ownership of a computer system has no significant relationship in terms of the number of problems completed in ASSISTments ($M=23$, $SD=20$). To further understand this, a comparison was made in terms of where they actually frequently access or use the computer and the Internet, whether at home, in school, or in computer shops. The p-value corresponding to the F-statistic of one-way ANOVA is lower than 0.01 which strongly suggests that one or more pairs of treatments are significantly different. To pinpoint which of them exhibits statistically significant difference, Tukey's HSD test was applied to each of the 3 pairs. There is a significant difference in the number of problems completed among the students who answered "School" as the place where they frequently use the computer and the Internet. This could imply the importance

of the role of the educational institution to provide computer and Internet access to students so that regardless if they have their own computer or not, they can still take advantage of computer-based learning environments.

The second area on the TID framework is the beliefs about one's technology skills. The relationship of the students' computer literacy and length of experience with computers and the number of problems solved in ASSISTments were analyzed and found no significant relationship for both, $r=0.01$, $p=0.92$ and $r=0.06$, $p=0.60$, respectively. This indicates that the students' background in using the computer did not hinder them in using ASSISTments.

In the third area, 56 students (64%) believed that technology is "5 – very important", 30 students (35%) gave a rating of "4 – important". No significant relationship was found between these ratings and the number of problems solved in ASSISTments. A combined 99% of "very important" and "important" means that even those who did not solve as many problems in ASSISTments still believe that technology plays an important role in education, business, personal development, and their future profession.

The last area is the belief about one's own motivation to learn more about technology. All of the participants (100%) expressed willingness to learn more about technology and its many uses but almost half (47%) of the students believe that they don't have sufficient access to technology to be used in education. This belief, though, has no relationship with their performance in ASSISTments ($rpb=0.15$, $p=0.16$).

Acknowledgements

The author thanks: the ASSISTments team for the platform in this pilot study; the faculty of the Institute of Computer Applications who assisted during the deployment and data gathering; and Dr. Didith Rodrigo of the Ateneo Laboratory for the Learning Sciences.

References

- Blanchard, E. G. (2012). On the weird nature of ITS/AIED conferences. *International Conference on Intelligent Tutoring Systems*, 280–285. Springer Berlin Heidelberg.
- Goode, J. (2010). The digital identity divide: how technology knowledge impacts college students. *New Media & Society*, 12(3), 497–513.
- Hargittai, E. (2010). Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the "Net Generation." *Sociological Inquiry*, 80(1), 92–113.
- Kam, M., Mathur, A., Kumar, A., & Canny, J. (2009). Designing digital games for rural children: a study of traditional village games in India. *SIGCHI Conference on Human Factors in Computing Systems*, 31–40. ACM.
- Keengwe, J., & Bhargava, M. (2014). Mobile learning and integration of mobile technologies in education. *Education and Information Technologies*, 19(4), 737–746.
- Khan, M., Hossain, S., Hasan, M., & Clement, C. K. (2012). Barriers to the Introduction of ICT into Education in Developing Countries: The Example of Bangladesh. *International Journal of Instruction*, 5(2), 61–80.
- Martin, D. (2000). *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency*. Routledge.
- Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 25(2), 177–203.
- Ogan, A., Walker, E., Baker, R. S., Rebolledo Mendez, G., Jimenez Castro, M., Laurentino, T., & De Carvalho, A. (2012). Collaboration in cognitive tutor use in Latin America: Field study and design recommendations. *SIGCHI Conference on Human Factors in Computing Systems*, 1381–1390. ACM.
- Rodrigo, M. M., Sugay, J. O., Agapito, J., & Reyes, S. J. (2014). Challenges to transferring western field research materials and methods to a developing world context. *Research and Practice in Technology Enhanced Learning*, 9(1), 71–82.
- Sanchez, J. J. C., & Aleman, E. C. (2011). Teachers' opinion survey on the use of ICT tools to support attendance-based teaching. *Computers & Education*, 56(3), 911–915.
- Sarkar, S. (2012). The role of information and communication technology (ICT) in higher education for the 21st century. *Science*, 1(1), 30–41.

Proposing Curriculum and Learning Environment Development for Global Liberal Arts Education Incorporating Future Work Skills

Tosh YAMAMOTO^{a*}, Juling SHIH^b, Chris PANG^c, Benson ONG^c

^a*The Center for Teaching and Learning, Kansai University, Japan*

^b*Department of Information and Learning Technology, National University of Tainan, Taiwan*

^c*School of Business Management, Nanyang Polytechnic, Singapore*

*ctltoosh@kansai-u.ac.jp

Abstract: This paper proposes a “de facto” standard curriculum for Tesseractiv© and Global Liberal Arts Education harnessed with ICT-enhanced learning environment targeting at global students’ audience, i.e., future generation that will be fully ready for Singularity in 2045. The purpose of the research is to develop a robust curriculum and associated contents to foster the future work skills for all stakeholders in the global society involving university students, adults with various ages in different social strata. In other words, in such curriculum, all stakeholders will learn together, where the younger generation can have dialogs with different generations as if they were discussing with themselves in the future, and the older generations can have dialogs as if they were discussing with themselves in the past when they were young. In this way, transcendence of cultural values, heritage, wisdom, and experiences in the human civilization will be guaranteed for the benefit of the future global society. Such a multi-facet curriculum involving all stakeholders will be designed across the border of the campus along with the society, as well as beyond the borders of nations. In addition to the curriculum and its related contents, this research proposal also includes the development of the learning environment for its learning environment, applying Jupyter notebook, Jupyter Hub, and Git Hub technologies at the level of simulation learning while guaranteeing Academic Integrity issues. The bird’s eye view of the entire research will be shared in the form of a poster.

Keywords: Global Liberal Arts Education, Curriculum Development, ICT-enhanced AGILE learning environment, transcendence in learning

1. Introduction

As in the ABET Model of education, the four years of the university curriculum bears the responsibility of raising the future generation that will strive for their life-long dream and contribute to the global society. Unfortunately, the current curriculum in the educational paradigm today is based on the Prussian (German) Educational System developed in 1806 (<https://feltd.wordpress.com/2010/09/16/the-prussian-german-educational-system/>). As a matter of fact, a Princeton University professor, in his book titled, “The Case Against Education”, Bryan Caplan, claims that the current education System is a waste of time and money. Such a curriculum or educational paradigm will only give rise to unemployment when graduates face the Singularity in 2045. Thus, a paradigm shift in education is a must. Changes toward the Singularity are already shifting outside the world of the university curriculum and the corporate-level human resources training workshop.

It should be emphasized that this project demonstrates an execution of active and Tesseractiv© learning in the form of Project/Problem-Based Learning in Global AGILE Teams.

2. Rationale behind the Paradigm Shift in Curriculum

Singularity in 2045 will bring the situation where almost a half of the jobs now will be replaced by intelligent robots. Oxford University predicts that 47% of the current jobs will disappear at the time of Singularity in 2045, which means that almost a half of freshmen today will lose their jobs when they are nearly fifty years of age. The Institute for the Future predicts and defines the future work skills that may be necessary in the future. The key concepts for the global education must be identified in order to design the mission as well as its associated curriculum.

It might be helpful to look at the key concepts proposed by the Institute for the Future, which focuses on the skills in workplace. The key concepts are: Sense Making, Social Intelligence, Novel Adaptive Thinking, Cross-Cultural Competencies, Computational Thinking, New Media Literacy, Transdisciplinarity, Design Mindset Visualized Communication), Cognitive Load Management, and Virtual Collaboration. It must be emphasized that none of the key concepts are currently incorporated in the education, still bound by the traditional and legacy curriculum. Thus, for the successful launch for the global education, such concepts must be incorporated in the curriculum.

The proposed curriculum must take such philosophical components as redefined stakeholders for education, andragogy, the realm of learning in the Bloom's Taxonomy Matrix, and the learner-centered Tesseractiv^o Education.

3. Research Method

3.1 Overview

This project will develop a robust curriculum for future generations to have the Tesseractiv^o Learning skills or competency to thrive and to show leadership in society. The ten crucial components for the future work skills defined by IFTF as well as the concepts of SDGs will be best incorporated in the curriculum being developed here. The research project consists of four sprints:

- Sprint I: Development for meta curriculum and Design for associated components.
- Sprint II: Design for Assessment Components
- Sprint III: Pilot tests and refinement of the entire system.
- Sprint IV: Dissemination: workshops for promotion of the research: Tesseractiv^o Curriculum

It should be emphasized that the research itself is an optimal showcase demonstrating problem/project-based learning with situational leadership in the social constructive model, which is best described as AGILE Learning Model.

3.2 Progress Report

So far, the training workshops or seminars have been conducted targeting at the same age group or the same skill level employees in series or sporadically at the university level or at the corporate level. The reason behind such trainings has been based on the concept that the outcome or results had been focused in order to sustain the status quo of the values that had been built while claiming the importance of the process of learning through trainings. For the system design and development, the progress has been published in the Kansai University's academic journal published in March 2019. In the article titled "A Proposal for ICT-Enhanced Learning Environment Fortified with BYOD Choreographies: Designing an Online Active Learning System to Foster the 21st Century Skills", the following system design was presented.

3.3 Case Study – Assessment of Learning Outcomes

The global collaboration between Kansai University (Japan) and Nanyang Polytechnic (Singapore) focused on assessing Tesseractiv^o Learning skills and competencies. It explored social entrepreneurship and UN SDGs with 7 teams of 74 participants over 8 weeks in 2019. Each international team comprised a Japanese and Singapore team. Teams use online collaboration tools (Skype, Flipgrid, Padlet, Google Apps) to produce SDG-targeted innovations that resulted in green services, eco-products and sustainability apps using business model canvases. The teams completed a pre- and post-test of entrepreneurial skills using FINCODA, which measured self-perceived innovation competence based on creativity, critical thinking, initiative, teamwork and networking. Content

analysis of pre and post teams' business model canvases (LEANSTACK) also coded for value proposition, value ideation, insights, foresight and readiness.

Despite language and cultural barriers, initial results are promising. Teams reported increases in innovation competence for creativity (8.7%), critical thinking (5.4%), initiative (4.1%), teamwork (2.3%) and networking (3.7%). Content analysis of business model canvases captured innovative competence increases for value proposition (10.6%), value ideation (10.9%), insights (23.3%), foresight (15.9%) and readiness (15.2%).

Acknowledgements

Special thanks must be mentioned to Dr. Meilun Shih, who have been actively participated in the collaborative learning between Kansai University and National Taiwan University; Dr. Akon Higuchi, of National Central University for pilot seminars for the curriculum, and Dr. Anthony Y. H. Liao, Director of AI Center at Asia University. Thanks is also given to the senior management from Nanyang Polytechnic for contributing to assessment design.

References

- Glazer, H. R., & Wanstreet, C. E. (2011). Connection to the academic community: Perceptions of students in online education. *Quarterly Review of Distance Education*, 12(1), 55-62.
- Graham, C. R. (2006). Blended learning systems. Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.). *The handbook of blended learning: Global perspectives, local designs*, 3–21. Retrieved January 20, 2019, from <https://books.google.com.mx/books?isbn=1118429575>.
- Horizon 2020. (2018) Funding, Tenders. Retrieved January 20, 2019, from <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>.
- Institute For The Future. (2011) *FUTURE Work Skills 2020*. Retrieved January 20, 2019, from http://www.iftf.org/uploads/media/IFTF_FutureWorkSkillsSummary_01.gif
- Kozma, R. B. (Ed.). (2003). *Technology, innovation and educational change: A global perspective*. Eugene: International Society for Technology in Education International Association for the Evaluation of Educational Achievement.
- Lajoie, S. P., Hmelo-Silver, C. E., Wiseman, J. G., Chan, L. K., Lu, J., Khurana, C., et al. (2014). Using online digital tools and video to support international problem-based learning. *Interdisciplinary Journal of Problem-Based Learning*. <https://doi.org/10.7771/1541-5015.1412>.
- Law, N., Pelgrum, W. J., & Plomp, T. (2008). *Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study (CERC Studies in Comparative Education)*. Hong Kong: Springer, Comparative Education Research Centre.
- Peñalver, M. J. P., Mas, L. E. A., & Fleta, B. M. (2018). Identification and classification of behavioural indicators to assess innovation competence. *Journal of Industrial Engineering and Management*, 11(1), 87-115.
- VISION 2020. (2014) *Vision 2020 – Education*. Retrieved January 20, 2019, from http://www.planningcommission.gov.in/reports/genrep/bkpap2020/14_bg2020.pdf.
- Waddoups, G. & Howell, S. (2002). Bringing online learning to campus: The hybridization of teaching and learning at Brigham Young University. *International Review of Research in Open and Distributed Learning*, 2(2). Retrieved Month day, year, from <http://www.irrodl.org/index.php/irrodl/article/view/52/108>.
- Yamamoto, T., Okunuki, M., & Watanabe, M. (2017). *Academic Writing as Corpus for Assessment of ePortfolio*. Session: Education Innovation with ICT. PNC 2017 Annual Conference and Joint Meetings at National at Cheng Kung University, Tainan, Taiwan.
- Yamamoto, T., Wu, W. V., Liao, A. Y. H. , & Tagami, M. (2017). *Proposing a New Phase of Liberal Arts Course Enhanced with Active Learning for PBL in the Social Constructivism Paradigm: Advanced Communication Based on Empathy & Trust Building*. PNC 2017.
- Yamamoto, T., Shih, M., and Watanabe, M. (2018). *What is the Educators' Mission for the Future Students? - Scenario Planning of Educational Design Targeting at the Singularity Era*. DLSA 2018. Hokkaido U.
- Yamamoto, T., Liao, A. Y. H., Wu, W. V., Shih, M., Shih, J., Chu, H. (2018). *A Proposal for the Global and Collaborative PBL Learning Environment Where All Global Members on Different Campuses Are 'On the Same Page' throughout the Process of Learning in the Project*. 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI). Publisher: IEEE. (DOI: 10.1109/TAAI.2018.00029).

Patterns of Student Learning Behaviour within Technology-supported Socialised Learning Contexts

HanYuNing LIN^{a*}, Mathew R PRYOR^b

^a*The University of Hong Kong, Hong Kong*

^b*The University of Hong Kong, Hong Kong*

*lynnlhyn@hku.hk

Abstract: With rapidly growing interest in the use of Information and Communication Technology (ICT) tools, student learning through interaction and collaboration becomes feasible and flexible in both online and offline learning environments. Cognitive development in high-density socialised contexts has been a distinct focus in educational research. In this study, we transformed a traditional lecture course at HKU into a communication-intense classroom with the aim of maximizing meaningful socialisation amongst learners. Semi-structured interviews with undergraduate students were conducted after the course and inductive content analysis was undertaken to identify patterns of student learning behaviours which influence the quality of their knowledge construction. Research findings indicated emerging patterns of group socialised learning behaviours were actively-involved, small-circle and passively-reacting; two types of peer pressure - positive and negative - had strong effect on an individual's learning behaviour within their group.

Keywords: Learning behavior pattern, socialised learning, technology support, peer pressure

1. Research Background and Research Questions

Researchers suggested that information technology affords innovation and diversity in learning and teaching compared to traditional classrooms (Lage, Platt, & Treglia, 2000). ICT tools give new opportunities for students to learn flexibly and interact closely; and for instructors to deliver courses more creatively in both online and offline environments. Prior research has confirmed students involve more intensely and distribute group work reasonably within technology-supported learning environments (Angeli, Valanides, & Bonk, 2003). Several studies have detected significant enhancement of students learning achievement, satisfaction and knowledge construction via online collaboration (Young, 2008; Zhu, 2012; Ku, Tseng, & Akarasriworn, 2013). However, further examination of how students behave within highly socialized groups is still needed.

In this study, we conducted an experiment to transform a traditional lecture-based common core course at HKU into a socialised-learning-design format based on Vygotsky's social constructivism theory (1978), in addition to Järvelä and Hadwin's Regulation of Collaboration Theory (2013). Within this socialized learning context we explored: 1) what patterns of student learning behaviours emerged; 2) what factor(s) influenced student learning behaviours; 3) how the factor(s) influenced such behaviours.

The course was conducted using online pre-class exercises and face-to-face in-class practices. Students were required to watch videos of the instructor teaching basic concepts on Open edX and perform group tasks to reinforce the knowledge they learnt from videos. As Lee (2000) stated, many students fail in online learning due to their loss of motivation or poor adaptation to new way of learning. Typically students work independently online, which requires them to have great control of their own study. We addressed this by assigning students to work on online pre-class tasks collaborative in groups, which encouraged them to regulate each other's learning by sharing the learning responsibility. Student groups were pre-chosen and with the intention of maximising background diversity. In class, the same groups sat together at tables and were undertook various interactive activities integrated with ICT tools (e.g. Mentimeter and Flipgrid) to share their ideas and apply the knowledge acquired pre-class to real-life situations. With Mentimeter, students shared their opinions simultaneously on a

projector screen in class and were thereby able to visualise others perspectives. On Flipgrid, students shared their group video projects, gain inspiration from other students' works and communicated with each other.

2. Methodology and Data Analysis

2.1 Participants and Semi-structured Interviews

14 out of 116 students voluntarily participated in post course semi-structured interviews. Guide questions were designed by colleagues with professional teaching and research backgrounds to avoid bias. To encourage students to reflect on group interaction, participants from the same course group were distributed into different interviews. To maximise the depth of conversation, each interview only included 1 to 2 participants and lasted for about 50 minutes. Interviews were audio recorded with participants consent.

2.2 Inductive Content Analysis

Content analysis is defined as a research method to make rational and sustainable inferences from data to their context, so as to offer understanding, new insights, and demonstration of facts (Krippendorff, 1980). Differentiating from deductive analysis, inductive content analysis generates categories from data rather than matching data with pre-set categories. It follows the procedure of open coding, creating categories and abstraction (Elo & Kyngäs, 2008). To avoid bias and pre-assumption of interview contents, cross-validations and discussions were carried out by the authors till agreement on the analysis results was reached.

3. Research Findings

3.1 Patterns of Group Socialised Learning Behaviours

Three patterns of students socialized learning behaviour in groups were identified from interviews. Actively-involved pattern refers to every group member making their own effort in group discussion, plan setting to achieve learning goals, checking project progress and evaluating work quality. Students who were less active at first also got progressively more involved. Small-circle pattern occurs where several students in a group form their own secure inner circle and interact only within that circle with other students in the group being left outside and not involved. Passively-reacting pattern happens where group members play solo roles within projects and have minor interaction with each other.

Table 1

An Example of Students Feedback on Patterns of Group Socialised Learning Behaviours

Actively-involved	“Every person in our group is willing to do things and we appreciate each other’s work quite a lot. We voice out our thoughts and do concrete works to show each other.”
Small-circle	“Sometimes we ended up submitting things that they just did within themselves and a point to me I wouldn’t appreciate the quality overall. I feel if the same thing we would do in a more collaborative way we would be able to submit a better quality of work.”
Passively-reacting	“Our group was not active in group activities and we didn’t have much communication. We just did our own work and submitted before deadline.”

3.2 Types of Peer Pressure

Two types of peer pressure were identified from data analysis. Positive peer pressure allowed students to feel self-driven and motivated under internal or external influences so that they made more contributions to their group project, provided assistance to group members in need, sought to improve the quality of group work, etc. Negative peer pressure diminished students self-efficacy and positive view of their capacity to contribute to group work due to certain worries or nervousness.

Table 2

An Example of Students Feedbacks on Typology of Peer Pressure

Positive pressure	“My group mates’ desire of improving the work drives me to engage more in the assignments and makes me feel it’s my responsibility to do so.”
Negative pressure	“Because I’m the type of person that I would like to finish an assignment as soon as I can, but my group mates do it in the last minute. So you have to learn how to compromise or at least communicate with other people without sounding annoying, which is very stressful.”

4. Discussion and Conclusion

Given the direct evidence from students’ feedback and data analysis, students who actively performed in class activities and interacted with peers tended to report having more positive pressure, self-driven contributions, engaging learning experiences, appreciative attitude and notable learning achievements. Students who were far less engaged seemed to suffer from negative pressure and impassive attitude but still considered that they obtained the required academic knowledge and skills.

In this study, to examine students learning behaviours in a socialised learning design context, we firstly transferred a traditional lecture course into a communication-intense classroom entailing collaborative online pre-class and offline in-class exercises. Then we undertook semi-structured interviews with students from the course and applied inductive content analysis to draw patterns of students socialized learning behaviours and the factor that has effect on how students behave in groups. From this we are able to give other instructors clear and multi-faceted understanding of how students behave to construct knowledge within highly socialized educational settings.

Acknowledgements

This research study was funded by HKU Teaching Development Grant (Project 18/685).

References

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing*, 62(1), 107-115.

Krippendorff K. (1980) Content Analysis: An Introduction to its Methodology. *Sage Publications*, Newbury Park.

Ku, H. Y., Tseng, H. W., & Akarasriworn, C. (2013). Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in human Behavior*, 29(3), 922-929.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, 48(1), 25-39.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, 31(1), 30-43.

Lee, C. Y. (2000). Student motivation in the online learning environment. *Journal of Educational Media & Library Sciences*, 37(4), 367-375.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Young, A. (2008). Structuring asynchronous discussions to incorporate learning principles in an online class: One professor’s course analysis. *MERLOT Journal of Online Learning and Teaching*, 4(2), 218-224.

Zhu, C. (2012). Student satisfaction, performance, and knowledge construction in online collaborative learning. *Journal of Educational Technology & Society*, 15(1), 127-136.

Investigating the effects of cognitive style on blended museum learning

Tien-Yu Hsu, Hsin-Yi Liang*, Jun-Ming Chen

Department of Operation, Visitor Service, Collection and Information Management, National Museum of Natural Science, Taichung, Taiwan

Abstract: Museums provide an ideal inquiry learning environment with fruitful learning resources for diverse learners. With digital applications in museums, there is a need to consider the influences of the museum context in the service design to avoid negative effects on museum learning. Besides, there is a need to provide proper learning supports to satisfy the different needs of diverse learners. In order to effectively support diverse learners learning, cognitive style has been the focus due to the significant effects on learning with digital applications. However, as there are limited studies on understanding the effects of cognitive style on museum learning, this study developed a game-based learning service to engage learners in interacting with museum exhibits and other people in a blended museum learning context. The learners' learning experiences were examined to evaluate the service design, and an independent *t*-test was used to identify the differences between learners with different cognitive styles, i.e., the Field dependent (FD) and Field independent (FI). The results show that FI and FD learners have different preferences concerning service design during their visits. Besides, the unmatched service design may influence FI learners better than it would FD learners.

Keywords: Museum learning, cognitive style, field dependence–independence, user-interface design, mental workload

1. Introduction

Museums provide an ideal inquiry learning environment with fruitful learning resources for diverse learners. In this study, we developed a game-based learning service named CoboFun to promote learners inquiring in a blended museum learning context. The various types of problem-solving tasks with fruitful learning resources are provided to engage learners interacting with museum exhibits and other people with digital applications and museum's physical environments. A learning activity was designed for the learners aged ten and above. A flexible game-based learning environment with different learning support design, such as hints and exploration map, were provided to support learners with different cognitive style, i.e., Field independent(FI) and dependent(FD) learners learning in museums. The learning experiences and mental workload were evaluated to examine the design of CoboFun. The results of the study can be used to improve the design of CoboFun and to address issues that do not meet the specifications.

2. Service and Learning Activities Design

A game-based learning service named CoboFun was developed to promote learners' interactive museum experiences during their visits. A learning quest named "Doctor D's Dinosaur Park" was designed for introducing the exhibitions of Life Science Hall on display of National Museum of Nature Science (NMNS) in Taiwan. A total of eight problem-solving tasks were integrated in the quest for constructing the thematic knowledge of the origin, evolution and extinction of life on earth. In the beginning, a simulated story is provided to motivate learners' participation and guide their actions to solve puzzles. The learner is able to use beacon technologies to identify their position and plan their route to the various locations with an exploration map. Each learning task provides a puzzle to engage learners interact with the nearby exhibits and other people. For example, Figure 1 and 2 display an example of task design. The learner is able to find the location of target exhibit and the answers from the descriptions and audio tour. The clues and hints are provided to support the learner identify the keys to solve the puzzle. The learner is able to select their preferred tasks, stop and continue gaming at any time

they want. To promote learner solving all the puzzles, the learner is able to collect Virtual Reality objects (Dinosaurs) after they finish each task. Besides, the learning summary provided illustrated explanations to summarize the thematic knowledge of each task. Once they complete the quest, the learner is able to view the VR video with their collected dinosaurs or take Augmented Reality photos as souvenir picture, and they are able to access extended learning materials related to the quests on a MOOCs platform. Figure 3 shows the service flow of CoboFun.



Figure 1. The learners interact with exhibits with Exploration Map

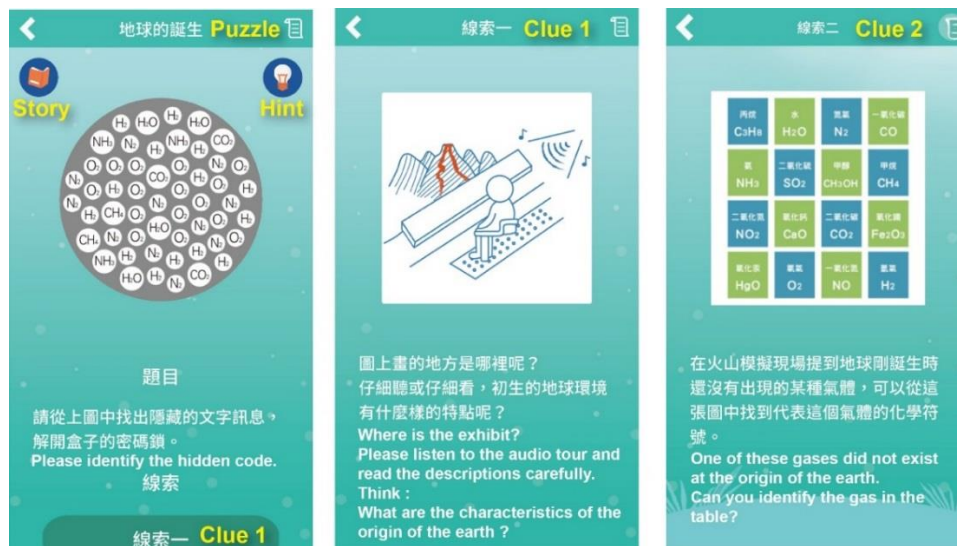


Figure 2. An example of task design of “Doctor D’s Dinosaur Park”

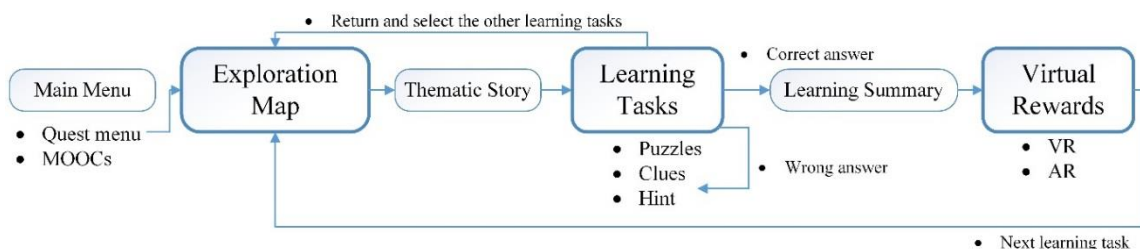


Figure 3. The game-based learning service flow of CoboFun

3. Evaluations

A total of 58 students (28 boys and 30 girls) in the sixth-grade of elementary school participated in this study. Descriptive analysis was applied to examine learners’ museum experiences and mental

workloads. We then used a Levene's test to check for homogeneity of variances among the raw measurements. An independent *t*-test was applied to compare the differences of learning experiences between FI and FD. The significance level was $\alpha = .05$ for each statistical analysis.

The results show both FI and FD learners considered that learning with CoboFun required neither low nor heavy mental workload. The results show the workloads of learning with CoboFun to be in the acceptable range, which hardly caused cognitive overload. Most learners considered that learning with CoboFun required more effort on mental demand than physical demand. Besides, the FD learners paid more effort to physical demand (Mean=4.13, SD=2.37) compared to the FI learners (Mean=3.31, SD=1.91).

Regarding the comparison between learners with different cognitive styles, the results echoed previous studies, in that FI and FD learners have different preferences and service design needs (Chen, 2002; Huang, Hwang & Chen, 2016). The *t*-test results demonstrate that the FD learners' preferences of service design differ from the FI learners. In particular, FI learners show noticeable preferences on icons with text labels ($t(56) = 1.93, p = .059$) and landmarks ($t(55) = 1.91, p = .061$) in interface design compared to FD learners. Although the *t*-test results did not reach a significant level, the results are similar to previous studies, which found that FI learners prefer navigational structure such as index or search to identify specific contents (Alhajri & Ahmed, 2016). In addition to the preferences regarding interface design, there is a significant difference in the preferences for corrective feedback design between FI and FD learners ($t(56) = 2.55, p = .014$). The results show FI learners prefer to use the tools to support their efficient locating of specific information, and they expected to have more elaborative corrective feedback compared to the FD learners. The different requirements of service design indicated the need to provide different learning supports to effectively support diverse learning in museums. Besides, the unmatched service design may influence FI learners' perceptions better than the FD learners because they prefer to work alone, without help from others. Although the FI learners were less satisfied with the service design of CoboFun, they judged their performance to be better than the FD learners did. The differences in self-judgments on learning performances may reflect that the FD learners considered that they performed worse compared to those without social supports.

4. Conclusions and Future Works

In this study, a game-based learning service named CoboFun was developed to engage the diverse learners learning in a blended museum context. A flexible learning environment with various types of learning supports and learning activities was provided to encourage diverse learners interacting with museum exhibits and the other learners. Besides, the learners' perceptions were examined to understand the different needs of the learners with different cognitive styles. The results of this study provides practical evidence to understand the learning differences between FI and FD learners. However, there are some limitations. Firstly, this study only evaluated learning perceptions. To better understand how and why diverse learners have different learning perceptions, there is a need for future studies to provide learning analysis of learning behaviors. Besides, the learning achievements can be evaluated as well. In addition to the effect of cognitive style, other human factors, such as prior knowledge and age, can be examined to provide better services and match diverse learners' needs in museum learning.

Acknowledgements

This work was supported by the National Science Council of Taiwan, R.O.C., under contract numbers MOST 108-2511-H-178-001.

References

- Alhajri, R., & Ahmed, A.-H. (2016). Integrating learning style in the design of educational interfaces. *Advances in Computer Science: an International Journal*, 5(1), 124-131.
- Chen, S. (2002). A cognitive model for non-linear learning in hypermedia programmes. *British Journal of Educational Technology*, 33(4), 449-460.
- Huang, Y.-M., Hwang, J.-P., & Chen, S. Y. (2016). Matching/mismatching in web-based learning: a perspective based on cognitive styles and physiological factors. *Interactive Learning Environments*, 24(6), 1198-1214.

Development of an Interactive Learning Module for Visualizing Self-Regulated Learning Skills

Harry B. SANTOSO^{a*} & Luthfi K. PUTRA^b

^a*Faculty of Computer Science, Universitas Indonesia, Indonesia*

^b*Faculty of Computer Science, Universitas Indonesia, Indonesia*

*harrybs@cs.ui.ac.id

Abstract: Interactive Learning Module (ILM) is a type of learning module where you can conduct a self-learning activity with certain study materials. The use of ILM can facilitate learners to demonstrate Self-Regulated Learning (SRL) skills. However, most ILMs can be used interactively but cannot record and visualize SRL aspects in the learning process. The purpose of this study is to develop an ILM prototype that can record learner's behavior from SRL perspective. This study also attempts to define qualitative requirements for an ILM that is able to record SRL aspects. The sample material used in this ILM prototype is Boolean logic that is one of the basic fundamentals in Computational Thinking.

Keywords: Interactive Learning Module, Self-Regulated Learning, Computational Thinking, Boolean Logic

1. Introduction

Recent studies have developed an Interactive Learning Module (ILM) using VR or AR technologies (Joo-Nagata, Abad, Giner, & García-Peñalvo, 2017; Papatthomas & Goldschmidt, 2017; Sampaio & Almeida, 2016). For example, a study was conducted to create a VR device where users could learn history by animating historical sites into VR (Joo-Nagata et al., 2017). Another example is a study by Sampaio and Almeida (2016) which was carried out to find a proper pedagogical strategy for ICT class that uses AR devices. From this point we could assume that AR and VR related research in education will grow in quantity in the near future.

Despite all of that research, there is no sign that there will be a piece of research related to the development of an ILM that could record its users' behavior and help them to formulate meaningful feedbacks, specifically from Self-Regulated Learning (SRL) perspective. SRL is a multidimensional construct that involves complex interactions such as cognitive strategy, motivation, and metacognition (Kauffman, 2004). Other studies also sought to define SRL as a process that involves construction of understanding related to a particular topic or domain using strategies and goals, as well as regulating and monitoring certain aspects of cognition, behavior, motivation, and behavior manipulation to achieve desired goals (Pintrich, 2000). The objective of the study is to build an ILM prototype that records users' behavior and map them to SRL aspects.

2. Method

This research is divided into five stages. First, researchers searched for relevant literature to study the concept and the techniques for designing and implementing the prototype. Second, researchers attempted to design and implement the system. Design techniques such as blueprints and content mapping were used in this step. This phase was intended to produce working software that would serve as a prototype. Third, the resulting working prototype was tested to representative users. Techniques such as usability testing and contextual interview were used here. An Indonesian version of System Usability Scale (SUS) questionnaire was used as a tool for gathering data (Sharfina & Santoso, 2006). Aside from the ten statements provided in SUS, the questionnaire also features additional open-ended questions to help researchers identify problems and to compare whether ILM materials are as good as school materials or not. Fourth, researchers analyzed participants' data from the SRL records and attempted to formulate conclusion from those data. Feedbacks from usability evaluation were used to

formulate qualitative requirements in this phase. Fifth, researchers presented the processed data and formulated a general conclusion from those data.

3. Implementation

3.1 Conceptual Design

The ILM prototype is divided into three parts: *game*, *learning module*, and *data*. The *game* part consists of eight stages. Each of the stages presents a Boolean Logic related game called *Dropzone*. In this game, players are challenged to drag a bunch of colored boxes into a part of window called *Dropzone*. The player has to put the right box into the *dropzone*. To know which box has to be put inside the *Dropzone*, they need to read the instruction that is written in *Boolean expressions*. The *learning module* part consists of materials that are divided into five parts. Those parts are *introduction module*, *proposition module*, *NOT module*, *AND module* and *OR module*. Each of the three *Boolean operator* modules (AND, OR and NOT) consists of three parts. Those parts are *definition*, *examples*, and *truth table*. The *definition* part explains the formal definition of each *Boolean operators*; the *example* part presents the users with usage examples related to the corresponding *Boolean operators*; and the *truth table* part shows the True-False configuration of each *Boolean logic operator*. The last one is the *data* part. In this part, users' game data are displayed. These data consist of the distribution of each SRL aspect and a *line chart* that displays SRL actions in a sequential manner.

3.2 Technology & Record Representation

The technologies used to implement ILM prototype are ExpressJS as a *web framework*, PostgreSQL that acts as *database driver*, and Heroku service that is used to host the prototype online. *Drag and drop* animation in the prototype is implemented using InteractJS. In the *game* section of the system, all actions performed by a user will be recorded in the database. The database will record each action in tuples that contain action name, timestamps, and the user who did it. Then, every action will be given tags to map them into one of the SRL aspects. There are four SRL aspects defined in this research, which are *planning strategies*, *monitoring strategies*, *cognitive action*, and *regulating strategies*. After being recorded and given tags, all of these actions are visualized in the *data* section. Figure 1 is an example of the line chart representation. Every dot in represents an action performed by the user.

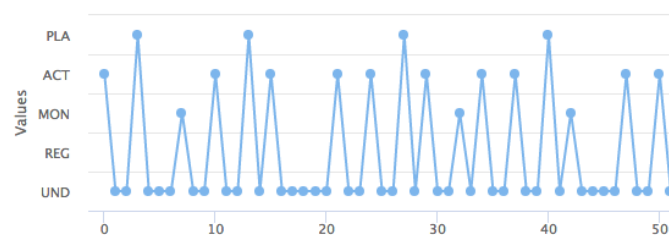


Fig. 1. SRL line chart representation with tags

4. Usability Evaluation

4.1 Participants & Task Scenarios

Thirty two sophomore students from 21 Senior High Schools in Jakarta, Indonesia were participated in the study. They came from a science-based class that uses a science curriculum to teach its students. They were involved in usability testing that was combined with a contextual interview. The tasks given in usability testing consisted of two tasks. The first task was to register into the system until the user successfully logged in into the system. The second task was to complete the tutorial provided by the system. After the usability testing was conducted, participants filled the SUS questionnaire provided.

4.2 SUS Score & Feedback from Contextual Interview and Questionnaire

The SUS score obtained by the system was 74.25. The adjective score obtained by the system was Excellent with grade C. From the first question in contextual interview that reads “What is the most disturbing thing in this application?”, we obtained this collection of problems from the participants: (1) bug from animation; (2) server took a long time to respond; (3) little time to understand a problem; (4) and complicated English. In response to the second question that reads “Which is better, Boolean logic material provided at school or from the ILM application?”, 87.5% of the participants said that ILM material was better, while 12.5% of them said that school material was better. Lastly, there are several recommendations offered by the participants in usability testing: (1) make a mobile version of the application; (2) improve study material and user interface; (3) provide other games to add options; (4) use the Indonesian language; and (5) replace colored boxes with real life objects.

5. Conclusion

From the data gathered and processed in this research, there are several facts that we could draw from this study. *First*, the participants’ plan is not well thought. Plan was not considered important by *usability testing* participants. *Second*, “trial and error” is more considered. This behavior can be seen from participants’ *cognitive action* score that is much bigger than that of other SRL aspects. This also means that participants did a lot of *cognitive action*-based actions compared to those of the other aspects. *Third*, users do not learn from their mistakes. They did not really reflect on their learning actions. Furthermore, from this study we found that many participants favored ILM material over school material. Learning activities provided on ILM are much more interesting than those provided at school. In addition, ILM methods of learning involve encouraging material understanding and immersion to game activity. The “intuitiveness” of ILM game activities creates an impression on the participants that the material provided on ILM is easy. This might be what makes ILM much more interesting than school material.

Acknowledgements

This research was supported by Hibah Publikasi Internasional Terindeks 9 (PIT 9) 2019 at the Universitas Indonesia (NKB-0008/UN2.R3.1/HKP.05.00/2019).

References

- Joo-Nagata, J., Abad, F. M., Giner, J. G., & García-Peñalvo, F. J. (2017). Augmented reality and pedestrian navigation through its implementation in m-learning and e-learning: Evaluation of an educational program in Chile. *Computers & Education, 111*, 1–17. doi:10.1016/j.compedu.2017.04.003.
- Kauffman, D. F. (2004). Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of Educational Computing Research, 30*(1–2), 139–161. doi:10.2190/ax2d-y9vm-v7px-0tad.
- Millard, D. L. (2000). Interactive learning modules for electrical engineering education. *Electronic Components and Technology Conference*, pp. 1042–1047. Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=853298>.
- Papathomas, P., & Goldschmidt, K. (2017). Utilizing virtual reality and immersion video technology as a focused learning tool for children with autism spectrum disorder. *Journal of Pediatric Nursing, 35*, 8–9. doi:10.1016/j.pedn.2017.01.013.
- Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). San Diego, CA: Academic Press.
- Sampaio, D., & Almeida, P. (2016). Pedagogical strategies for the integration of augmented reality in ICT teaching and learning processes. *Procedia Computer Science, 100*, 894–899. doi:10.1016/j.procs.2016.09.240.
- Sharfina, Z., & Santoso, H. B. (2016). An Indonesian adaptation of the System Usability Scale (SUS). Paper presented at the 8th International Conference on Advanced Computer Science and Information Systems (ICACSIS) 2016, Malang, Indonesia.

Investigating STEM-based Learning Package for Enhancing Programming and Problem Solving Skills

Phiraphong PROMJAN^a & Nadh DITCHAROEN^{b*}

^aMajor in Science Education (Computer), Faculty of Science, Ubon Ratchathani University, Thailand

^bDepartment of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Thailand

*nadh.d@ubu.ac.th

Abstract: This research aims to investigate the effectiveness of a STEM-based learning package to enhance the skills of problem-solving and programming. We design and develop the learning package which is consisted of the traffic model, self-directed worksheets, and the Sphero robot. In our design process, the students' knowledge integration and real-life based problem-solving skills were taken into account. Participants are high school students who are aiming major in science in the future. Research instruments and collection include skill assessments, satisfactory questionnaires and observation during the activities. The findings revealed that students had better understanding the program workflow, had improved programming skills, and had the high satisfaction levels from the learning. The average of overall students' satisfaction was 4.38 (SD=0.63).

Keywords: Collaboration, Problem-Solving Skill, Programming Skill, STEM Education, Learning Package

1. Introduction

Programming skills are important in the time of technological advancement. Programming skills are promoted in various organizations around the world. Programming knowledge and skills are required to develop everything such as internet of thing (IOT), machine learning, artificial intelligence (AI) and intelligent systems.

In Thailand, formerly, the high school curriculum was not established to acquire skills of basic programming, computational thinking and problem-solving skills. As a result, students' learning achievement in the computer science study in higher education was low because these students did not have sufficient understanding and knowledge of advance programming (Sochara, 2010). Likewise, other skills required in the computer science field, including creative thinking, problem solving and critical thinking were not focused to be taught in school (Gultekin, 2005). Therefore, students were lacking these important skills which are needed to flourish in life in the 21st century.

Currently, based on the Thai Education Policy for the high schools, the educational focus is to integrate learning methods using the principles of the STEM education. The STEM education consists of four disciplines Science, Technology, Engineering, and Mathematics. This teaching and learning method emphasizes on students' learning projects (Siripattrachai, 2013, Ioannou & Bratitsis, 2017). The STEM method is believed to promote the integrated learning in Thai high schools in order to develop essential computer competencies and encourage computing science literacy among students.

With the principles of the STEM education, teachers still have to choose appropriate instruments to support learning activities. The learning instruments must be appropriate for the characteristics of students which are hands-on, intellectually-stimulating, and entertaining. The Sphero Robots will be a reasonable choice for this circumstance. The Sphero robots are approachable, playful appearance and simple to use, despite containing complex technologies inside (Sphero Inc, 2012).

In order to comprehensively promote the higher order thinking skills and knowledge in computing science, we developed the learning package to provide experiences, improve student behavior, and also facilitate teacher in learning management. In this study, we aim to investigate how STEM-based learning package can enhance programming and problem solving skills by using Sphero robots with specific traffic model and self-directed worksheets.

2. Method

2.1 Participants and Classroom Study

This study involves 30 high school students who are aiming major in science in the future. They were divided into groups of 5 persons. Subsequently, the learning package was provided to these student groups. Figure 1 illustrates the learning flow of this preliminary experiment using our learning package. After the activity was done, students were asked to self-assess their programming and problem solving skill, then to take the questionnaire.

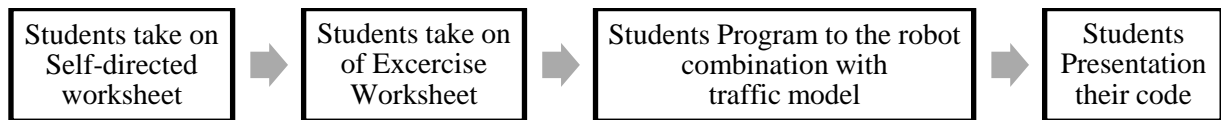


Figure 1. The Learning activity process.

2.2 Research Instrument

The research instrument is the learning package which is consisted of:

- The traffic model, the designed field for robots which is developed based on simulation map from the actual location.
- The self-directed worksheet, the programming practice consisted of 3 sections including concepts of programming, concepts of loop function, and concepts of condition which followed the Thai basic education core curriculum in technology (computing science) subject.
- The exercise worksheet, it provides engineering design process to solve problem real-life simulated situation requiring students to travel from the start point to the end point.
- The Sphero robot used in our experiment was the Sphero SPRK2 robot coding via the Sphero Edu application.

2.3 Measurements

2 types of questionnaires were designed and developed. One was for evaluating students' satisfaction. Another was for evaluating students' skills which included 12 items (6 items for programming skill and 6 items for problem solving skill). They both were developed according to 5-point Likert scale including strongly agree=5, agree=4, don't know=3, disagree=2, and strongly disagree=1 (Likert, 1932).

3. Results and Analysis

3.1 Programming and Problem Solving Skill Assessment

From analysis of students' self-assessment data, the results showed that the average score was 4.26 (SD=0.75). Out of the 12 items, 11 items had average score above 4.00. While, the highest score was the item "the students were able to understand the workflow of the program better" (Mean = 4.53, SD = 0.64), the lowest score was the item "the students were less able to apply knowledge to daily lives" (Mean = 3.87, SD = 0.83).

3.2 Students Satisfaction

The result of students' satisfaction toward the learning package was averagely 4.38 (SD=0.63). Out of the 10 items, 9 items had average score above 4.00. Only one item had average score below 4.00 that was "Proper study time" (Mean = 3.87, SD = 0.92). The highest score was the item "The teaching and learning activity are interesting" (Mean = 4.67, SD = 0.49).

4. Conclusion and Discussion

4.1 How the Activity is Related to STEM Education?

We observed students during doing the activity followed the step of the engineering design process. Figure 2 shows the students' engagement in the activity.

- *Ask*: Each group was able to identify the problems and determine which route would lead them to the destination with the least amount of time.
- *Imagine*: Students brainstormed and shared ideas about the best-possible route.
- *Plan*: Students designed and calculated speed time and distance.
- *Create*: Students programmed the robot following the plan.
- *Experiment*: Students tested the program, observed and adjusted.
- *Improve*: Students modified the code following the adjustments and reattempted the program.



Figure 2. (a) Students brainstorm for the practice worksheet, (b) Students try their code on the traffic model, and (c) Students improve the program.

4.2 How did the Activity Enhance Programming Skill?

Students were able to understand and explain concepts and how to process the program on the self-directed worksheet without teacher assistance. Likewise, students were able to control the robot to complete mission by applying knowledge that gained from the self-directed worksheet.

Acknowledgements

The authors acknowledge the Project for the Promotion of Science and Mathematics Talented Teachers (PSMT), the Institute for the Promotion of Teaching Science and Technology (IPST), Ministry of Education, Thailand. They are also grateful to Faculty of Science, Ubon Ratchathani University for use of facilities.

References

- Sochara, C. (2016). The Development of Instructional Package on The Basic Commands for Controlling a Robot in The Robotics Course for 10th Grade Students. *4th Rajabhat University National and International Research and Academic Conference (RUNIRAC IV)*, 218. [in Thai]
- Gültekin, M. (2005). The Effect of Project Based Learning on Learning Outcomes in the 5th Grade Social Studies Course in Primary Education. *Educational Sciences: Theory & Practice*, 5(2).
- Ioannou, M., & Bratitsis, T. (2017). Teaching the notion of Speed in Kindergarten using the Sphero SPRK robot. In *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference*, 311-312.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55.
- Siripattrachai, P. (2013). STEM Education and 21st Century Skills Development. *Executive Journal*, 33(2), 49-56. [in Thai]
- Sphero Inc. (2012). Sphero Edu TEACHER RESOURCE GUIDE K-12. Retrieved from https://dmmedia.sphero.com/email-marketing/Sphero-Edu/SpheroEdu-k12-teacher-resource-guide-v1_updated050818.pdf

MAIN CONFERENCE C3
ADVANCED LEARNING TECHNOLOGIES (ALT)
LEARNING ANALYTICS AND DIGITAL INFRASTRUCTURE

- FULL PAPER -

CONTINGENCY THEORY OF ADAPTIVE PRACTICES THROUGH THE LENS OF EYE TRACKERS	214
TANG WEE TEO, SENG CHEE TAN	
SIMULATION OF ONLINE LEARNING INTERACTION RELATION NETWORK BASED ON BA MODEL	224
PENG ZHOU, HUAN LI, MEIYI GUO, CHENCHEN SHANG & DI WU	
CLUSTERING MODELS FOR TOPIC ANALYSIS IN GRADUATE DISCUSSION FORUMS	232
GOKARN MALLIKA NITIN, SWAPNA GOTTIPATI, VENKY SHANKARARAMAN	
TEACHING ANALYTICS: A MULTI-LAYER ANALYSIS OF TEACHER NOTICING TO SUPPORT TEACHING PRACTICE...242	
ALWYN VWEN YEN LEE, AIK LING TAN & SENG CHEE TAN	
DO MY STUDENTS UNDERSTAND? AUTOMATED IDENTIFICATION OF DOUBTS FROM INFORMAL REFLECTIONS	252
SIAW LING LO, KAR WAY TAN & ENG LIEH OUH	
TOWARDS SUSTAINABLE LEARNING MATERIALS FOR MOOC IN POOR NETWORK ENVIRONMENTS.....	263
AHMED SHAMS & CHRISTOPH MEINEL	
PROPOSITIONAL-LEVEL ANALYSIS OF COLLABORATIVE LEARNING WITH KIT-BUILD CONCEPT MAP	273
YUSUKE HAYASHI, TOSHIHIRO NOMURA, AND TSUKASA HIRASHIMA	

- SHORT PAPER -

AN EXERCISE RECOMMENDATION METHOD FOR K-12 STUDENTS BASED ON THE SYLLABUS	282
ZHIZHUANG LI, ZHENGZHOU ZHU, TENG YAN	
AN ANALYSIS OF LEARNING BEHAVIOR PATTERNS WITH DIFFERENT DEVICES AND WEIGHTS	288
CHENGJIU YIN, JUAN ZHOU, LINGYU LI, KOUDAI YAMAGUCHI, NORIKO UOSAKI, & HIROAKI OGATA	
RESEARCH ON THE DEVELOPMENT OF A PERSONALIZED LEARNING ASSESSMENT MODEL: BUILDING CONNECTIONS BETWEEN KNOWLEDGE COMPONENTS AND COGNITIVE LEVELS	294
XIAOLING PENG, BIAN WU	
MODELING SELF-PLANNING AND PROMOTING PLANNING SKILLS IN A DATA-RICH CONTEXT	300
HUIYONG LI, YUANYUAN YANG, RWITAJIT MAJUMDAR, GÖKHAN AKÇAPINAR, BRENDAN FLANAGAN & HIROAKI OGATA	
LEARNING SUPPORT SYSTEM FOR SOFTWARE COMPONENT DESIGN BASED ON TESTABILITY.....	306
YASUHIRO NOGUCHI, DAIKI IHARA, SATORU KOGURE, KOICHI YAMASHITA, TATSUHIRO KONISHI & YUKIHIRO ITOH	
IDENTIFYING READING STYLES FROM E-BOOK LOG DATA.....	312
IVICA BOTICKI, HIROAKI OGATA, KARLA TOMIEK, GOKHAN AKCAPINAR, BRENDAN FLANAGAN, RWITAJIT MAJUMDAR & NEHAL HASNINE	
EXPLORING THE RELATIONSHIPS BETWEEN STUDENTS' ENGAGEMENT AND ACADEMIC PERFORMANCE IN THE DIGITAL TEXTBOOK SYSTEM	318
GÖKHAN AKÇAPINAR, MOHAMMAD NEHAL HASNINE, RWITAJIT MAJUMDAR, BRENDAN FLANAGAN & HIROAKI OGATA	
USING CLICKSTREAM TO UNDERSTAND LEARNING PATHS AND THE NETWORK STRUCTURE OF LEARNING RESOURCES: USING MOOC AS AN EXAMPLE	324
MING GAO, JINGJING ZHANG, DI SUN & JIANG ZHANG	

IDENTIFYING AND COMPARING ELEMENTARY STUDENTS’ PROBLEM-SOLVING BEHAVIOR PATTERNS USING LAG SEQUENTIAL ANALYSIS	330
Bo YANG, BAOPING LI	
A TOOL FOR LEARNING OF COGNITIVE PROCESS BY ANALYSIS FROM EXEMPLAR DOCUMENTS	336
WASAN NA CHAI, TANETH RUANGRAJITPAKORN, AND THEPCHAI SUPNITHI	
E-BOOK LEARNER BEHAVIORS DIFFERENCE UNDER TWO MEANINGFUL LEARNING SUPPORT ENVIRONMENTS .	342
JINGYUN WANG, ATSUSHI SHIMADA & FUMIYA OKUBO	
- POSTER PAPER -	
APPLICATION OF PROGRAMMING LEARNING SUPPORT SYSTEM TO OBJECT-ORIENTED LANGUAGE.....	348
SATORU KOGURE, KENTO OGASAWARA, KOICHI YAMASHITA, YASUHIRO NOGUCHI, TATSUHIRO KONISHI & YUKIHIRO ITOH	
LEARNING EVIDENCE ANALYTICS FRAMEWORK (LEAF) IN PRACTICE: A2I2 BASED TEACHER ADOPTION APPROACH	351
RWITAJIT MAJUMDAR, JAYAKRISHNAN M. WARRIEM, HIROYUKI KUROMIYA, GÖKHAN AKÇAPINAR, BRENDAN FLANAGAN & HIROAKI OGATA	
EXTENDING PROGRAM VISUALIZATION SYSTEM BASED ON TEACHER’S INTENT OF INSTRUCTION TO SUPPORT LEARNING DYNAMIC DATA STRUCTURES.....	354
KOICHI YAMASHITA, YUTA HIRAMATSU, SATORU KOGURE, YASUHIRO NOGUCHI, TATSUHIRO KONISHI & YUKIHIRO ITOH	
AN INDEX SYSTEM OF EDUCATION INFORMATION RESOURCES SELECTION BASED ON ANALYTICAL HIERARCHY PROCESS	357
JIAHUI RUAN, YILING HU, XIAOLING PENG	
DEVELOPING E-BOOK PAGE RANKING MODEL FOR PRE-CLASS READING RECOMMENDATION	360
CHRISTOPHER C.Y. YANG, GÖKHAN AKÇAPINAR, BRENDAN FLANAGAN & HIROAKI OGATA	
DISCOVERING THE IMPACT OF STUDENT COMMUNITIES IN EDUCATIONAL RECOMMENDATIONS.....	363
YONG ZHENG	
AN INVESTIGATION OF A MEDICAL TERMINOLOGY LEARNING ENVIRONMENT WITH A ROBOT AND A TABLET .	366
YOKO SAKAMOTO & NOBUHIRO SAKATA	
RECONSTRUCTION OF CONCEPT MAP TO PROMOTE LEARNERS’ COMPREHENSION ON NEW KNOWLEDGE.....	369
DIDIK DWI PRASETYA, TRIYANNA WIDIYANINGTYAS, TSUKASA HIRASHIMA, & YUSUKE HAYASHI	
BEYOND JUST FOLLOWING DATA: HOW DOES VISUALIZATION STRATEGY FACILITATE LEARNING ANALYTICS DESIGN?.....	372
JIAXIN CAO & YANJIE SONG	

Contingency Theory of Adaptive Practices Through the Lens of Eye Trackers

Tang Wee TEO^{a*}, Seng Chee TAN^b

^a*Natural Sciences and Science Education (Academic Group), Multi-centric Education Research and Industry STEM Centre, National Institute of Education, Nanyang Technological University, Singapore*

^b*Learning Sciences & Technologies (Academic Group), National Institute of Education, Nanyang Technological University, Singapore*

*tangwee.teo@nie.edu.sg

Abstract: In this paper, we report on a study of adaptive practices, as revealed by a teacher's eye gazes, in response to the contingencies that arose during a lesson. From the literature, four categories of adaptive practices, namely, *adaptive recognitions*, *adaptive anticipations*, *adaptive deliberations*, and *adaptive insights* were used for initial coding. In this study, eye tracking technologies are positioned as a mediator between the contingencies that arise in the classroom and the adaptive practices undertaken by teachers. The two research questions are: (1) What classroom events and/or objects, as revealed by the eye fixations, invoke the adaptive practice(s) of recognitions, anticipations, deliberations and/or insights during a lesson? (2) What events unfold following the enactment of the adaptive practice as informed by the eye fixations? The findings in this paper were based on a 29-minute lesson video of a biology lesson during which the teacher was wearing eye trackers. The four contingencies that arose during the lesson include: (1) students engaging in personal talks, (2) students not taking down notes, (3) students not looking confident when answering, and (4) student raised hand to seek clarification. This study offers new insights into the nature of teachers' adaptive practices in classroom teaching, with two new sub-categories of adaptive practices being identified. The findings suggest that eye-tracking technologies can help to generate new empirical insights on the nature of adaptive practices that teachers adopt in the classroom.

Keywords: eye tracking, contingency theory, adaptive practices

1. Introduction

Teachers engage in adaptive practices as they continually interpret the contextual-specific information related to the subject matter, ecological factors, and students (e.g., expressed emotions, actions, responses) in classroom teaching. As such, it can be challenging for teachers to be aware of the often implicit adaptive practices (Mylopoulos & Scardamalia, 2008) because of the many things that are simultaneously ongoing in class. Post-lesson reflections based upon recall are reconstructed memories that are partial and subjected to personal bias (Dempsey, 2010). Lesson videos offering meso-level insights into the classroom events (Tobin, 2017) may not capture the specific "object" (i.e., person or thing) that invoke the specific adaptive practice. This paper reports on a study of adaptive practices (Männikkö & Husu, 2019; Martin, Nejad, Colmar, & Liem, 2012), as revealed by teachers' eye gazes when enacting a lesson, with the goal to help them gain a deeper understanding of their adaptive teaching practices. According to Männikkö and Husu (2019), there are four categories of adaptive practices, namely, *adaptive recognitions*, *adaptive anticipations*, *adaptive deliberations*, and *adaptive insights*. In this study, eye trackers are used to objectively and accurately collect and analyze visual behavior (Tobii, 2019). Here, eye tracking technologies are positioned as a mediator between the contingencies that arise in the classroom and the adaptive practices undertaken by teachers. The theory of contingency (Donaldson, 2001) is applied here to explain the situatedness of the "organization", that is, the teacher. As key agents in upholding the rules, norms, regulations, and conventions of the school that they teach, teachers represent the embodied forms of the organizations (Freedman & Holmes, 2003). The research questions addressed are:

1. What classroom events and/or objects, as revealed by the eye fixations, invoke the adaptive practice(s) of recognitions, anticipations, deliberations and/or insights during a lesson?
2. What events unfold following the enactment of the adaptive practice as informed by the eye fixations?

The findings illuminate insights on what teachers “see” prior to enacting a specific type of adaptive practice in response to a classroom contingency. Here, we do not make a judgment of the (in)appropriateness of the adaptive practice in response to what they “see”. We also do not stop at identifying what they “see” or “should see” as these have been extensively reported in the teacher-noticing literature (see e.g., Sherin, Jacobs, & Philipp, 2011; Smith, 2012). While most previous studies have adopted eye-tracking technologies as a research tool, we use it as a *pedagogical mechanism* to tease out elements that feed into the choice of adaptive practices. By *pedagogical mechanism*, we refer to eye-tracking technologies as a system of parts—comprising multiplistic factors including attention span, personal preference on what one chooses to look at, beliefs about what events or objects deserve attention, the presence of distractors, and so on—that intersects with a lesson pathway comprising the classroom contingency and adaptive practice of a teacher. In integrating the social theory of contingency, adaptive practices and eye-tracking technologies, this paper builds on the existing work and enrich the scholarly discourse of each of the three fields to hone the epistemic quality (Kelly & Licona, 2018) of classroom teaching. Hence, this work makes practical contributions to classroom teaching.

2. Adaptive Practices and Contingency Theory

2.1 Adaptive Practices

Adaptive practices refer to teachers’ competencies in responding to the knowledge about students by adjusting their curriculum and teaching practices to achieve improved learning outcomes (Beltramo, 2017; Hammerness, Darling-Hammond, Bransford, Berliner, Cochran-Smith, & McDonald, 2005; Hatano & Oura, 2003; Lin, Schwartz, & Bransford, 2007). When teachers teach, they employ a set of “core practices” within their subject discipline (e.g., conduct science inquiry lessons in laboratory lessons), but also adapt these “core practices” flexibly contingent on the exigencies of their school or classroom contexts (Lampert, Boerst, & Graziani, 2011). Many scholars have argued for the necessity for teachers to engage in adaptive practices due to the complexity and fluidity of the social contexts in which education is embedded (Brown, 2004; Emdin, 2016; Tobin & Roth, 2006). Many studies about adaptive practices have focused on identifying the areas of expertise that teachers should have in order to engage in these practices effectively. The literature, for example, suggests the importance for teachers to develop a strong base of pedagogical content knowledge, a vision of ideal teaching, and a deep understanding and familiarity with their students (Fairbanks, Duffy, Faircloth, He, Levin, & Rohr 2010). However, deeper dialogue about the contingent nature of adaptive practices remain superficial and vague, hence, it is unhelpful to teachers who seek to unpack and understanding more deeply about their own teaching practices. Recently, Männikkö and Husu (2019) have identified adaptive practices of 17 primary school teachers, based upon recall interviews, and inductively coded these practices as fixed or open orientations of teaching. Table 1 summarizes the types of adaptive practices and alignment to the types of teaching orientation reported in their work.

Table 1

Types of adaptive practices and teaching orientations

Fixed orientation (intrapositions)	Open orientation (extrapolations)
Adaptive recognitions refer to teachers' established actions based on knowledge about students or rules that have been set. E.g., <ul style="list-style-type: none"> • Recalling what students have previously learnt or completed • Identifying rules in problem-solving that students should adhere to 	Adaptive deliberations refer to teachers' flexible actions derived from appraising and interpreting the ongoing events. E.g., <ul style="list-style-type: none"> • Making inferences about students' performance • Making appraisals on students' performance
Adaptive anticipations refer to teachers' customary practices based on beliefs or habits. E.g., <ul style="list-style-type: none"> • Identifying habits that students should have cultivated • Invoking basic beliefs that shape one's decisions in teaching 	Adaptive insights refer to teachers' new understanding about teaching and suggestions about new practices. E.g., <ul style="list-style-type: none"> • Wondering about imagined practices that one could adopt • Deriving new understandings about students

Based upon our interpretation, the key difference between the two columns in Table 1 lies in the degree of the contingency nature of the adaptive practices. In the case of the fixed orientations of teaching, the adaptive practices are *intrapositions* from existing knowledge or information about students. By intraposition, we mean looking inward at the “subject” of interest (i.e., student or self) to decision making on the next set of adaptive practices to engage. On the contrary, adaptive practices in the open orientations of teaching seem to be *extrapolations* from the interpreted new and possible, previous, information arising from the existing state of affairs. As such, the open orientation adaptive practices illuminate the contingent nature of teaching and hence, the adaptive nature of practices of teaching.

2.2 Contingent Theory and Adaptive Practices

The theory of contingency is a major theoretical lens used to describe organizations (Donaldson, 2001). The essence of contingency theory is that “organizational effectiveness results from fitting characteristics of the organization, such as structure, to contingencies that reflect the situation of the organization” (Donaldson, 2001, p. 1). In this paper, we position teachers and the teaching profession as the “organization” of embodiments of the institutional values, thinking, norms, and practices. Figure 1 illustrates how we view contingency at play in a classroom.

Education contexts are filled with contingencies that teachers have to respond to. According to Donaldson (2001), there are three underlying contingencies, namely, task uncertainty, task interdependence, and size. The uncertainty in tasks could be due to the nature of the tasks themselves, the technology, technology change, innovation, and environmental instability. In this case, the task can refer to the lesson to be delivered and the uncertainty may arise due to changes in student-related factors such as student behaviours, student performance, and change in classroom ecological conditions. Task uncertainty may be reinforced as a result of engaging in technology and innovation as opposed to direct teaching. Task interdependence may be in the form of prior knowledge required for students to learn a new topic or the set of experiences that a teacher must afford in order for students to participate fully in the subsequent activities. The size refers to the number of students in the class. It tends to be more challenging for teachers to handle larger class sizes due to greater diversities and hence, higher possibilities of unknown events happening.

Teachers will construct understandings of the contingencies that emerge during a lesson and act upon mediators such as their thinking—shaped by their beliefs, personal bias, and assumptions (Day, Pope, & Denicolo, 2013)—and external stimuli that affect what they hear, see, and feel. Informed by the literature on teacher noticing, we argue that teachers' responses to contingencies are, first and foremost,

affected by what they *see*. What they see will impact the choice of practices that they will subsequently adapt.

In order to elicit what teachers see, we harness the affordances of eye-tracking technologies to help us distil the pedagogical data or signals that teachers act upon in their adaptive practices. In the next section, we explain the research design and discuss the findings of the study.

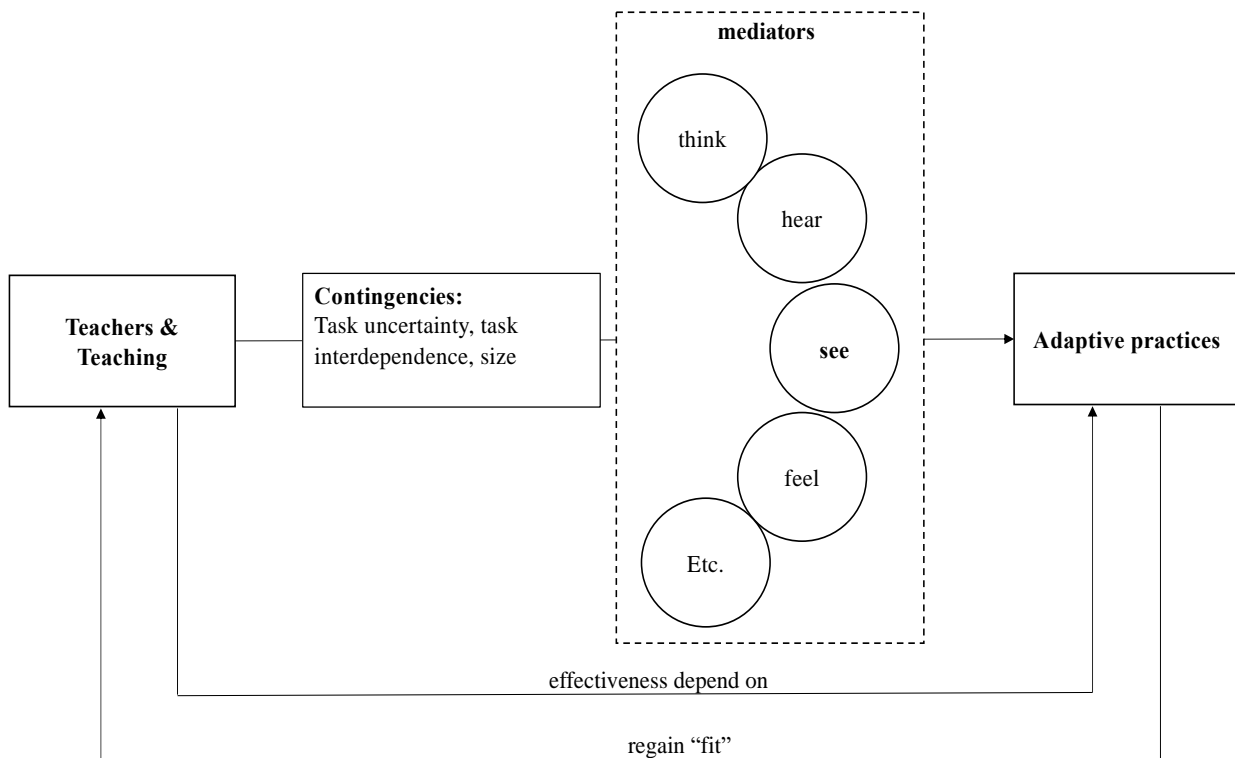


Figure 1. The relationship between teachers and teaching, and adaptive practices using contingency theory

3. Methods

3.1 Participants

The data reported in this paper were drawn from a larger eye-tracking study involving 10 teachers in Singapore secondary schools (for students of Grades 7-10, aged 13-16). This paper reports a biology teacher, Karina (pseudonyms are used in this paper) in her lessons with Grade 9 students in an all-girls secondary school. There were about 40 students in the class. At the time of the study, Karina had 16 years of teaching experience and had taught at the school for 15 years. We have chosen to focus on her as the main research participant for analysis in this paper because she was the most experienced among the teachers. Our assumption was that she would be more adept at adapting her practices in response to the contingencies that emerged during the lesson.

3.2 Data collection

The data collection involved videoing, eye-tracking and interviews with the teachers. Two (1 Front 1 back) video cameras were set up at the back and front of the classroom for the teacher and student views. During the lesson, the teacher wore the eye-tracker and data were wirelessly recorded via the transmitter to the laptop using the D-Lab software. As the eye-tracker was light-weight and portable, the level of discomfort for the teacher was minimized. Prior to the lesson, the teacher had explained to the students the purpose of the device and study hence, they were preempted on the changes around the classroom. The eye tracker recorded the movement of the eye pupils and simultaneously video recorded the view in

front of the eye tracker. Each lesson duration was about 30 minutes. We also interviewed Karina after each lesson observation and recording. The interview lasted about 60 minutes and was video recorded. Videos of the eye gazes on objects, the teacher's view lesson video, and student's view lesson videos were synchronized and combined for analysis. The interview was transcribed for analysis. For the purpose of this paper, only one of Karina's lesson videos was reported for detailed analysis.

3.3 Data analysis

The combined video data from one lesson were analyzed stepwise. In Phase 1, we identified the significant episodes during which adaptive practices were adopted. This is a form of event-oriented inquiry (Tobin, 2014) which helped us to focus on what is relevant and significant in illuminating the contingencies and subsequent adaptive practices in the classroom. A total of four episodes were identified. In Phase 2, we first adopted prescriptive coding using Männikkö and Husu's (2019) four categories and eight sub-categories (see Table 1) of adaptive practices to those episodes. Three out of the four categories (i.e., *adaptive recognitions*, *adaptive anticipations*, and *adaptive deliberations*) were identified and used to code the four episodes. Adaptive insight was not observed in the lesson analyzed. As such, adaptive practices identified in Karina's video data were not all the same as the ones identified in Männikkö and Husu's study. This could be because: (a) adaptive practices identified through recall interviews may not be the same as that enacted in practice; and/or (b) the adaptive practices in the lesson analyzed were different from the ones in Männikkö and Husu's study. However, only two of the eight sub-categories (i.e., *rules* and *habits*) of the adaptive practices were identified in two episodes. For the remaining two episodes, we have identified two new sub-categories (i.e., *alternative pedagogical tools* and *re-constructed explanations*) of adaptive practices. In Phase 3, we returned to the parts of the video coded for specific adaptive practice and code the "object" (e.g., student, table, book, ceiling, door, window) at which Karina was looking at when the adaptive practice was engaged. In Phase 4, we return to the part of the video coded in Phase 1 and 2 to identify the contingency that arises to draw the teacher's eye gaze. The constant comparative approach (Glaser, 1965) was employed to ensure that the data were analyzed in a rigorous manner and validity was achieved. Additionally, the data were independently coded and checked by two researchers. Discrepancies were negotiated and the data were recoded until 100-percent consistency was achieved.

4. Findings and Discussion

In Table 2, we summarized: (a) the episodes during which adaptive practices were enacted, (b) the specific type of adaptive practices adopted, (c) the "object" on which the eye gaze fell upon, and (d) the contingencies that arose that triggered the respective adaptive practices.

Table 2
Summary of episodes, adaptive practices, “object” of eye gaze, and contingencies

Episode	Adaptive Practice	“Object” of eye gaze	Contingencies
1 Karina disciplining two students engaging in personal talks	Adaptive recognition (rules): Pausing a lesson to discipline students	Two students	Students engaging in personal talks rather than paying attention to the lesson
2 Students taking down notes from the whiteboard	Adaptive anticipations (habits): Teachers complimented students who took notes from the whiteboard without being told	Whiteboard and students	Noticing that some students were not taking down notes
3 Having students guess whether the upper side of the leaf is the phloem or xylem	Adaptive deliberations (repetitions): demonstrating the movement of the hand again	Teacher’s lower arm and students	Students generally did not look confident when answering
4 Providing alternative answer to students	Adaptive deliberations (reconstructed explanations): Explaining using another set of words	Student who posed query	Student raised hand to ask for clarification

In what follows, we describe the four episodes during which adaptive practices were enacted and then provide our discussion of the episode.

Episode 1:

Time stamp	Descriptions of Karina’s eye gazes
00:25	Karina saw Nurhasni chewing.
00:35	Karina told Nurhasni to swallow her food (as they had just returned from recess) or to throw it out in the trash bin. She thanked her in advance for complying.
1:17	Karina started teaching. She asked, “What is the meaning...?” (paused) and stared at Nurhasni and Melissa (sitting next to each other at the second row in the column furthest away from Karina).
1:50	Nurhasni stood up to throw something into the trash bin. Karina’s eye gaze followed Nurhasni to the front of the class where the trash bin was located as she continued teaching.
2:08	Karina looked at Kelly (sitting in front of Nurhasni) as she turned around to talk to Nurhasni. Karina then looked at Nurhasni.
2:18	Karina looked at Divya (middle column, middle row) and told her to stop playing with her stapler.
2:31	Karina looked at Nurhasni as she was talking to Melissa. Now Nurhasni has turned her body 90-degrees to face Melissa.
3:31, 3:36, 3:48, 3:49-3.50, 3:58-4:00, 4:01	Karina’s eye gazes were on Nurshani on these instances.

Time stamp	Descriptions of Karina's eye gazes
4:23	Karina looked at Melissa and asked if the textbook belonged to her or Nurhasni. Melissa said the textbook was hers. Nurhasni said she left her textbook at home. Karina told Nurhasni, "You need to bring your textbook yah?"
4:48, 4:50, 4:53-4:54	Karina's eye gazes were on Nurshani on these instances.
4:58	Karina requested to the researcher to pause the data collection and removed her eye tracker. The whole class looked in Karina's direction. As Karina walked towards Melissa, her jaws dropped and looked in fear. Karina stood in front of Nurhasni and spoke with her (the content of the conversation was unrecorded).
5:27	Karina walked back to her teacher's desk.
5:29	Melissa seemed embarrassed as she shielded her face with her left hand. Subsequently, Melissa and Nurhasni did not engage in private talks. The whole class had eyes on Karina. Karina's eye gazes left Nurhasni and Melissa and were distributed to the rest of the class until the end of the lesson. She did more content teaching thereafter.

In Episode 1, Karina was seen adopting the practice of *adaptive recognitions* of rules. She kept a close eye on Melissa and Nurhasni as she noticed the two of them engaging in frequent private talks. As such, her eye gazes frequently landed on the two girls. She did not openly reprimand them until much later (timestamp at 4:58) when she took a drastic decision to remove her eye trackers, probably to respect the privacy of the students while scolding them or to not have the eye tracker hinder her direct face-to-face interaction with the students). In this case, the flow of her lesson was disrupted and she stopped her lesson to address a recurrent discipline issue. Her adaptive practices were enacted as a result of her expectations of proper behavior (habits) when she was teaching. She was very attentive to the misbehaviours and hence, her eye gazes frequently landed on students who were engaging in non-lesson related activities. The contingencies that arose, in this case, were the lack of alignment of student behaviours to the classroom norms which Karina had set. Following her adaptive practice, the two students paid attention for the rest of the lesson and stopped engaging in personal talks.

Episode 2:

Timestamp	Descriptions of Karina's eye gazes
10:04	Karina was teaching and she said, "This is a very important idea. Good, I see some of you writing." The students sitting in the front row could be seen picking up their pen to write. She stopped talking and the students were taking notes.

Karina adopted the practice of *adaptive anticipations of habits*. Instead of telling students that they should be taking down notes without being told to do so, she praised the students who had taken the initiative to do it. In doing this, she was instilling in students the good habits of independent learning without constant and direct reminders. She praised those who did it as a means to also imply that those who did not do so, did not know how to take ownership of their own learning. The contingencies of this practice were her observations of the different student actions—those who were writing and those who were not writing anything and simply staring at the teacher. Subsequently, the whole class of students was taking notes. Karina paused the lesson for them to complete their writing.

Episode 3:

Timestamp	Descriptions of Karina's eye gazes
15:44	Karina drew schematic diagrams of the cross-sections of roots on the whiteboard.
16:17	Karina demonstrated the drawing of schematic diagrams of the cross-section of the stem.
16:20	Karina recapped on the concept of xylem and phloem which students had prior knowledge.
16:58	Karina labeled the words "xylem" (on her inner lower arm) and "phloem" (on her outer lower arm). She said that the whole hand is a "vascular bundle".
18:19	She modeled the extension of a leaf as it grew and asked if the xylem or phloem was facing the top of the leaf. Not all students answered and they generally did not sound confident.
18:29	Karina demonstrated one more time.
18:43	Karina asked again if the xylem would be at the top or bottom of the leaf. This time, more students answered that it would be on top. The collective voice was louder and more students sounded confident of their answer.
18:46	Karina reiterated, "The xylem ends up being on top."
19:13	Karina said, "That is how you remember."

Karina was seen making the *adaptive deliberations of repetitions* and using her hand as a pedagogical tool to help students identify a convenient and vivid way of remembering the content. Her decision to repeat her demonstration was contingent on her observation and interpretations of the lack of confidence in the students' responses, probably from the loudness and number of students answering. Karina had used this method to help students remember the content many times before and she thought it would be useful for this group of students as well. Hence, she had applied this method again to help students derive at the answer confidently and concretize their learning in a visual manner.

Episode 4:

Timestamp	Descriptions of Karina's eye gazes
21:54	Karina looked up at the clock (indicated 11.30am) and decided that she would not carry on teaching. She highlighted the importance of knowing how to draw the cross-sections of the root, stem, and leaf. Karina asked the students if they had any questions.
22:30	A student asked a question, "What is trans-locate?" It was a word that Karina had used earlier.
22:32	Karina explained, "'Trans-locate' means to transfer from one location to another. In this context, it refers to where it is made to where it is needed."

Karina adopted the *adaptive deliberation of reconstructing explanations* to help a student understand a jargon more easily. The student expressed difficulty understanding the meaning of the word "trans-locate". Instead of providing a formal definition, Karina first explained what it meant in layman terms, and then contextualized it to the topic that she was teaching to help the student understand the term more easily. In this case, she had deliberated on the method of making the word clear, the choice of words to use and process to scaffold it from decontextualized to contextualised meaning making. Her responses were contingent on the type of question that the student asked, that is, the clarification of word meaning.

5. Implications and Conclusion

In this paper, we reported on the use of eye-tracking technologies as a mediator of the pathway from the emergence of classroom contingencies to the adaption of practices in a lesson. In our analysis of a 29-minute lesson video, we identified the contingencies that arose in the lessons and were captured by the eye-trackers as: (1) students who engaged in personal talks, (2) students not taking down notes, (3) students not looking confident when answering, and (4) student raised hand to seek clarification. For (1) the eye gazes were fixed on students and adaptive recognitions of rules were practiced. For (2), the eye gazes were fixed on the whiteboard and students, and adaptive anticipations of habits were practiced. For (3) the eye gazes were fixed on the teacher's own lower arm and students, and adaptive deliberations of repetitions were practiced. For (4) the eye gazes were fixed on the student asking a question, and adaptive deliberations of reconstructed explanations were practiced. The adaptive practices led to improved discipline and enhanced understanding.

This study illuminated the process in which the eyes performed as a mediator in the emergence of classroom contingencies and the types of adaptive practices that result in improved and positive outcomes. Insights on the nature of teachers' adaptive practices in classroom teaching can be drawn. Two new sub-categories of adaptive practices, namely, adaptive deliberations of repetitions and adaptive deliberations of reconstructed explanations, that have not been reported in Männikkö and Husu's (2019) study have been identified here. This implies that eye-tracking technologies can afford new empirical insights on the nature of adaptive practices that teachers adopt in the classroom. This knowledge can be used to inform teacher preparation courses to get them ready for the types of contingencies and hence, potential useful practices that they can adapt in the classrooms. As such, we argue that eye-tracking technologies can potentially contribute to the literature on teaching and teacher education. Teacher educators can consider using eye-tracking technologies to help preservice teachers identify and refine their adaptive practices when teaching (e.g., microteaching scenarios).

6. Limitations

This paper is based upon the case study of one teacher to analyze the occurrences that took place immediately before the next action or spoken word. While we do not aim to generalize the findings, the adaptive practices identified can add to the database of known adaptive practices that teachers adopt and form the base for coding other datasets. Additionally, we acknowledge that while the data collected were meant to be objective and accurate, in comparison to other types of data that were collected retrospectively (e.g., stimulated recall interviews), the object of interest in the eye gaze was contingent on the nature of the adaptive practice identified by the researchers. Hence, inter-rater reliability in the form of independent coding by two or more researchers was necessary to ensure validity.

Acknowledgements

This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (ARF 02/17 TSC) and administered by National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE and NIE.

References

- Beltramo, J. L. (2017). Developing adaptive teaching practices through participation in cogenerative dialogues. *Teaching and Teacher Education, 63*, 326-337.
- Brown, B. A. (2004). Discursive identity: Assimilation into the culture of science and its implications for minority students. *Journal of Research in Science Teaching, 41*(8), 810-834.
- Day, C., Pope, M., & Denicolo, P. (2013). *Insights into teachers' thinking and practice*. London, UK: Routledge.
- Dempsey, N.P. (2010). Stimulated recall interviews in ethnography. *Qualitative Sociology, 33*, 349-367.
- Donaldson, L. (2001). *The contingency theory of organizations*. Thousand Oaks, CA: Sage.

- Emdin, C. (2016). *For white folks who teach in the hood ... and the rest of y'all too: Reality pedagogy and urban education*. Boston, MA: Beacon Press.
- Fairbanks, C. M., Duffy, G. G., Faircloth, B. S., He, Y., Levin, B., Rohr, J., & Stein, C. (2010). Beyond knowledge: Exploring why some teachers are more thoughtfully adaptive than others. *Journal of Teacher Education*, 61(1–2), 161–171.
- Freedman, P., & Holmes, M. S. (2003). *The teacher's body: embodiment, authority, and identity in the academy*. Albany, NY: State University of New York Press.
- Glaser, B. (1965). The Constant Comparative Method of Qualitative Analysis. *Social Problems*, 12(4), 436-445.
- Hammerness K, Darling-Hammond L, Bransford J, Berliner D, Cochran-Smith M, McDonald M and Zeichner K. (2005). How teachers learn and develop. In Darling-Hammond, L. & Bransford, J. (Eds), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 390-441). San Francisco: Jossey-Bass.
- Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher*, 32(8), 26-29.
- Kelly, G. J., & Licona, P. (2018). Epistemic practices and science education. In M. Matthews (Ed.), *History, Philosophy and Science Teaching* (pp. 139–165). Springer.
- Lampert, M., Boerst, T. A., & Graziani, F. (2011). Organizational resources in the service of schoolwide ambitious teaching practice. *Teachers College Record*, 113(7), 1361-1400.
- Lin, X., Schwartz, D. L., & Bransford, J. (2007). Intercultural adaptive expertise: Explicit and implicit lessons from Dr. Hatano. *Human Development*, 50, 65-72.
- Martin, A.J., Nejad, H., Colmar, S., & Liem, G.A.D. (2012). Adaptability: Conceptual and empirical perspectives on responses to change, novelty and uncertainty. *Australian Journal of Guidance and Counselling*, 22, 58-81.
- Männikkö, L., & Husu, J. (2019), Examining teachers' adaptive expertise through personal practical theories, *Teaching and Teacher Education*, 77, 126-137.
- Mylopoulos, M. & Scardamalia, M. (2008). Doctors' perspectives on their innovations in daily practice: implications for knowledge building in health care. *Medical Education*, 42(10), 975-981.
- Sherin, M., Jacobs, V., & Philipp, R. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Abingdon, OX: Taylor & Francis.
- Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology*, 16, 53-81.
- Tobii (2019). What is eye tracking? Retrieved from <https://www.tobii.com/blog/what-is-eye-tracking/>
- Tobin K. (2014) Using Collaborative Inquiry to Better Understand Teaching and Learning. In: Bencze J., Alsop S. (eds) *Activist science and technology education. Cultural Studies of Science Education* (p. 127-147). Springer, Dordrecht.
- Tobin, K. (2018). Methodological bricolage. In S. M. Ritchie & K. Tobin (Eds.), *Eventful learning*. Rotterdam, The Netherlands: Sense Publishers.
- Tobin, K., & Roth, W. M. (2006). *Teaching to learn: A view from the field*. Rotterdam: Sense Publishers.

Simulation of Online Learning Interaction Relation Network Based on BA Model

Peng ZHOU, Huan LI*, Meiyi GUO, Chenchen SHANG & Di WU

*National Engineering Research Center for E-learning,
Central China Normal University, Wuhan, China*

*Huan Li@corresponding.author

Abstract: To explore the knowledge dissemination in online learning interaction, the online learning interaction relation network should be constructed firstly. However, the random network, small-world network and scale-free network proposed in the current research can not describe the interactive relationship of online learning interaction. Therefore, how to build the online learning interaction relation network has become a fundamental research problem to be solved. This study updates the BA model algorithm to construct a simulated interaction relation network of online learning. The graphs of degree distribution show that the interaction relation network simulated in this study conforms to the actual rule of online learning interaction of instructors and learners.

Keywords: Online learning, learning interaction, learning relation network, BA model

1. Introduction

Online learning gains more and more attention with its massive resources and instant interactivity (Mikalef, Giannakos, Chorianopoulos, & Jaccheri, 2013; Kuiper, Volman, & Terwel, 2005). In online learning environment, knowledge and information spread by participants' interaction relation path. To explore how knowledge and information disseminate in online learning interaction, the online learning interaction relation network should be built. In online learning interaction relation network, the nodes and edges are significantly different from the social network generated in social media. The online learning interaction relation network can not be described by random network, small-world network and scale-free network (Barabási & Albert, 1999; Erds & Rényi, 1960; Watts & Strogatz, 1998). Therefore, how to generate the online learning interaction relation network to simulate the dissemination of knowledge and information is the fundamental problem in the field of online learning interaction.

As one algorithm for generating the scale-free network, Barabási-Albert (BA) model has great advantages in explaining the formation of social networks because of its two characteristics of growth and priority connection mechanism (Barabási & Albert, 1999). Krawczyk, Kułakowski, and Hołyst (2018) proposed an algorithm to imitate a series of consecutive conflicts between leaders in social groups by using the fragments of scale-free Barabasi-Albert networks. DasGupta, Mobasheri, and Yero (2019) constructed a large number of synthetic networks generated by the Barabasi-Albert preferential attachment model to shed light on privacy violation properties of eight real social networks. In the field of online learning, relation network structure of instructor-learner has not received sufficient attention. This study explored the instructor-learner online learning interaction of one Chinese university, and designed an algorithm for generating an online learning interaction relation network based on the BA model.

Firstly, this paper reviewed the literature of online learning interaction and social network model. Then, we designed an algorithm to generate the online learning interaction relation network based on the BA model. Finally, the degree distribution diagram of generated relation network was presented. The result showed that the interaction relation network simulated in our study conformed to the online learning interaction of instructors and learners in real world.

2. Literature Review

2.1 Online Learning Interaction

Interaction was defined as a dialogue or discourse or event between two or more participants and objects which occurred synchronously and/or asynchronously mediated by response or feedback and interfaced by technology (Muirhead & Juwah, 2004). In the online learning environment, learning interaction was achieved through collaborative behaviors, from learners' sharing the diverse perspectives of the other group members, to being able to seek feedback and clarify ideas through the group's communication, either electronic or through other forms of communication stimulated by the electronic group communication (Wilson & Stacey, 2004).

Online interaction might benefit learning from 3 aspects. (a) Interaction might improve learners' satisfaction. Hong (2002) discovered that interaction may improve learners' satisfaction, and learners who highly perceived the learner-instructor interaction were more satisfied with the course. (b) Interaction might improve learning outcome. Kurucay and Inan (2017) found that learners working collaboratively achieved significantly higher than those working individually. Gunawardena, Linder-VanBerschoot, LaPointe, and Rao (2010) reported that interaction between learners was a significant predictor of achievement. Jung, Choi, Lim, and Leem (2002) concluded that learner-to-instructor interaction that included academic and social communications increased achievement. (c) Interaction might enhance learners' sense of community. Nistor, Daxecker, Stanciu, and Diekamp (2015) investigated the correlation between interaction and the sense of community, and found that intensive interactions within the community could lead to stronger emotional connections between members, and a similar conclusion was proved by Luo, Zhang, and Qi (2017). Online learning interaction could promote learning significantly, but how knowledge and information disseminated in online learning environment was still a question to be studied.

2.2 Social Network Model

To explore the knowledge dissemination of online learning interaction, it is necessary to study the network structure of online learning interaction. The existing research explored the network structure by social network theory.

Social network theory pointed out that social network was a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors (Wasserman & Faust, 1994). In learning contexts, online social networking behavior was related to learning and academic success by creating systems of information, contacts and support (Yu, Tian, Vogel, & Kwok, 2010). Thoms used social network to analyze how social media was chosen in distance learning, they also used social network analysis to build the "read" networks and "reply" networks, and they found that higher network diameters were more characteristic of performing learning networks (Thoms & Eryilmaz, 2014). Hernández-García, González-González, Jiménez-Zarco, and Chaparro-Peláez (2015) explored the relationship between learning analysis parameters and learner outcomes, and showed how the visualization of social learning analysis could help observe the visible and invisible interactions that occur in online distance education.

Common complex network models include: the Erdos-Renyi network, the Watts-Strogatz small-world network and the scale-free network. The Erdos-Renyi network is a complex network that is built through a random process (Erds & Rényi, 1960). It is based on a "natural" construction method: assume that there are n nodes, and assume that the probability of connection between each pair of nodes is constant $0 < p < 1$. The Watts-Strogatz small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps (Watts & Strogatz, 1998). The scale-free network is a complex network with a degree distribution obeying or close to a power law distribution (Barabási & Albert, 1999). However, these three social networks can not describe the interactive relationship of online learning interaction, so it is necessary to introduce a new relational construction model.

2.3 Relational Construction Model

The Barabási-Albert model was a model proposed by Barabási and Albert (1999) to explain the scale-free characteristics of complex networks. This model was found to be a consequence of two generic mechanisms: (a) networks expand continuously by the addition of new vertices, and (b) new vertices attach preferentially to sites that are already well connected (Song, Havlin, & Makse, 2005). Xie et al. (2012) used a model of Barabasi-Albert scale-free networks to study how the presence of such groups within social networks affects the outcome and the speed of evolution of the overall opinion on the network. Jiang, Chen, and Liu (2014) used the Barabasi-Albert scale-free network to model the dynamic information diffusion process in social networks, which showed that the proposed game theoretic model could well fit and predict the information diffusion over real social networks.

Based on the two characteristics of growth and priority connection mechanism of Barabási-Albert model, this study improves its algorithm and simulates the relation network of online learning interaction.

3. Graph Model of Interaction Network

Social network is usually presented in the form of a graph composed of nodes and edges. In social network, nodes represent individuals or organizations, edges represent their social relationships including friendships, classmate relationships, business partnerships, ethnic beliefs, etc. In online learning interaction relation network, there are two types of nodes, one is the learner node and the other is the instructor node. Edges in online learning interaction relation network represent the interactive relationship of instructors and learners.

In online learning, instructors mainly build knowledge, guide learners, exchange information, and feedback results. Learners access information from network, exchange information from peers, participate in discussions, and reflect on the learning process. There are three types of edges in online learning: instructor-instructor edge, instructor-learner edge, and learner-learner edge. Interaction between instructor-instructor mainly takes place in the online teaching and research community, where instructors can conduct online collaborative learning, share learning resources and research results, and establish knowledge connections. Interaction between instructor-learner can be divided into cognitive interaction and emotional interaction. Instructors and learners conduct cognitive interaction through questions, rebuttals, and assessments, and they conduct emotional interaction by expressing thank-you words, inspiring each other, and emitting emoji. Interaction between learner-learner usually occurs when learners are engaged in group discussions and peer-to-peer evaluations.

For a vertex of social network, the number of heads ends adjacent to a vertex is called the in-degree and the number of tails ends adjacent to a vertex is its out-degree (Bondy & Murty, 1976). Whereas node degrees characterize individual nodes, one can define a degree distribution to quantify the diversity of the whole network (Albert, 2005). The degree distribution $P(k)$ of a network is defined to be the fraction of nodes in the network with degree k , and the value of k is $k=1, 2, 3, \dots$. There is a power law: $P(k) = Ak^{-\gamma}$. Where A is a constant that adds up the $P(k)$ value to 1, and the exponential index of degrees is usually in the range $2 < \gamma < 3$ (Albert & Barabási, 2002).

Due to the different nature of the two types of nodes in the online learning interaction relation network, the degree distribution of the instructor and the degree distribution of the learner are also different. In order to more accurately analyze the degree distribution of instructors and learners, we separately study the three types of edges of the online learning interaction relation network, which is shown in Figure 1. From the perspective of the instructors alone, the network formed by instructor-instructor interaction is a scale-free network, and its degree distribution is $P(k_I)_1 = Ak_I^{-\gamma}$. Similarly, the network formed by learner-learner interaction is also a scale-free network, and its degree distribution is $P(k_L)_1 = Ak_L^{-\gamma}$.

From the perspective of the connection between the instructor and the learner, the Tie of L-I is the edge that connects the two networks above. In the network formed only by instructor-learner interaction, the degree distribution of the instructors is $P(k_I)_2 = Ak_I^{-\gamma}$, and the degree distribution of the learners is $P(k_L)_2 = Ak_L^{-\gamma}$. Combining these two situations, the degree distribution of the

instructors is $P(k_I) = P(k_I)_1 + P(k_I)_2$, and the degree distribution of the learners is $P(k_L) = P(k_L)_1 + P(k_L)_2$.

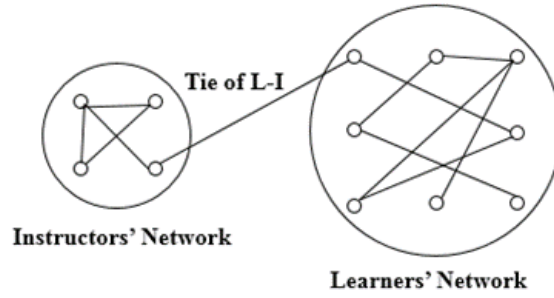


Figure 19 · The online learning interaction relation network

4. Algorithm of Interaction Network

In order to generate the online learning interaction relation network to simulate the dissemination of knowledge and information, we improve the algorithm of BA model, which is shown in Figure 2. Firstly, the $mi0$ instructor nodes and the $ml0$ learner nodes are gathered together to form the initial isolated network. Next, we add new nodes to the initial network to make the network size reach N nodes, in which the probability of adding new instructor nodes is mi . If the newly added node is an instructor node, it is first connected to ki instructor nodes and then to $ni-ki$ learner nodes in the existing network by Roulette Method. Similarly, if the newly added node is a learner node, it is first connected to kl learner nodes and then to $nl-kl$ instructor nodes in the existing network by Roulette Method. Then, the online learning interaction relation network is generated. Finally, the degree distribution of the whole network is shown in the form of degree distribution graph.

Algorithm 1 Starting from the existing network created by $m0$, BA scale-free network is generated by using growth and priority connection mechanism.

Input: Initial nodes which represents the numbers of instructors, $mi0$; initial nodes which represents the numbers of learners, $ml0$; Number of new edges generated each time an instructor node is introduced, ni ; Number of new edges generated each time a learner node is introduced, nl ; isolated networks generated by $mi0$ and $ml0$, A ; network scale after growth, N .

Output: Degree distribution graph.

```

1: 0 or 1 randomly generates  $matrix[N-mi0-ml0,1]$  with probabilities of  $1-mi$  and  $mi$ , which
   extends  $nodes[mi0+ml0,2]$  to  $nodes[N,2]$ ;
2: Set the number of interactions between instructors and instructors,  $k_i$  and the number of
   interactions between learners and learners,  $k_l$ ;
3: for  $k = mi0 + ml0 + 1$  to  $N$  do
4:   Set the total interaction frequency,  $p$  and the interaction frequency between instructors
   and instructors,  $pi$  and the interaction frequency between learners and learners,  $pl$ ;
5:    $pp \leftarrow cumsum(p)$ ;  $ppi \leftarrow cumsum(pi)$ ;  $tpl \leftarrow cumsum(pl)$ ;
6:   if New node is an instructor then
7:      $nCount \leftarrow k_i$ ;
8:     for  $i = 1$  to  $nCount$  do
9:       Choose an instructor node by Roulette Method,  $jj_i$ ;
10:       $A[k, jj_i] \leftarrow 1$ ;  $A[jj_i, k] \leftarrow 1$ ;
11:    end for
12:     $nCount \leftarrow ni - k_i$ ;
13:    for  $i = 1$  to  $nCount$  do
14:      Choose a learner node by Roulette Method,  $jj_l$ ;
15:       $A[k, jj_l] \leftarrow 1$ ;  $A[jj_l, k] \leftarrow 1$ ;
16:    end for
17:   else
18:      $nCount \leftarrow k_l$ ;
19:     for  $i = 1$  to  $nCount$  do
20:       Choose a learner node by Roulette Method,  $jj_l$ ;
21:       $A[k, jj_l] \leftarrow 1$ ;  $A[jj_l, k] \leftarrow 1$ ;
22:     end for
23:      $nCount \leftarrow nl - k_l$ ;
24:     for  $i = 1$  to  $nCount$  do
25:       Choose an instructor node by Roulette Method,  $jj_i$ ;
26:       $A[k, jj_i] \leftarrow 1$ ;  $A[jj_i, k] \leftarrow 1$ ;
27:     end for
28:   end if
29: end for
30: Generate the degree distribution graph of instructors and learners.

```

Figure 20. The improved algorithm of BA model

5. Static Properties of Interaction Network

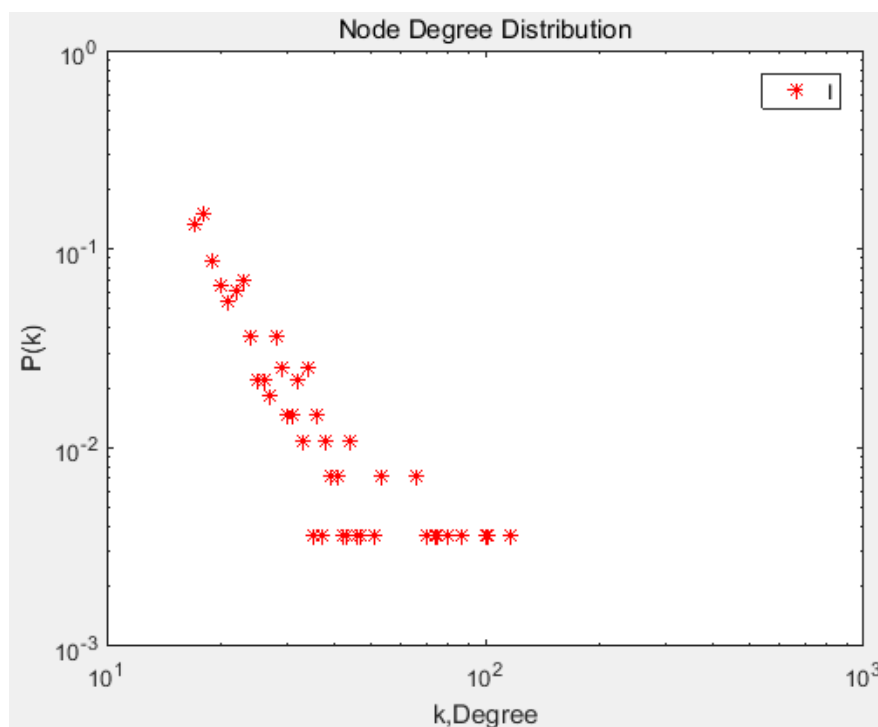
The degree distribution of the online learning interaction relation network generated by our improved algorithm of BA model is shown in Figure 3, and the parameter setting of the algorithm is shown in Table 1.

Table 6

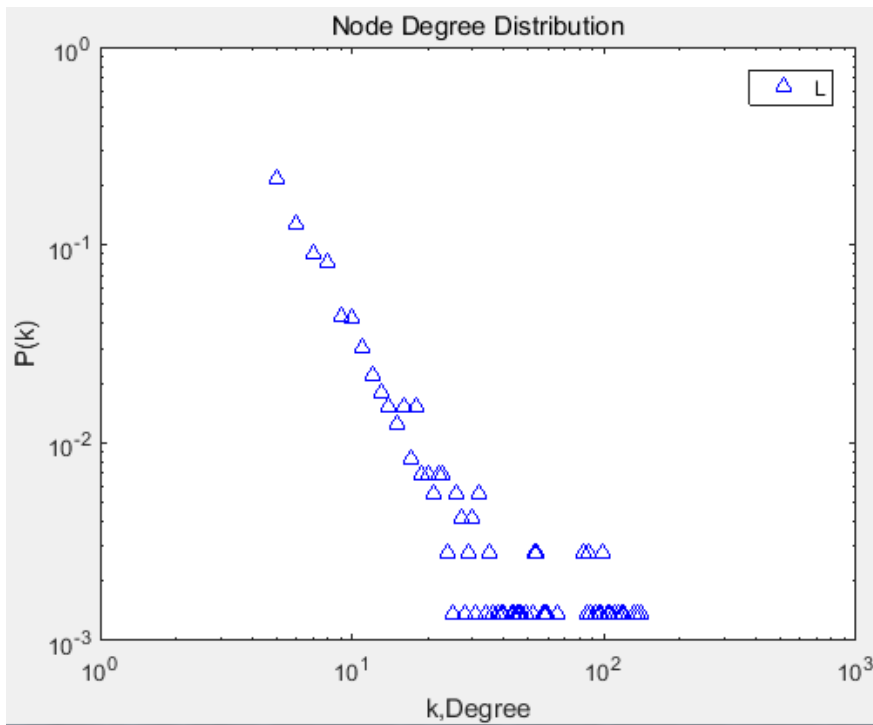
The Parameter Setting

Parameter	Definition	Value
$mi0$	The initial number of instructor nodes	3
$ml0$	The initial number of learner nodes	100
mi	The probability of a newly introduced node being an instructor node	0.3
ki	The number of instructor nodes to connect in the existing network when the newly introduced node is an instructor	2
ni	The number of total nodes to connect in the existing network when the newly introduced node is an instructor	17
kl	The number of learner nodes to connect in the existing network when the newly introduced node is a learner	2
nl	The number of total nodes to connect in the existing network when the newly introduced node is a learner	5
N	The scale of the network after the increase	1000

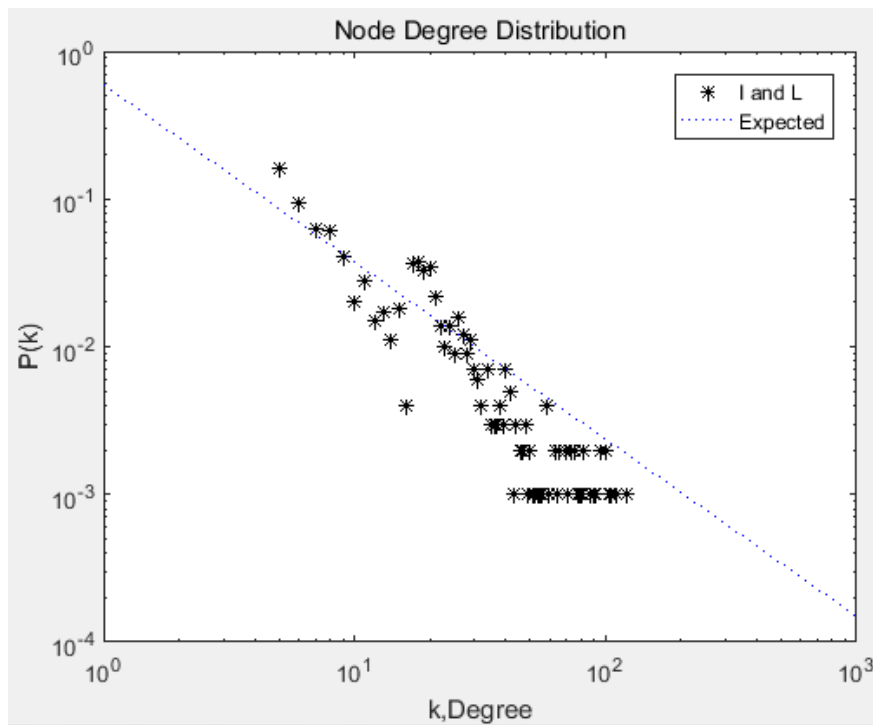
As shown in Figure 3, (a) is the degree distribution of instructors, (b) is the degree distribution of learners, and (c) is the degree distribution of instructors and learners. It can be seen that the degree distribution of the instructors and the learners are both consistent with the characteristics of the scale-free network, and the average degree of the instructors is higher than the average degree of the learners, which is in line with the actual rule of online learning interaction. Although the online learning interaction relation network consists of two types of nodes and three types of edges, it is still a scale-free network.



(a)



(b)



(c)

Figure 21. The degree distribution of the online learning interaction relation network

6. Conclusion

The purpose of this study is to simulate the interaction relation network in the field of online learning setting. Based on the improved algorithm of BA model, we generate the online learning interaction relation network and show its degree distribution in the form of a degree distribution graph. The result shows that the interaction network generated in our research can greatly simulate the dissemination of knowledge and information during online learning interaction, which has practical significance for further study of the online learning interaction of instructors and learners. In the future, we will try to improve other algorithms to simulate a network that is more in line with the rule of online learning interaction. At the same time, the interactive relationship and interactive content in online learning interaction are also the key content we will explore.

Acknowledgements

This study was supported by the Project of Science and Technology Development Institute. The Research on Evaluation Index and Method of Teacher and Student's Stickiness in Cyber Learning Space(CCNU18KFY04).

References

- Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science*, 118(21), 4947-4957.
- Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
- Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.
- Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications (Vol. 290)*. London: Macmillan.
- DasGupta, B., Mobasheri, N., & Yero, I. G. (2019). On analyzing and evaluating privacy measures for social networks under active attack. *Information Sciences*, 473, 87-100.
- Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1), 17-60.
- Gunawardena, C. N., Linder-VanBerschot, J. A., LaPointe, D. K., & Rao, L. (2010). Predictors of Learner Satisfaction and Transfer of Learning in a Corporate Online Education Program. *American Journal of Distance Education*, 24(4), 207-226.
- Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, 47, 68-80.
- Hong, K. (2002). Relationships between students' and instructional variables with satisfaction and learning from a Web-based course. *The Internet and Higher Education*, 5(3), 267-281.
- Jiang, C., Chen, Y., & Liu, K. R. (2014). Evolutionary dynamics of information diffusion over social networks. *IEEE Transactions on Signal Processing*, 62(17), 4573-4586.
- Jung, I., Choi, S., Lim, C., & Leem, J. (2002). Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction. *Innovations in education and teaching international*, 39(2), 153-162
- Krawczyk, M. J., Kułakowski, K., & Hołyst, J. A. (2018). Hierarchical partitions of social networks between rivaling leaders. *PloS one*, 13(3), e193715.
- Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, 75(3), 285-328.
- Kurucay, M., & Inan, F. A. (2017). Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course. *Computers & Education*, 115, 20-37.
- Luo, N., Zhang, M., & Qi, D. (2017). Effects of different interactions on students' sense of community in e-learning environment. *Computers & Education*, 115, 153-160.
- Mikalef, K., Giannakos, M. N., Chorianopoulos, K., & Jaccheri, L. (2013). Does informal learning benefit from interactivity? The effect of trial and error on knowledge acquisition during a museum visit. *International Journal of Mobile Learning and Organisation*, 7(2), 158-175.
- Muirhead, B., & Juwah, C. (2004). Interactivity in computer-mediated college and university education: A recent review of the literature. *Journal of Educational Technology & Society*, 7(1), 12-20.
- Nistor, N., Daxecker, I., Stanciu, D., & Diekamp, O. (2015). Sense of community in academic communities of practice: predictors and effects. *Higher Education*, 69(2), 257-273.

- Song, C., Havlin, S., & Makse, H. A. (2005). Self-similarity of complex networks. *Nature*, 433(7024), 392.
- Thoms, B., & Eryilmaz, E. (2014). How media choice affects learner interactions in distance learning classes. *Computers & Education*, 75, 112-126.
- Wasserman, S., & Faust, K. (1994). Social network analysis in the social and behavioral sciences. *Social network analysis: Methods and applications, 1994*, 1-27.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440.
- Wilson, G., & Stacey, E. (2004). Online interaction impacts on learning: Teaching the teachers to teach online. *Australasian Journal of Educational Technology*, 20(1), 33-48.
- Xie, J., Emenheiser, J., Kirby, M., Sreenivasan, S., Szymanski, B. K., & Korniss, G. (2012). Evolution of opinions on social networks in the presence of competing committed groups. *PLoS One*, 7(3), e33215.
- Yu, A. Y., Tian, S. W., Vogel, D., & Kwok, R. C. (2010). Can learning be virtually boosted? An investigation of online social networking impacts. *Computers & Education*, 55(4), 1494-1503.

Clustering Models for Topic Analysis in Graduate Discussion Forums

Gokarn Mallika NITIN, Swapna GOTTIPATI*, Venky SHANKARARAMAN
School of Information Systems, Singapore Management University, Singapore
mallikag.2015,*swapnag,venky@smu.edu.sg

Abstract: Discussion forums provide the base content for creating a knowledge repository. It contains discussion threads related to key course topics that are debated by the students. In order to better understand the student learning experience, the instructor needs to analyse these discussion threads. This paper proposes the use of clustering models and interactive visualizations to conduct a qualitative analysis of graduate discussion forums. Our goal is to identify the sub-topics and topic evolutions in the discussion forums by applying text mining techniques. Our approach generates insights into the topic analysis in the forums and discovers the students' cognitive understanding within and beyond the classroom learning settings. We developed the analysis model and conducted our experiments on a graduate course in Information Systems. The results show that the proposed techniques are useful in discovering knowledge from the forums and generating user-friendly visualizations. Such results can be used by the faculty to analyse the students' discussions and study the strengths and weaknesses of the students' cognitive knowledge on course topics.

Keywords: Topic analysis, Online Discussion Forums, Clustering Models, Topic evolutions

1. Introduction

Online discussion forums are advantageous as they provide an equitable space and flexibility for all students. They offer an alternative means for encouraging interactions between the students and the instructors. Discussion forums are classified into three types; standard forum for general use, single simple discussion forum and question and answering forum. General use forums are useful for any random discussions posted by any user with or without the goal of testing or learning more about the topics covered in the class. Simple discussion forum is focused on a single topic or a subject. Question and answer forums are focused on several diversified topics and aimed at supporting the learning process. Most of the student forums are usually designed as Q&A forums (Andy, 2013) where the instructor posts a questions and the students submit their answers.

To measure the students' learning, instructors always want to know what the students are discussing. For effective learning through discussions, it is important for the instructors to intervene in the threads (Simon et al. 2016). Effective intervention requires an understanding of the content and analysis of the forums (Chaturvedi, 2014). Use of analytics on discussion forums is focused on collecting, analysing, and displaying the "traces" that learners leave behind, with a purpose to improve learning (Atapattu, Falkner et al. 2016). For example, if the students are discussing less on certain topic, instructors can post additional supporting hints for the students to continue the discussions. Emergent topics or declining topics require more instructor intervention. Another example is that if the students are digressing from the topic, instructors can control the discussions or encourage further learning points in the digressed topics.

In this paper, we propose a solution for discovering the topical insights from the discussion forums based on a text analytics approach. In particular, understanding the topics and sub-topics that are emerging in the discussions provide useful insights into the student learning process. For example, if the students have discussed only the main topics that were covered in class, it indicates that the students are bounded to in-class learning and have not taken efforts to do further research on their own. If more sub-topics, that were not covered in the class, emerge from the main topic, it indicates the out of class learning process of the students. In this digital era, it is important for students to learn beyond the classroom, and further scaffold this learning, by instructors intervening to identify the links between the

various sub-topics and providing a summary of the topical evolutions. However, a manual approach to this is very time-consuming, since the instructor has to read all the posts and generate topics, sub-topics and the evolutions. Using automated tools to help gain insights from the discussion forum posts holds great promise for providing adaptive support to individual students and collaborative groups.

We use data from the online discussion forum of a masters course, “Text Analytics and Applications” taught at the School of Information Systems, Singapore Management University. A study by Burge (1994) of Master of Education students enrolled in a web-based distance program identified challenges that related to peer interaction, difficulties associated with handling and managing large quantities of information and discussion fragmentation. Therefore, we design the discussion forum with the controlled and challenging threads, so that the students can appreciate and participate in the organized discussions. The knowledge generated from such posts can be applied to their project and exam preparations.

This paper is structured as follows. Section 2 will review the background of discussion forum analysis and topics analysis of posts. Section 3 provides a background of text analytics. Section 4 describes the research problem statement along with the context. Section 5 presents the solution models with details of the text analytics techniques that are used in the models. In section 6, we describe findings, analysis and answer our research questions, and we conclude in Section 7.

2. Related Work

2.1 Discussion Forum Analysis

Several researchers have studied interactions in the classroom since the 1960s to help quantify verbal behaviour. Applications of interaction analysis include improvement of teaching style and pupil achievement through reflection, using the classification of interaction type (Amidon, 1968). Additionally, an adapted form of Flanders’ system of Interaction Analysis was used to understand and provide feedback on teaching behaviour in a foreign language classroom, to support future classroom planning and improve content delivery (Wragg, 1970). Lively online discussions can be facilitated by requiring participants to not only post their work but also comment and respond to each other’s submissions. Classroom live discussions capturing and analysing is also an important research work for better learning process (Venky et al. 2018). As a result, the discussions become more than just an assignment; students learn from each other and become more engaged in the learning process.

Learning analytics is focused on collecting, analysing, and displaying the “traces” that learners leave behind, with a purpose to improve learning (Duval, 2011). The system developed by Leony et al. (2012) captures and visualizes the events of learning through the use of a dashboard which serves as a presentation layer to display important analytics insights. Lisa Lobry (2004) defined social, cognitive, and system responses that can be identified in the student postings. Scholars have argued that, theoretically, asynchronous discussion forums should be able to improve learning outcomes because of their unique technological affordances. As Allen et al. (2013) noted, because of the asynchronous nature of the technology, the very nature and method of discussions are different. “This means that students can think, edit, research, and post on a topic, even a couple of days after the original post”.

2.2 Topic Analysis in Online Discussion Forums (ODFs)

Topics are latent and embedded in the textual data. There are multiple methods for topic discovery, including the Hierarchical Dirichlet Process that was used by Ma et al. (2016) to model topic evolution in online news articles published by Reuters. Latent Dirichlet Allocation model used by Feng Jian et al. (2018) to automatically extract topics from different time slices and thereby extract the evolutionary relationship among sub-topics in Microblogs. Ezen-Can et al. (2015) used clustering techniques to group discussion topics. They proposed the k-medoids algorithm and defined each forum post as a data point. To determine the number of clusters, they rely on the Bayesian Information Criterion. Atapattu, Falkner et al. (2016) proposed the ideas of using topic-wise classification of discussion threads on MOOC discussion forums. This work facilitates the instructors to locate and navigate the most influential topic clusters as well as the discussions that require intervention by connecting the topics with the corresponding weekly lectures. Li et al. (2016) proposed keyGraph (Ohsawa, 1998), an advanced mining technique, that helps to assess the students’ knowledge discovery process and aids the instructors to create a new approach for transformative learning and teaching in education. Coffrin et al.

(2015) proposed visualization techniques to understand the patterns of the students' engagement in MOOC discussions. Similar work by Vytasek et al. (2017) used topic modeling approach to discover topics and sub-topics from the entire discussion forum. Our goal is similar but the forum is designed with more organized Q&A forums format. Further, we also emphasized on the topics evolutions and visualizations in our work..

In our project, we adopt similar techniques from Ezen-Can et al. (2015). We make use of the k-means clustering algorithm to group documents or contributions during the discussions based on cosine similarity and TF-IDF vector space. Each document is assigned to a single cluster, while documents from the time slice (in our experiment - weeks), can represent multiple clusters. Further, we propose techniques for extracting the sub-topics and discovering the evolution of the topics which is a new contribution to the research area of analysing discussion forums.

3. Background

The goal of text analytics is to extract high-quality information from collections of documents. Text Mining and Natural Language Processing techniques are useful for data processing and discovering useful patterns. Text Analytics techniques comprise of multidisciplinary fields like information retrieval, extraction, natural language processing, and text mining. Some of the issues that should be considered during text mining are tokenization, stop word list, lemmatization, etc. A brief description of the key components and techniques used in our solution follows.

Tokenization: Tokenization is a common text data pre-processing step that deals with the splitting of data into smaller units. These units can be paragraphs, sentences, phrases of n-grams (n number of words), and single words. Every dataset might have a different delimiter used to make the distinction between these units. Some common delimiters include commas, semicolons, tabs, new line characters and space.

Stop Word Removal: When doing text mining, many of the frequently used words in English are useless and add “noise” to the document. Words such as “could”, “and”, “if”, “the”, etc. that are classified as pronouns, conjunctions and prepositions, do not add value or carry information of importance to the model. Therefore, these words are referred to as stop words and are removed at the pre-processing stage.

Stemming: Stemming is the process of retrieving the stem or root form of a word in a heuristic approach, in the hopes of achieving the common base form or root form of the word. For the purpose of our application, words are stemmed using the Porter Stemming algorithm proposed by Porter (1980).

Lemmatization: Lemmatization although similar to stemming in that the ultimate goal is to retrieve the base form of a word, it is more complex than stemming because it requires Parts of Speech categorization of words before lemmatization to retrieve the lemma or the canonical form of the word by removing the inflectional suffix or prefix only. The lemmatizer used is based on the WordNet Database. In our preliminary experiments, we observed that lemmatization is the best pre-processing step for our results. Therefore, our focus will be on using the dataset cleaned with the help of the WordNet Lemmatizer (Christiane).

TF-IDF Document Representation: For any statistical computations on text data such as similarity scoring, a vector space representation of the text data is required. This representation consists of each document being evaluated as a term-frequency (TF) and inverse document frequency (IDF) weighted vector. TF-IDF is a statistical weight that ensures the rarer the term, the higher the weight of the score, by checking the frequency of occurrence of the term in a document (TF) as compared to how significant that term is with respect to the whole corpus or collection (IDF). Both these measures, in our solution, aid in generating the aspects or topics from the discussions. One way to combine a word's term frequency and inverse document frequency into a single weight is a TF-IDF. Each document in the dataset is then represented as a document-term matrix. For a term i in document j , the TF-IDF representation can be written as:

$$w_{i,j} = tf_{i,j} \times \log \frac{N}{df_i}$$

where $tf_{i,j}$ is the number of occurrences of i in j , df_i is the number of documents containing i , and N is the total number of documents.

Document similarity score: The similarity score between two documents determines the co-occurrence of a primary topic in the two documents. We compute this score by computing the cosine

angle between two documents which are modelled as vectors in a vector space suggested by Christopher Manning (2008).

K-means clustering: Clustering algorithms are exploratory data analysis tools that have proved to be essential for gaining valuable insights on various aspects and relationships of the underlying textual data. Clustering algorithms are used to find groups of similar objects in the data. The k-means clustering algorithm finds its clusters by initializing with k set of seeds to which each of the documents in the corpus are assigned (one to each) on the bases of closest similarity, in our case the cosine similarity. In the subsequent iterations, until convergence, the new seeds are defined by the cluster centroids of the current clusters derived (Aggarwal & Zhai, 2012).

Agglomerative Clustering: Agglomerative algorithms find the clusters by initially assigning each object to its own cluster and then repeatedly merging pairs of clusters until either the desired number of clusters has been obtained or all the objects have been merged into a single cluster leading to a complete agglomerative tree (Christopher Manning, 2008). The key step in these algorithms is the method, also referred to as clustering function, used to identify pairs of clusters to be merged iteratively.

4. Research Problem and Methodology

4.1 Research Problem

Recall that our motivation for this paper is to identify the sub-topics and the evolution of topics within the discussion forum. We study the two research questions.

RQ1: How the clustering technique performs in discovering sub-topics?

RQ2: Which visualizations are suitable for sub-topic and topic evolution representation?

4.2 Online Forum Settings

As part of the text analytics course for graduate students, the instructor designed a weekly question and answering forum. The information systems graduate courses are business-IT courses. We observed that the student were reluctant to use the discussion forum if it was a mere repetition of the course content. Therefore, we had to come up with a design, where questions were asked that promoted the student to do research and then participate in the forum. Moreover, in our experience, we also observed that the questions that related to topics beyond the class, were found to be more interesting and motivated the students to be active participants in the forum. Table 1 shows the discussion forum settings for the course. Note that the questions are a mix of business and technical aspects which align with the course objectives.

Table 1: Weekly topic and the related questions. Underlined phrases show the main topics.

Week	Discussion Forum Thread
0	<u>General discussions</u> General discussions including concepts, labs, class etc.
1	<u>Text Mining Introduction</u> What are applications of Text mining in education domain?
2	<u>Text pre-processing and NLP</u> How search engines (Bing or Google) use NLP? What are examples of applications of chatbots in different industries?
3	<u>Document Similarity</u> Explain the differences between the bag of words & vector space model.
4	<u>Text Classification</u> What are examples of text classification in the industry (Government, healthcare, banks, etc.)? What are various evaluation measures for text classification?
5	<u>Text Clustering</u> What are visuals for the displaying cluster results - Free draw and upload? Explain one clustering evaluation measure with an example.
6	<u>Information Extraction</u> What are applications of HMM models (Or any other Sequence Model)? What are examples of information Extraction in Industry (e.g. Finance, Retail, Travel,

	Healthcare, Media, Education etc.)
9	Sentiment analysis Discuss the technique to handle negation in opinions. Discuss technique to handle sarcasm in opinions. Discuss the technique to handle suggestions in opinions.

4.3 Participants

Out of the 55 students enrolled in the course, 37 students participated in the discussion forums. More than 50% of the students have past industry experience or were currently working in the industry.

5. Solution Design

The data from the discussion forums can be noisy and requires significant cleaning before we apply the mining techniques. To generate the topics, commonly used technique is LDA topic model. However, the limitation with this technique is the requirement of large datasets for better performance. Therefore, we took the clustering approach that also aids in the identifying the topics of discussions. With the several clustering techniques available from the machine learning research, we propose to study two popular models and evaluate them on their performance for discussion forum analysis. Finally, choosing visualizations is based on the previous research on the topic evolution and social network studies. Figure 1 depicts the solution overview showing the four stages; data processing, clustering, evaluations and visualizations.

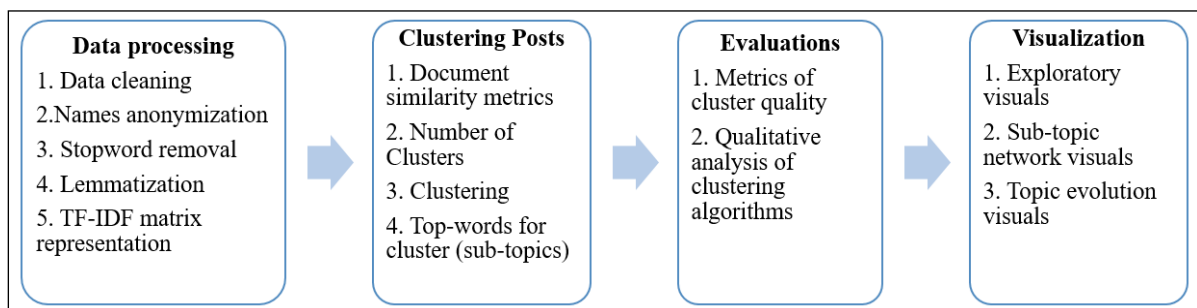


Figure 1: Solution model for topical analysis in discussion forums

Data processing: The first stage begins with the data preparation for representing each post as a matrix that becomes the input for the clustering algorithms. The posts consist of noisy data such as dates of posting, names, etc. We remove such data by regular expression method. Each discussion post is then represented as a TF-IDF matrix after the processing of stopwords removal and lemmatization.

Clustering Posts: We provide the choice of two clustering algorithms for the users; k-means and agglomerative clustering. Document similarity scores for k-means algorithm from scikit learn is based on Euclidean distance (Gavin, 2017). The objective function is to minimize the within cluster sum of squares between the documents and centroids. In case of agglomerative clustering, the tool supports l1 norm (cityblock distance) and l2 norm (Euclidean distance). In our preliminary analysis, the cosine similarity performance has not performed well. Hence, even though the user interface is developed with the choice of different metrics, in this paper, we focus on Euclidean distance metric. The challenge of choosing the best number of clusters can be estimated using the elbow method. For each k value, initialise k-means and use the inertia attribute to identify the sum of squared distances of samples to the nearest cluster centre. As k increases, the sum of squared distance tends to zero. Plot sum of squared distances for k in the range specified. If the plot looks like an arm, then the elbow on the arm is optimal k. The final part is labelling the clusters. Since, clustering is unsupervised learning, to label each cluster, we use the high frequent words to label the cluster. If the top words are coherent, qualitatively, the cluster is of better quality, indicating the performance of the clustering algorithm.

Evaluations: The quantitative analysis for clusters can be based on true labels or the non-human clustering performance metrics such as Calinski and Harabaz score (Yanchi, 2010). We also perform the qualitative analysis by comparing both the clustering methods based on the top words representing the clusters. The clusters with coherence and non-repetitive representative words are considered to be of high-quality clusters.

Visualizations: For exploratory analysis to study the statistics of student’s discussion, pie chart or bar chart are the best choice. To represent the topics and sub-topics, the network-based graphs are suitable (Aric et al. 2008). They are the advanced charts which are not only user-friendly but also suitable to represent the hierarchies and relationships. Finally, for the topic evolution visualization, we propose interactive line graphs, which are user-friendly charts with hovering feature.

6. Experiments and Findings

To examine our research questions from a more objective standpoint, we used exploratory analysis to study the general statistics on the student participations on various topics. For answering our RQ1, we conduct clustering evaluation as discussed in Section 5, and for answering RQ2, we develop visuals using python networkX and matplotlib packages (Aric et al. 2008).

6.1 Exploratory analysis results

Figure 2 shows the distribution of discussions by topics posted by the instructor. The figure shows the proportions of posts over the main topics described in Table 1.

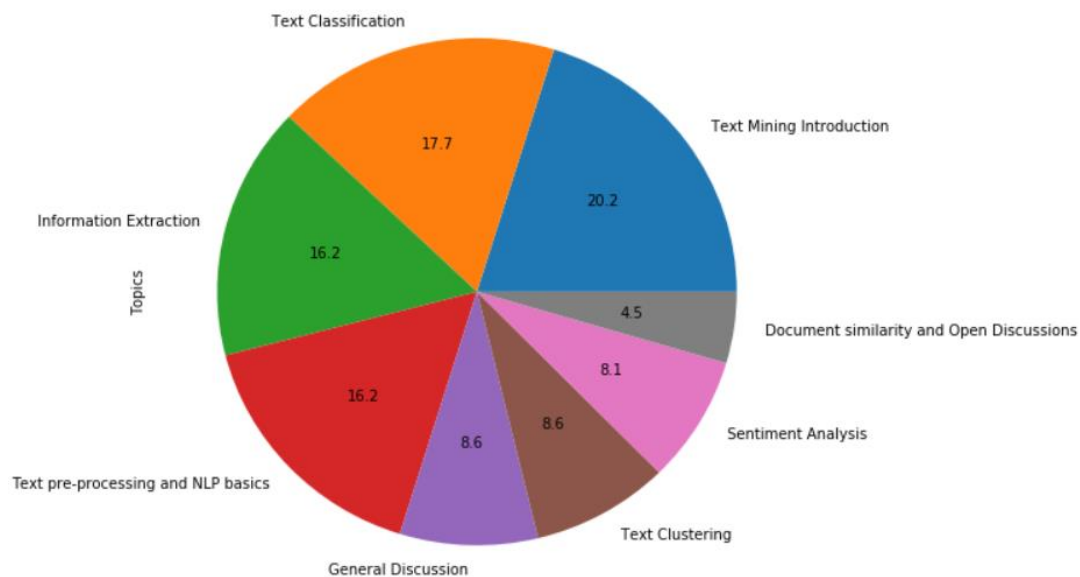


Figure 2: Overall participation in topics.

We observe that the first four topics have similar contributions more than 16% and the last four topics are around 8% except for the topic, “document similarity” which is lower than 5%. This is the most challenging technical topic during the early weeks of the course, and hence, the number of posts is lower. From the perspective of the type of questions, the different types of questions, namely understanding, analysis and discuss, received a similar number of posts from the students. The average number of words per participant is 676, which is quite high, thus indicating students’ interest in proving detailed explanations.

6.2 Clustering Evaluations

As described in our solution, the k-means clustering and agglomerative clustering are used to cluster the posts. Major limitations with k-means the selection of the number of clusters and randomness in the clusters. On the bright side, it can be implemented easily and on well segregated data, the clusters can be very coherent. To define the number of clusters, we employed the elbow method

with the k-means algorithm. The results show that the best number for clusters can be around 8 or 20. Since we are studying the sub-topics, we choose the bigger number, 20.

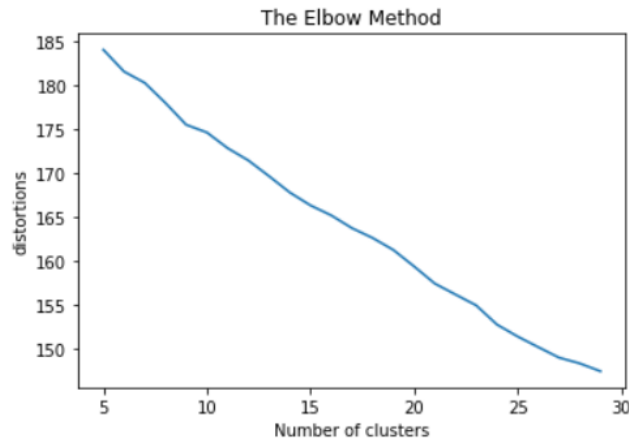


Figure 3: Elbow plot for k-means algorithm based on inertia.

Once the number of clusters is determined using the elbow method, we set the parameters for both clustering algorithms as below.

K-means settings: number of clusters is 20, initialization method is k-means++, number of time the k-means algorithm will be run with different centroid seeds is 20, and iterations is set to 500.

Agglomerative Cluster Top words in each cluster	Coherent/ Repetitive	K-means Top words in each cluster	Coherent/ Repetitive
['machine', 'tweet', 'comments', 'replying', 'google']	C	['machine', 'training', 'model', 'dataset', 'data']	C
['discuss', 'technique/s', 'opinions', 'handle', 'cluster']	C	['doubt', 'clarified', 'arun', 'comment', 'mislead']	R
['sentiment', 'negation', 'sarcasm', 'analysis', 'expressions']	C	['customer', 'chatbots', 'service', 'customers', 'chatbot']	C
['text', 'mining', 'classification', 'government', 'precision']	C	['doubt', 'clarified', 'arun', 'comment', 'mislead']	R
['in-class', 'question', 'thanks', 'students', 'compilation']	C	['sentiment', 'negation', 'analysis', 'scope', 'polarity']	C
['chatbots', 'customer', 'service', 'questions', 'banks']	C	['government', 'sites', 'classification', 'banks', 'examples']	R
['data', 'plagiarism', 'clinical', 'manually', 'website']	C	['patient', 'text', 'healthcare', 'doctors', 'analytics']	C
['doctors', 'symptoms', 'flu', 'doctor', 'analytics']	C	['words', 'slide', 'word', 'document', 'general']	C
['sequence', 'models', 'applications', 'state', 'model']	C	['doctor', 'medical', 'records', 'hospital', 'doctors']	C
['automated', 'model', 'learning', 'features', 'selection']	C	['patient', 'text', 'healthcare', 'doctors', 'analytics']	C
['doc', 'list', 'documents', 'docs', 'words']	C	['text', 'mining', 'students', 'education', 'analytics']	R
['words', 'training', 'dictionary', 'data', 'document']	C	['text', 'mining', 'students', 'education', 'analytics']	R
['extraction', 'finance', 'metadata', 'industry', 'used']	C	['text', 'mining', 'students', 'education', 'analytics']	R
['search', 'google', 'nlp', 'words', 'bing']	C	['google', 'search', 'nlp', 'engines', 'bing']	C
['news', 'readers', 'articles', 'organizations', 'travel']	C	['customer', 'chatbots', 'service', 'customers', 'chatbot']	R
['evaluation', 'clusters', 'measures', 'clustering', 'various']	C	['doc', 'documents', 'list', 'docs', 'w.lower']	C
['medical', 'records', 'problem', '']	C	['government', 'sites', 'classification', '']	R

'institutions', 'doctor']		'banks', 'examples']	
['feedback', 'increase', 'customer', 'teacher', 'sort']	C	['text', 'mining', 'students', 'education', 'analytics']	R
['chat', 'robots', 'chatbot', 'technology', 'customer']	C	['named', 'article', 'entity', 'recognition', 'articles']	C
['events', 'recognition', 'drug', 'speech', 'gait']	C	['discuss', 'sarcasm', 'opinions', 'technique/s', 'handle']	C

Agglomerative Settings: number of clusters is 20, metric used to compute the linkage is “l2”, and linkage criterion for distance calculation is set to “complete”, the maximum distance between all observations.

The representation for each cluster is the list of high-frequency words. To evaluate the clustering outcomes from k-means and agglomerative, we compare the clusters using qualitative analysis, as shown in Table 2. If the top words are coherent and non-repetitive, we label as ‘C’, and repetitive clusters are labeled as ‘R’.

Table 2: Qualitative analysis of agglomerative and k-means clustering. C represent coherently and R represents repetitive.

From the table, it is evident that agglomerative clustering has out performed k-means clustering. Hence, for final solution design evaluations, we use agglomerative clustering. This answers our first research question, RQ1.

6.3 Visualizations

Figure 4 shows the network graph for topics and sub-topics visualization. The topics are represented by yellow circles and the sub-topics are represented by red circle. The legend for the sub-topics is shown to the right. For the purpose of this paper, we are showing only 12 sub-topics instead of 20 for simplicity.

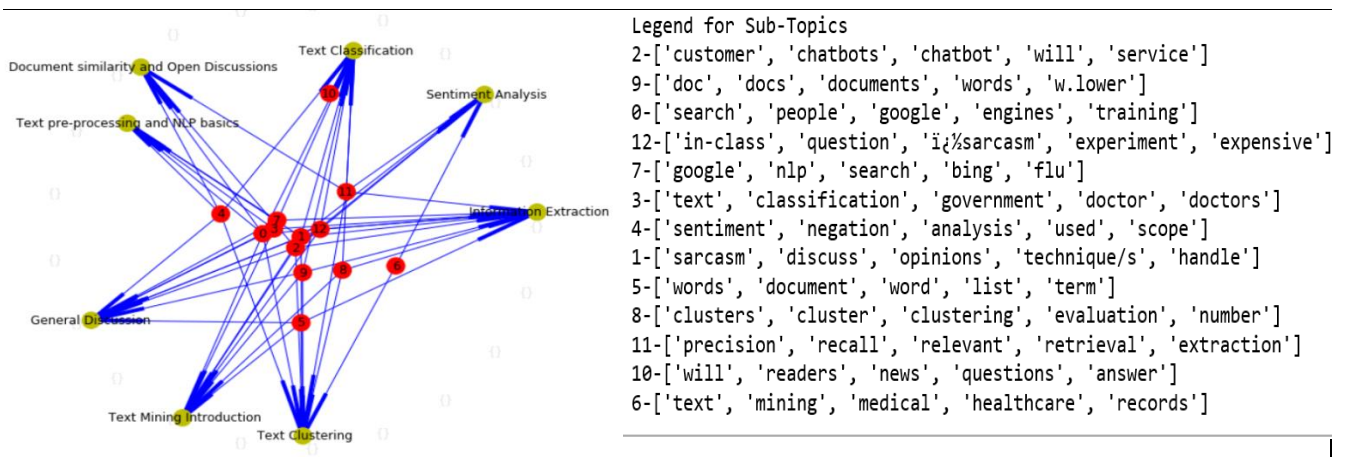


Figure 4: Topics and Sub-topics network graph. The legend is displayed to show the sub-topics, red circles.

From Figure 4 we observe that sub-topics can be part of more than one main topic. For example topic, 6 - “medical and healthcare”, appears under “text classification”, “text mining introduction”, and “clustering”. This shows how students are connecting the sub-topics over various topics via the discussion posts. From such graphs, the instructor can identify the missing sub-topics and submit the posts under the main topic to lead the students in the learning process.

Figure 5 shows the sub-topic evolution over time. It depicts topic evolution over the weeks, given the percentage makeup of each week. A threshold value of minimum percentage makeup can be set using the slider below the graph. The interactive nature of the graph also enables the users to study each topic in detail and aid the instructor to decide on the need for intervention if the student misses the sub-topics. This chart is based on the results from the 12 clusters shown in Figure 4.

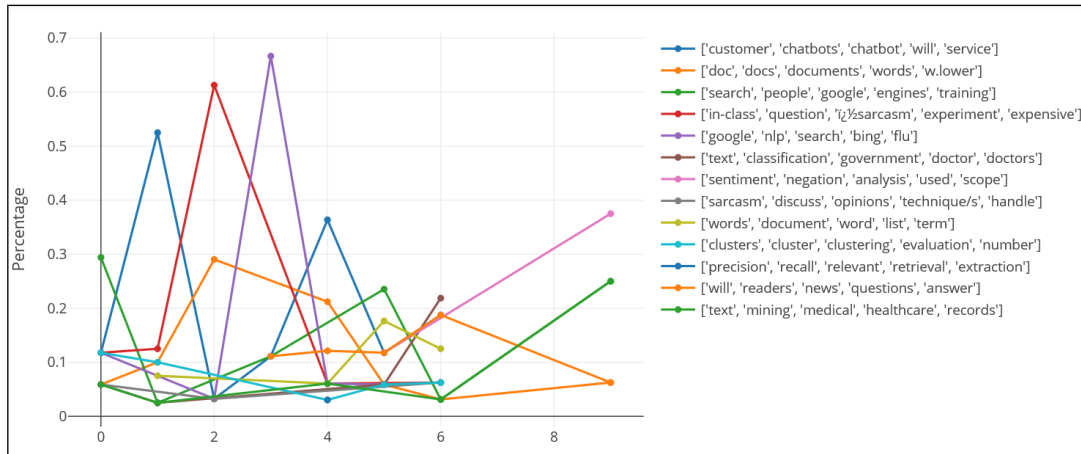


Figure 5: Sub-Topic Evolution over the weeks. Interactive graph with hovering features.

We observe some interesting patterns in this graph in terms of the short-lived vs repeated topics. Topics on “chatbot” and “healthcare” have occurred several times over the weeks. This is due to the examples that the students choose to apply the concepts in the given domains. The chart is also interactive that the selection of topics can be done and study the sub-topics comparison in details. This answers our RQ2 on types of visualizations for the topic of network analysis and evolution. We observed that standard non-interactive graphs can provide overwhelming information to the users where as the interative graphs enable the users to analyse the visuals and gain insights in effective manner.

6.4 Discussions

Our proposed solution worked well for the discussion forum within the chosen information systems graduate course. Our experiments show that agglomerative clustering model performs better than k-means clustering for IS technical courses. However, this may not be true for other courses. Therefore, our solution design provides the flexibility for the users to choose the algorithm, similarity techniques and a number of clusters. In this solution, we explored clustering techniques to discover the sub-topics. An interesting future work is to incorporate LDA models to extract the sub-topics by considering the tokenisation by sentences which can provide large data for LDA learning. It is a more suitable algorithm if the discussion posts tend to have multiple subtopics. Another interesting technique for clustering latent class analysis which can overcome some of the limitations of hierarchical clustering methods. The second interesting future work is the summarization based on the topics and sub-topics. We are currently working on the topic-based summarization models to generate automated summaries to the instructors that can be shared with the students.

7. Conclusion

In this paper, we propose a clustering based solution model for discovering sub-topics from the students’ discussions in an online asynchronous discussion forum. Our experiments show that agglomerative clustering model performs better than k-means clustering for IS technical courses. To generalize the solution model across other courses, we provide the flexibility to the users to choose the clustering settings. The second contribution is the discovery and visualization of topic-evolution over the time. The interactive model enables the users to study how students are connecting the previous topics over the period of the course and to what proportion of students post such topical connections every week of the course.

References

Allen, M., Omori, K., Burrell, N., Mabry, E., & Timmerman, E. (2013). Satisfaction with distance education. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 143–154). New York, NY: Routledge.

- Amidon E. J. (1968). Interaction Analysis. Theory Into Practice, Workshop in the Analysis of Teaching: 7:5, 159-167.
- Andy, C. (2013), Reports Forum Graph, Moodle. https://moodle.org/plugins/report_forumgraph
- Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart. (2008) "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), Gael Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008
- Atapattu, T., Falkner, K. and Tarmazdi, H. (2016). Topic-wise classification of MOOC discussions: A visual analytics approach. Proceedings of the 9th International Conference on Educational Data Mining (EDM), Raleigh, NC, USA.
- Burge, E. (1994). Learning in computer conferenced contexts: The learners' perspective. Journal of Distance Education, 9(1), 19-43.
- C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy. (2015). Visualizing patterns of student engagement and performance in MOOCs. Pages 83–92, New York, New York, USA, 2014. ACM
- Charu C. Aggarwal and Cheng Xiang Zhai. (2012). *Mining Text Data*. Springer Publishing Company, Incorporated. 2012
- Chaturvedi, Snigdha and Goldwasser, Dan and Daumé III, Hal. (2014). Predicting Instructor's Intervention in MOOC forums Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014
- Christiane Fellbaum (1998) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- De Bruyn, Lisa Lobry (2004) 'Monitoring online communication: can the development of convergence and social presence indicate an interactive learning environment?', Distance Education, 25: 1, 67 — 81
- Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 9-17). ACM.
- Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modelling for understanding MOOC discussion forums: a learning analytics approach. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 146-150).
- Gavin Hackeling. (2017). *Mastering Machine Learning with Scikit-Learn - Second Edition: Apply Effective Learning Algorithms to Real-World Problems Using Scikit-Learn* (2nd ed.). Packt Publishing.
- Jian, F., Yajiao, W., & Yuanyuan, D. (2018). Microblog topic evolution computing based on LDA algorithm. Open Physics, 16(1), 509-516.
- Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012). GLASS: a learning analytics visualization tool. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 162-163). ACM.
- Li, S. & Wong, G. 2016, Educational Data Mining using Chance Discovery from Discussion Board. In Proceedings of GCCCE'16. The Hong Kong University of Education (pp. 712-715).
- M.F. Porter. (1980) An algorithm for suffix stripping, Program, 14 (3):130-137, 1980
- Ma, T., Qu, D., & Ma, R. (2016). Online topic evolution modelling based on hierarchical Dirichlet Process. IEEE International Conference on Data Science in Cyberspace, pp. 400–405.
- Ohsawa, Y., Benson, N. E., & Yachida, M. (1998) KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on (pp. 12-18).
- Simon, Y. K. Li, and Gary K. W. Wong. (2016) Visualizing the asynchronous discussion forum data with topic detection. In *SIGGRAPH ASIA 2016 Symposium on Education: Talks* (SA '16). ACM, New York, NY, USA, Article 17, 3 pages. DOI: <https://doi.org/10.1145/2993363.2993367>
- Venky S, Swapna Gottipati, Ramaswami Seshan and Chirag Chhablani. (2018). Class Discussion Management and Analysis Application. In proceedings of 26th International Conference on Computers in Education (ICCE). 2018.
- Vytasek, J. M., Wise, A. F., & Woloshen, S. (2017). Topic models to support instructors in MOOC forums. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (pp. 610-611). ACM.
- Wragg, E.C. (1970). Interaction Analysis in the Foreign Language Classroom. The Modern Language Journal: 54:2, 116-120.
- Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. (2010) Understanding of Internal Clustering Validation Measures. In *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10)*. IEEE Computer Society, Washington, DC, USA, 911-916. DOI=<http://dx.doi.org/10.1109/ICDM.2010.35>

Teaching Analytics: A Multi-layer Analysis of Teacher Noticing to Support Teaching Practice

Alwyn Vwen Yen LEE^{a*}, Aik Ling TAN^b & Seng Chee TAN^c

^{abc}*National Institute of Education, Nanyang Technological University, Singapore*

*alwyn.lee@nie.edu.sg

Abstract: This paper, as part of a larger ongoing study, presents the use of a multi-layer approach to analyzing teacher noticing for the improvement of teaching practices. Situated in the field of teaching analytics, the use of multimodal sensors and analytics, especially for teacher noticing research, has provided affordances to discover deep insights for improving teaching practices. We collected data from a case study of one teacher over three lessons of science teaching in a secondary school. Multimodal sensors including an eye-tracking device, a microphone, and multiple video cameras were deployed in a classroom. The various sources of data were integrated and a multi-layer analysis was performed to uncover insights into the teaching practice. The findings show that a novice teacher in our case study was able to attend to events in her classroom, with some interpretations and sense-making of the events; some necessary actions were taken based on the teacher's analysis but in some instances, necessary action was found to be lacking. Prior knowledge and the wealth of experiences or the lack thereof, together with visual cues in the environment, can affect the decision of novice teachers in executing certain actions in a classroom.

Keywords: Teacher noticing, teaching analytics, multi-layer analysis, teaching practice

1. Introduction

Research for improving teaching and learning processes in science education are prevalent in various forms. For example, it can exist as interventions to improve teachers' practice (Kiemer, Gröschner, Pehmer, & Seidel, 2015), or in more recent times, developed as part of online tools (Kuosa et al., 2016) to enhance visualization and innovative equipment for acquiring data (e.g., eye-tracking device in eye movement research by Jarodzka, Holmqvist, & Gruber, 2017). These research help to gather insights into the improvement of teaching and learning, whether in face-to-face situations or within online learning environments. Although educational data mining and analytical algorithms are automating processes and aiding processes to be easily implemented, there is still a need to make learning activities visible, so that both teachers and students can develop skills in monitoring their own learning and see directly how invested efforts can improve learning (Bienkowski, Feng, & Means, 2012). The learning analytics approach has thus been applied to further these efforts in understanding teaching and learning processes, often through visualizations, methods, and digital tools that measure, collect, analyze and report data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (SOLAR, 2011).

Teaching-oriented learning analytics, also known as "Teaching Analytics", refers to an approach that supports teaching practice through the use of analytics (Vatrapu, Teplovs, Fujita, & Bull, 2011) and improve educational designs prior to the delivery (Sergis & Sampson, 2017). This form of analytics contains similar analytical processes that follow closely that of learning analytics processes, beginning with the gathering and analysis of data, followed by interventions to seek newer understanding of classroom situations. As a result, teaching analytics projects also tend to follow the theoretically-grounded Learning Analytics Cycle (Clow, 2012), which conceptualizes learning analytics work as four linked aspects: learners (aspect 1) generate data (aspect 2), that is used to produce metrics, analytics or visualizations (aspect 3), with key emphasis on 'closing the loop' by feeding back this product to learners through one or more interventions (aspect 4). However, as pointed out by Prieto, Sharma, Dillenbourg, and Jesús (2016), due to the abundance of online trace data generated on digital platforms and the ease of implementing triggers or interventions in an online environment, teaching

analytics have become mostly focused on online or blended learning situations and often lack the crucial step of supporting teaching interventions based on learning analytics insights. In face-to-face situations such as classroom-based instruction, classroom data is captured in the forms of video data, field observations and discourse data, which may be further analyzed to aid the crafting of interventions to support teaching. One main challenge is how can the identification of teaching practices during lessons be improved and the insights be applied to support teaching practices?

This paper presents a multi-layer approach to analyzing teacher noticing. We collected data, from a case study of one teacher over three lessons of science teaching with secondary school students, using multimodal sensors, including an eye-tracking device, a microphone, and multiple video cameras around a classroom. The various sources of data are integrated into a single data set before a multi-layer analysis is applied to generate insights that could be drawn from each level of the analytical lens.

2. Literature and Related Work

2.1 Teacher Noticing

Noticing is a natural part of human sense-making and can aid interpretation based on one's orientations and goals. Noticing entailed by teaching, thus also referred to as teacher noticing, is specialized to its purposes, such as noticing things that are central not to personal goals, but to professional ones (Sherin, Jacobs, & Philipp, 2011). Therein lies a crucial and often overlooked factor that has to be considered in studies that involve teacher noticing, which is the contrast between experienced and newly-trained observers in noticing and interpreting events that happened in classrooms. For example, a disruptive student in class might use unruly behavior to hide the insecurity of approaching a difficult subject, thus affecting the progress of the lesson. A novice teacher might be unable to handle the situation and choose to ignore the situation over time, but an experienced teacher will pick up on visual cues and question the student to provide further support.

Early research focused on how teacher noticing affects pre-service teachers' abilities to attend to student thinking and learning in lessons (e.g., Levin, Hammer, & Coffey, 2009), or on the use of evidence to support claims of teaching effectiveness (e.g., Santagata & Yeh, 2014). There are currently some but limited attention paid to the use of multimodal sensors to capture and provide pre-service, newly-trained, or novice teachers with more insights to support their teaching practice (Prieto et al, 2016). In general, literature (Jacobs, Lamb, & Philipp, 2010; Sherin et al., 2011) have shown that teacher noticing involves three processes, namely, 1) attending to particular events in an instructional setting, 2) interpretation and making sense of the events in the instructional setting, and 3) deciding necessary actions based on the analysis. In this paper, a wearable eye-tracking device was used to record, across multiple lessons of instruction, how a novice teacher attended to classroom events, made sense and interpreted the events, and eventually took action.

2.2 The Affordance of Technology and Eye-tracking Device to Understand Teacher Noticing

A large number of technological tools exist to aid teachers in analyzing student-related aspects, ranging from visualizations of student progress in activities (Phillips & Popovic, 2012) to devices that alert and guide teachers to students who need help (Alavi & Dillenbourg, 2012). Recent studies (Matuk, Linn, & Eylon, 2015) tapped on the features of technology to afford customizations that support individualized guidance to teachers and reduced the need for teachers to dynamically adapt to students. The use of eye-tracking technology is currently more pervasive due to advancements of technology and as more researchers take an interest in investigating teachers' professional vision (Wolff, Jarodzka, Bogert, & Boshuizen, 2016; Stickler, Smith, Shi, Caws, & Hamel, 2016). Although eye tracking is an invaluable tool to investigate the execution of professional tasks via eye movements (Jarodzka et al., 2017), our take on the use of eye-tracking technology coincides with what research in other fields such as medicine have reported (Kok & Jarodzka, 2017): that even though eye movements reflect cognitive processes, these cognitive processes cannot be directly inferred from eye-tracking data. Therefore, for us to interpret eye-tracking data properly and frame it within a teaching analytics context, this paper incorporated the use of multiple recording devices to collect data for analysis with eye-tracking data

using a multi-layer analysis. The interpretation of eye-tracking data is further supported by evidence obtained through reflective dialogues with teachers.

2.3 Multimodal Learning Analytics (MMLA)

There are opportunities and challenges in using multimodal learning analytics (MMLA; Ochoa, Worsley, Chiliza, & Luz, 2014; Worsley, 2016), a nascent subfield in learning analytics. As an emergent trend of capturing, modeling, and interpreting multimodal interactions that occur in both digital and face-to-face learning environments, a variety of data sources can be used to complement common digital analyses. However, given a large amount of multimodal data that could be obtained from various data sources, there remains a substantial amount of pre-analysis and conceptual considerations for data fusion. The variety of natural rich modalities may be redundant in some circumstances. In this study, a combination or fusion of modalities is designed to give more affordance to analysts and provide deeper insights into supporting teaching practice.

Prieto et al. (2016) cautioned that using multiple layers of analysis might not always provide better results, sometimes such methods only lead to marginal gains. In our case study, we consciously leverage different affordances of each modality so that newer or previously obscure insights may be discovered when MMLA is used with a multi-layer analysis to interpret and make sense of the data, based on the provided data in each layer of analysis.

3. Methods

3.1 Multimodal Data Collection

Our case study uses multimodal data from three sources — an eye-tracking device that is worn by the teacher during the lesson, a microphone attached close to the teacher's mouth, and two cameras around the classroom (placed at the front and back of the classroom) to capture activities within the classroom. To minimize the influence of cameras on the behavior and performance of both teachers and students, we oriented the teacher with the equipment and explained the study's non-assessable objectives before the start of the lesson. A pre-lesson briefing was then conducted by the teacher to inform students of the nature of the equipment and that the video recordings will neither impact nor be used to assess their behaviors during instruction. Participants have the right to not participate or withdraw from the study at any time and this action will not result in punitive measures or loss of benefits.

In this paper, the outputs from the various recording devices can be individually collected and analyzed as part of a layer of analysis that is shown in Table 1, or if needed, different combinations of multimodal data can also be brought together for analysis. Video footages from the two cameras placed around the classroom are collectively referred to as "video data". Audio data recorded through the microphone was transcribed and referred to as "discourse data". The eye-tracking device was accessed through a D-Lab software platform in order to download a "Point-of-View" (POV) video footage and a list of gaze point coordinates, which are collectively referred to as the "eye-tracking data".

3.2 Multi-layer Analysis and Measures

There are three layers of analysis. We named the layers of analysis as shown in Table 1, mainly for the ease of reference in this paper. The multi-layer analysis begins with a basic analysis that scrutinizes student activity in a classroom, based on what is observed visually from the two sources of video footage. This is followed by an intermediate analysis based on a combination of the video data and a transcript consisting mainly of teacher's instructions interjected with some student interactions. The third final layer of analysis is based on data from all modalities in the case study, including video data, a transcript of discourse, and eye-tracking data (POV video and a list of gaze points).

Table 1

Analytic components of each layer in the multi-layer analysis

Layers of Analyses (starting from top to bottom)	Analytic components in each layer		
Basic Layer	Video Data		
Intermediate Layer	Video Data	Discourse Data	
Final Layer	Video Data	Discourse Data	Eye-tracking Data

Similar to the processes that were mentioned in the literature, within each layer of analysis, there are three interrelated and cyclical processes that we examined in our case study. We focus on three aspects: 1) whether the teacher was attending to particular events in an instructional setting, 2) how the teacher interpret and make sense of the events in the instructional setting, and 3) take the necessary and appropriate actions based on the teacher's analysis.

To analyze the first aspect, we focus on the location that the teacher chose to look at, which was determined using the POV video footage (see Figure 1). The gaze point was plotted onto the video, using the list of coordinates obtained from the eye-tracking device.

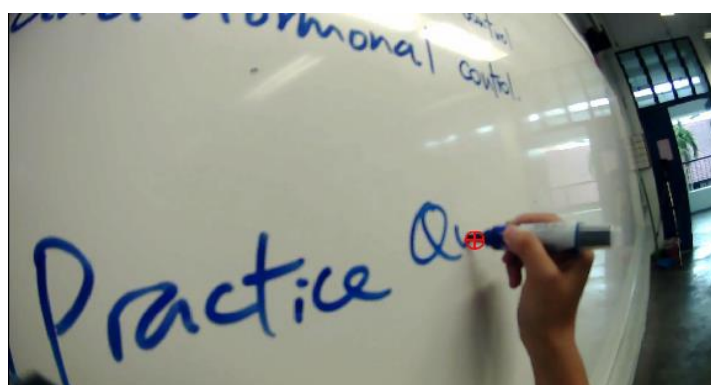


Figure 1. Screenshot from a Point-of-View (POV) video when the teacher was writing on a whiteboard. The gaze point is represented by a red crosshair-like symbol.

After the data were processed and transcribed, we engaged the teacher in reflective dialogues to understand the teacher's interpretation of events, how the teacher made sense of the events, and on actions that took place after the analysis has occurred. In the following subsections, we describe the processing and integration of different data sources in the various layers of analyses.

3.2.1 Using video data – Basic layer of analysis

Two video cameras were used to capture the teacher's practices and student activities during lessons in a classroom setting. The two cameras were placed at the front and back of the classroom to provide overlapping fields of vision and to ensure that most, if not all of the classroom activities, were captured for analysis. Based on the visual data, two researchers noted and concurred the general direction and any specific objects or subjects (students in this case study) that the teacher was looking at or interacting with during the period of analysis. In this layer of analysis, we are solely interested in the insights from video analysis of movements and physical positioning of the teacher. We also expect the subsequent intermediate layer of analysis to provide clearer insights into a teacher's practice and conversations with students.

3.2.2 Using video and discourse data – Intermediate layer of analysis

The audio conversations between the teacher and students were recorded using a microphone that was attached close to the teacher's mouth. The audio recording was then synchronized with the video data. The audio recording was transcribed and actions that were observed from the video data were integrated into the transcript, thus forming a more detailed description of the teacher's instructions and

interactions with the students in the classroom. This transcript was then used to interpret the teacher's actions.

3.2.3 Using video, discourse, and eye-tracking data – Final layer of analysis

Using the D-Lab software, we were able to record and export eye-tracking data, consisting of the POV video and a list of gaze points. These data allowed us to interpolate and determine areas that the teacher was looking at any point in time during the lesson, by focusing on significant gazes that hover around or fixate on a subject or object in the classroom environment for more than two seconds. Together with the other sources of data, we interpreted the teacher's noticing habits and interactions with the students, alongside any subsequent actions that took place.

3.2.4 Reflective dialogue for further understanding and validation

Recordings of what the teacher pay attention to during classroom instructions were used during reflective dialogues as concrete reference points for recall and discussion. We relied on Video-Stimulated Recall (VSR; Sturtz & Hessberg, 2012), that is, after the teacher had watched video-recorded segments of classroom instruction, we conducted a reflection discussion with the teacher. Through this method, we gathered insights into the interpretation and sense-making processes that the teacher demonstrated at the time of events in the instructional setting, along with the thought process on the necessary actions that were taken based on the teacher's analysis. Eventually, from our observations, we would also be able to delve deeper into the understanding of the unarticulated thinking and decision-making processes that effective teachers engage in while conducting instruction (Martinelle, 2018).

4. Case Study

4.1 Dataset and Settings

The data for our study was collected based on the teacher's availability and scheduled lessons in a secondary school. Overall, a total of three lessons (30-60 minutes) were collected, totaling over two hours' worth of multimodal data. The teacher is considered a beginning (novice) teacher in the local context, with less than three years of experience. Three classes of between 28 to 34 students attended each lesson in which the teacher wore a wearable eye-tracking device with an attached lapel microphone. A researcher helped to set up two cameras to capture students' activities from the front and back of the classroom. The recorded lessons began with the students greeting the teacher and ended when the teacher dismissed the class. All three lessons had a similar lesson plan to teach science and are mainly teacher-based instruction, interspersed with some group work and individual in-class assignments.

The teacher has a free reign on their actions with no restrictions other than to avoid touching the eye-tracking device in order to retain a high accuracy of measurements obtained through an initial calibration and setup process. An example of the teacher during the lesson is shown in Figure 2, with some identities whited out for anonymity.



Figure 2. Back view from one of the cameras capturing teacher-based instruction in a classroom. The teacher was wearing an eye-tracking device that drew power from a backpack.

4.2 Choice of Events for Analysis

Considering the amount of multimodal data collected, we initially browsed through the collective video data to identify several events that were interesting for investigations and might contain some insights that are only discernible with the multi-layer analysis. In this paper, we presented two such events that stood out and we further probed the teacher during the reflective dialogue for understanding and validation of the events and actions that were revealed in the video data.

4.3 Findings and Discussion

In this section, we first present the summarized findings from our multi-layer analyses in the form of a table, followed by our interpretations from each layer of analysis. Insights obtained from the reflective dialogue were then used to corroborate our findings. The first event occurred at the early stages of a lesson when the teacher was awaiting students to settle into the classroom.

Table 2

Findings from a multi-layer analysis of the first event

Layer of Analysis	Analyses and Interpretations
Basic Layer – Video Analysis	<ul style="list-style-type: none"> The teacher monitored students entering the classroom and gesticulated at an incoming student (Student C) who stopped at the entrance, seemingly due to a seating issue. Student C proceeded to sit at an empty space and the teacher proceeded with instructions involving content presented on a screen.
Intermediate Layer – Video and discourse analyses	<ul style="list-style-type: none"> The teacher was apparently checking with Student C, who was distracted by another Student J who was sitting at Student C’s original seat. The teacher then probed the reason for the displacement of Student C and the matter was settled amicably when Student C proceeded to sit at a different seat. Discourse analysis showed that the teacher first emphasized the lack of time and was in a hurry to start her lesson immediately before she observed the incoming Student C. The students involved in the debacle were left to their own devices to settle the seating issue as the teacher did not provide any instructions to resolve the issue or offer any solutions to the students. Student C eventually decided to sit at an empty seat to end the stalemate.
Final Layer – Video, discourse, and eye-tracking analyses	<ul style="list-style-type: none"> In addition to previous analyses and observations, when students were entering the classroom, the teacher’s gaze was on the incoming students, but it stopped and shifted to Student C, who suddenly stopped at the doorway. The teacher’s gaze then tracked the direction of the seat that Student C was looking at, before focusing back to Student C. The

problem was then apparent to the teacher, as she asked Student C “who took your seat” and proceeded to scan through the neighboring students to check on their responses to the question. The teacher then recalled from previous experiences that another Student J has taken Student C’s seat and queried Student J. During this interaction, the teacher’s gaze was alternating between Student C and Student J.

- After getting a vague clarification from Student J, Student C had already settled in a new seat and the teacher saw there were no remaining students who were standing and decided to end the issue. The teacher momentarily paused for a moment with her gaze fixated on a blank space between students, before proceeding to explain the content on the screen.
-

From the multi-layer analyses in Table 2, we noted that the teacher attempted to get students to settle into the class before she could begin covering the required content within a limited amount of time. During the reflective dialogue, we showed the teacher a video snippet of the event and asked her for the most significant thing that she can recall about what happened. The teacher replied that she was surprised that Student J and her friend beside her were sitting in Student’s C seat in the middle of the classroom, rather than at the side of the classroom. This event unsettled her slightly and she mentioned that Student J and her friend were known to be unenthusiastic students, and as such, she should have focused attention more on them. Therefore, from this event, it was apparent that the teacher was swift in detecting visual cues within the classroom, but have difficulties handling unexpected students’ behaviors. Although she was undeterred by the turn of events that she did not previously encounter, the teacher’s composure was affected before the lesson even started. The reflective dialogue with the teacher also revealed that there was prior knowledge of certain students that might have caused inaction on the teacher’s part. The next event took place in the middle of a lesson when the teacher was instructing using content from a presentation slide.

Table 3

Findings from a multi-layer analysis of the second event

Layer of Analysis	Analyses and Interpretations
Basic Layer – Video Analysis	<ul style="list-style-type: none"> • The teacher was instructing using presentation slides, with frequent head turning movements at the slide’s content and gesticulating to better explain a structure found within a leaf. She gave attention to the students seated at the front by looking at them, but there was no further engagement with any student. Several front row students were staring blankly with heads resting on their hands and yawning at several instances. The students’ general behaviors reflect disinterest and boredom during this part of the lesson.
Intermediate Layer – Video and discourse analyses	<ul style="list-style-type: none"> • The teacher was explaining the concept of photosynthetic cells in a plant, using vertical hand gestures to simulate the efficient packing of palisade mesophyll cells. • Discourse analysis also showed that there were multiple instances where the phrase “maximize the rate of photosynthesis” was overly emphasized and constantly appeared in a sequence of sentences. • Similar to the first event, the teacher was rushing through the content and therefore, gave minimal wait times after questions were posed for students to think, resulting in students giving garbled or no response. This also resulted in some students preferring to copy content from the presentation slides onto their textbooks rather than pay attention to the teacher. The teacher noticed this trend and admonished the students, asking the students not to copy, as she would provide notes at a later time, but students would still need to fill in the blanks in the notes.
Final Layer – Video,	<ul style="list-style-type: none"> • When explaining the content from the slides, the teacher was making

discourse, and eye-tracking analyses

several glances from the students and back to the screen to refer to some technical terms, a possible indication that she was unsure of the instructed content.

- The teacher repeatedly emphasized “maximum rate of photosynthesis”, and her glance was not on any object, but at a space right above the students’ heads that hovered for a significant period of time.
 - There was a lack of engagement and eye contact with students during this event, possibly leading students to feel bored and uninterested, with some students preferring to copy notes into their textbooks instead of listening to the teacher.
-

The multi-layer analyses in Table 3 show that when the teacher was teaching, she was also having issues in engaging and maintaining this engagement with the students. During the reflective dialogue, we showed the teacher a video snippet of this event and similarly asked her for the most significant thing that she can recall about the event. The teacher explained that during instruction, she was often thinking while speaking and did this quite often during lessons, as she tried to find the appropriate words to use during the lessons, followed by consideration of necessary actions to take. As such, she often did not look at the students and hence had little situational awareness of students and their activities while she was speaking. This piece of information corroborates with the evidence in our study that shows the teacher seemed to be distracted during the instruction, as she gazed at blank spaces above students’ heads while instructing and thinking. As the teacher was also thinking while instructing, she might have not been paying attention to her own talk as she repeatedly emphasized a series of phrases, with little recollection of what she just spoke.

With this in mind, the teacher further added that she could hardly observe or interpret the level of students’ understanding during lessons, or whether the students were engaged with the topic at hand. This means that the teacher most probably did not notice the students’ general level of disinterest and boredom during lessons. She also acknowledged that she found it challenging to process her observations and was unable to think sufficiently and react in a timely manner to unexpected events, which might have led to inaction as observed in the first event (seating issue). Surprisingly, the teacher mentioned that she noticed this trait in herself when she was supervising the hockey team as well.

We interpreted the teacher’s response as a way of how the teacher tried to identify her areas for improvement in teaching and to circumvent her limitations to adapt to the situation. The use of multimodal sensors and multi-layer analysis helped to identify certain behaviors of concern during teaching, which the teacher was then able to relate to certain shortcomings in her teaching practice. After the reflective dialogue, the teacher was able to seek help from educational specialists in modifying her teaching practice to suit her needs. The lecture method of instruction chosen by the teacher was a problem for her as she was unable to process information and think sufficiently fast on her feet. It was then suggested that she could explore choosing strategies that depended more on students working in groups to learn. Using this strategy, the teacher would do less talking and she can then focus on making sense of students’ learning and interactions.

From the various analyses using our multi-layer approach, the findings from the two events helped to affirm that prior knowledge and experiences or the lack thereof, and visual cues in the environment could affect the decision of novice teachers in executing certain actions in a classroom, similar to learners in other studies (e.g., Luo, Koszalka, & Zuo, 2016). In the first event, the appearance of a visual cue prompted her to stop her current activity and shift her gaze to investigate the cued event, much like how visual cues are an effective tool for grabbing attention (Jamet, Gavota, & Quaireau, 2008). Such swift responses can, however, be detrimental for the teacher when dealing with constant distractions in a classroom. It might be advisable for her as a novice teacher to learn from expert teachers, as shown in other studies (Carter, Cushing, Sabers, Stein, & Berliner, 1988), to weigh the importance and form connections between pieces of information, so as to better represent management and instructional situations as meaningful problems that can eventually be solved.

Overall, our multi-layer analysis can provide users with combinations of analyses that provide different types of information and different levels of details for improvement. From the various layers of analyses, users can expect to obtain a certain fidelity of data insights. A summary of the insights that a teacher might obtain from the respective layers of the multi-layer analysis is presented in Table 4.

Table 4

A summary of insights from the respective layers of analysis

Layer of Analysis	Insights
Basic Layer – Video Analysis	Video analysis can provide a basic understanding of events and actions that took place in the video, but without specific attention to detail.
Intermediate Layer – Video and discourse analyses	The additional discourse analysis provides more details regarding the events, follow-up actions and possible reasons for doing so. There are clearer indications from the discourse that signal the intention of actions that the teacher may or may not take.
Final Layer – Video, discourse, and eye-tracking analyses	Eye-tracking data can provide clearer statistical and visual indications, such as the number of glances and noticing patterns, of why certain actions took place and if any preceding factors (such as subjects or objects) were considered before eventual actions were taken by the teacher.

5. Conclusion

This study is part of a larger ongoing study that investigates the effects of using a multi-layer analysis for explaining and improving teaching practice. Situated in the field of teaching analytics, the use of multimodal sensors and analytics, especially for teacher noticing, has provided affordances to discover insights for supporting teaching practice.

In essence, we noted the teacher attending to particular events, with some interpretations and sense-making of the events, followed by the taking or lack of necessary actions that were taken based on the teacher's analysis. Our basic analysis was able to provide an elementary understanding of the setting and actions that occurred during the event, while the discourse analysis in the intermediate analysis provided more details on possible reasons as to why certain actions took place, with some insights into the teacher's possible intention in conducting follow-up actions. In the final analysis, the teacher's intent in making certain actions were noticeably clearer, as we noted in the first event, where the teacher noticed visual cues in a logical and sequential manner (such as following Student C's gaze to the seat and scanning of neighboring students) that allowed her to better decide the best course of response. In the second event, the multi-layer analysis provided a similar sequence of explanations that identified the teacher's need to change her teaching practice in order to adapt to her students' needs.

Moving forward, we are in the midst of recording data from teachers with different experiences and will apply this multi-layer approach to discover deep insights for improving teaching practices.

Acknowledgments

The authors gratefully acknowledge the support and research funded by the Office of Educational Research, National Institute of Education, Singapore (project AFR 02/17 TSC). The university's ethics clearance (IRB-2018-01-030) has been obtained. The research team would also like to thank the teachers and students who participated in this study.

References

- Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, 5(3), 264-274.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. U.S. Department of Education, Office of Educational Technology.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of teacher education*, 39(3), 25-31.

- Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134–138). ACM.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International journal of science education*, 25(6), 671-688.
- Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for research in mathematics education*, 41(2), 169-202.
- Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and instruction*, 18(2), 135-145.
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of eye movement research*, 10(1), 1-18.
- Kiemer, K., Gröschner, A., Pehmer, A. K., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and instruction*, 35, 94-103.
- Kok, E. M., & Jarodzka, H. (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114-122.
- Kuosa, K., Distante, D., Tervakari, A., Cerulo, L., Fernández, A., Koro, J., & Kailanto, M. (2016). Interactive visualization tools to improve learning and teaching in online learning environments. *International journal of distance education technologies (IJDET)*, 14(1), 1-21.
- Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education*, 60(2), 142-154.
- Luo, H., Koszalka, T., & Zuo, M. (2016, July). Investigating the Effects of Visual Cues in Multimedia Instruction Using Eye Tracking. In *International Conference on Blended Learning* (pp. 63-72). Springer, Cham.
- Martinelle, R. (2018). Video-stimulated recall: aiding teacher practice. *Educational Leadership*, 76(3), 55-70.
- Matuk, C. F., Linn, M. C., & Eylon, B. S. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional science*, 43(2), 229-257.
- Ochoa, X., Worsley, M., Chiluiza, K., & Luz, S. (2014). Mla'14: Third multimodal learning analytics workshop and grand challenges. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 531-532). ACM.
- Phillips, V., & Popović, Z. (2012). More than child's play: Games have potential learning and assessment tools. *Phi Delta Kappan*, 94(2), 26-30.
- Prieto, L. P., Sharma, K., Dillenbourg, P., & Jesús, M. (2016). Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 148-157). ACM.
- Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video-and practice-based approach. *Journal of Mathematics Teacher Education*, 17(6), 491-514.
- Sergis, S., & Sampson, D. G. (2017). Teaching and learning analytics to support teacher inquiry: A systematic literature review. In *Learning analytics: Fundamentals, applications, and trends* (pp. 25-63). Springer, Cham.
- Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge.
- Society for Learning Analytics Research (SOLAR). (2011). In *Proceedings of the first international conference on learning analytics and knowledge*. Banff, Alberta, Canada.
- Stickler, U., Smith, B., Shi, L., Caws, C., & Hamel, M. J. (2016). Using eye-tracking technology to explore online learner interactions. *Language-learner computer interactions: Theory, methodology and CALL applications*, 163-186.
- Sturtz, J., & Hessberg, K. (2012). Examining teacher development: The role of teacher thinking, observation, and reflection. *Contemporary social studies: An essential reader*, 547-563.
- Vatrapu, R., Teplov, C., Fujita, N., & Bull, S. (2011). Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In *Proceedings of the first international conference on learning analytics and knowledge* (pp. 93-98). ACM.
- Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. (2016). Teacher vision: expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, 44(3), 243-265.
- Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. In *12th International Conference of the Learning Sciences: Transforming Learning, Empowering Learners, ICLS 2016* (pp. 1346-1349). International Society of the Learning Sciences (ISLS).

Do my students understand? Automated identification of doubts from informal reflections

Siaw Ling LO*, Kar Way TAN & Eng Lieh OUH

School of Information Systems, Singapore Management University, Singapore

*slllo@smu.edu.sg

Abstract: Traditionally, teaching is usually one directional where the instructor imparts knowledge and there is minimal interaction between learners and instructor. With the focus on learner-centered pedagogy, it can be a challenge to provide timely and relevant guidance to individual learners according to their levels of understanding. One of the options available is to collect reflections from learners after each lesson to extract relevant feedback so that doubts or questions can be addressed in a timely manner. In this paper, we derived an approach to automate the identification of doubts from students' informal reflections through features analysis, word representation and machine learning. Using reflections as a feedback mechanism and aligning it to the weekly course content can pave the way to a promising approach for learner-centered teaching and personalized learning.

Keywords: Doubt identification, learner-centered pedagogy, text analytics

1. Introduction

One of the main goals of an instructor is to ensure that most students, if not all, are able to understand and articulate the key concepts of each lesson clearly. It is challenging to verify that students understand the class materials in an informal way rather than an assessment. An approach is to have students write a reflection after each lesson. Since the reflection is a free-form text response, it is not uncommon to find combinations of articulation of learning points, questions and statements reflecting doubts related to the topic of the week. For example, students may include the following phrases, "I am still confuse about...", "It would be good if you can go through [a topic] again", "I am quite unsure when [an example] is a sample or a population".

We define doubt as a statement, which can potentially be a question or simply a statement that requires more clarification of a given topic. A doubt can be different from a question since it may not be expressed in the form of 5W1H (who, what, where, when, which, how) or end with a question mark. With the amount of informal free text collected from each class, it is essential to find an automated method to effectively extract questions or doubts so that the key concepts can be clarified in a timely manner.

In this study, we explored the effectiveness of feature analysis, word representation and machine learning approaches (such as text classification and sentiment analysis) in identifying doubts expressed in free-form students' reflection. We collected reflection data from two classes taught by two different instructors. Instructor 1 had two modes of data collection: free-form reflections and a self-assessed level of understanding captured in a Likert scale rating of 1-5 (with 1 being the lowest and 5, the highest). Instructor 2 collected only free-form reflections.

This study consisted of two parts. The first part focused on mining doubts from free-form students' reflections by automatically identifying doubts based on analyzing various text features and their role in constructing a machine learning model. Essentially, this is a binary classification problem, which differentiates reflections that contain and do not contain doubts, through assessment of suitable features and word representations. Since doubts can potentially be a question, question patterns were identified as one group of important features. Another feature was the sentiment of the reflections since sentiment analysis has been used extensively in feedback analysis studies (Dhanalakshmi, Bino, &

Saravanan, 2016; Gottipati, Shankararaman, & Gan, 2017). Thus, it is of interest to investigate if the sentiment of students' reflection plays a role in identifying doubt. The second part of the study evaluated the possibility of using the self-assessed level of understanding Likert scale rating to identify students who may need more help and also to assist in doubt detection.

To evaluate the effectiveness of our model, our selected top models trained using the data of Instructor 1 were applied to the set of reflections collected by Instructor 2. This second data set was collected without the corresponding self-assessed rating. The purpose of this verification was to evaluate the performance of applying our model to another qualitative data set since it is a common practice for instructors to collect qualitative feedback after a lesson or course without an accompanying rating scale. Our results showed that our proposed model can successfully extract doubts from students' reflections and interestingly, most of the contents did not contain the standard question patterns such as question mark or 5W1H. The findings exemplify the need to differentiate doubts from questions. We believe that identification of doubt contributes towards providing relevant feedback in a learner-centered environment that tailors to the needs of each individual learner.

The contributions of our work can be summarised in three folds. Firstly, we propose an automated doubt identification model using neural embedding word representation that shows promising results based on real-world student reflections. Secondly, we observed from our results that the model built with features inclusive of positive sentiment performed the best. Specifically, the model was constructed using reflections annotated with positive sentiment without doubts and reflections labelled as doubt. However, sentiment analysis alone was not sufficient to detect and identify doubts. Finally, our analysis suggests that self-assessed ratings correlate with doubts and the rating can be used in a learner-centered pedagogy to extract reflections and contents that can be useful to instructors.

In the next section, we will discuss some related work in detecting questions and feedback analysis. This is followed by the scope of data, methods used, our findings and results in Sections 3, 4 and 5, respectively. In Section 6, we discuss our observations of the findings and future plans before conclusions are drawn in Section 7.

2. Background and Related Work

2.1 Detecting Questions in Online Content

Question identification/detection serves many purposes but is very challenging for online content. Online questions are usually long and informal, and standard features such as question mark or 5W1H words are likely to be absent. Both rule-based and learning-based are common approaches to address this challenge. Rule-based approach such as the paper proposed by Efron and Winget (2010) designed several rules from heuristics or observations to check whether a tweet is a question or not. Learning-based approach proposed by Cong, Wang, Lin, Song, and Sun (2008) involves sequential patterns-based classification method to detect questions in a forum thread, and a graph-based propagation method to detect answers for questions in the same thread. Another learning-based approach explored question characteristics in community question answering services, and proposed an automated approach to detect question sentences based on lexical and syntactic features (Wang & Chua, 2010; Efron & Winget, 2010). Since this study focused on identifying doubts and its relevant features that can impact the accuracy of a model, the features identified by the rule-based approach proposed by Efron and Winget (2010) was adopted to assess whether common questions lexical and syntactic features (e.g., question mark, 5W1H, question patterns) can be used to identify doubts in reflections.

2.2 Analysis of Student Feedback

Analyzing student feedback can help to improve student's learning experience. A large part of feedback comes in the form of textual comments which pose a challenge in quantifying and deriving insights. Gottipati et al. (2017) have presented a conceptual framework for student feedback analysis that provides the necessary structure for implementing a prototype tool for mining student comments and highlights the method to extract the relevant topics, sentiments and suggestions from student feedback. Shankararaman, Gottipati, Lin, and Gan (2017) have provided an automated solution for analyzing

comments, specifically extracting implicit suggestions which are expressed as wishes or improvements from the students' qualitative feedback. Dhanalakshmi et al. (2016) have explored opinion mining using supervised learning algorithms to find the polarity of the student feedback based on predefined features of teaching and learning. Opinion mining, especially in the aspect of sentiment analysis and polarity study, has been the cornerstone of student feedback analysis and thus, it was of interest to explore if sentiment analysis played an important role in identifying doubts from the weekly reflections. However, we would like to highlight that this study is different from the common end-of-term student feedback analysis. The end-of-term student feedback focuses on providing qualitative analysis on the course delivery or the instructor while this study works on weekly reflections and feedback so that timely clarification of key concepts can be offered to students before introducing new knowledge or concepts.

3. Scope

3.1 Data Description and Collection

The data for this study was the course: Analytics Foundation, at the Singapore Management University. This is an undergraduate level foundation class involving data analytics for students from various disciplines (Business, Accountancy, Social Science, Economics, Law and Computing). The course required students to understand the algorithms underlying machine learning models, tune the parameters and apply the models to various problem contexts. Since the students had varied backgrounds, they encountered different challenges in understanding algorithms and the application of the learning models.

The course was conducted in a seminar-style learning environment with about 45 students in each class, over 3 hours of class engagement per week. Two course instructors collected weekly reflections as part of students' learning journal. Reflections collected by Instructor 1 consisted of two elements while Instructor 2 collected only one part. The details of the reflections data can be found in Table 1. The intent of the reflections was to provide weekly feedback to the instructors on the level of understanding in each class. Reflections were collected across 10 instructional weeks over a 16-week semester.

The reflection data collected under Instructor 1 served as the core data for our study. This dataset consisted of both free-form text and self-assessed rating. The free-form text from Instructor 2 (covering only one week of data) was used as a verification.

Table 7. Details of reflection data collected

	Instructor 1	Instructor 2
Part 1	Free-form text for entering the key learning points from each lesson.	Free-form text for entering constructive feedback to improve the lesson.
Part 2	An objective question for students to self-assess their level of understanding. Five levels of understanding in the form of Likert scale from 1-5 was collected: 1. Disagree – I am totally lost and I need a consultation 2. Somewhat disagree – I am lost (and I am paying attention) but I'll catch up on my own 3. Not applicable- I was not in class for a valid reason 4. Agree – Understand quite a lot 5. Strongly Agree – Understand almost fully	NIL
Number of students	44 [from one class] (Total reflections collected: 375)	89 [from two classes] (Total reflections collected: 71)

Weekly reflections were collected from students after each lesson and students were encouraged to provide constructive free-form feedback comprising specific learning points for that lesson. The students had up to 5 days to complete the reflections. Some students completed the reflection immediately after class while some students preferred to revise the course content before completing the reflections.

4. Automated Doubt Identification Approach

The focus of this research was to establish an approach to provide timely feedback to the students based on their reflections. Our automated doubt identification approach analyzed individual reflections to extract questions and doubts, thus providing a means to make decisions on the course of action for learner-centered learning. For example, a student whose reflection was identified to contain doubt may receive additional guidance on the given topic. A summary of our approach is depicted in Figure 1.

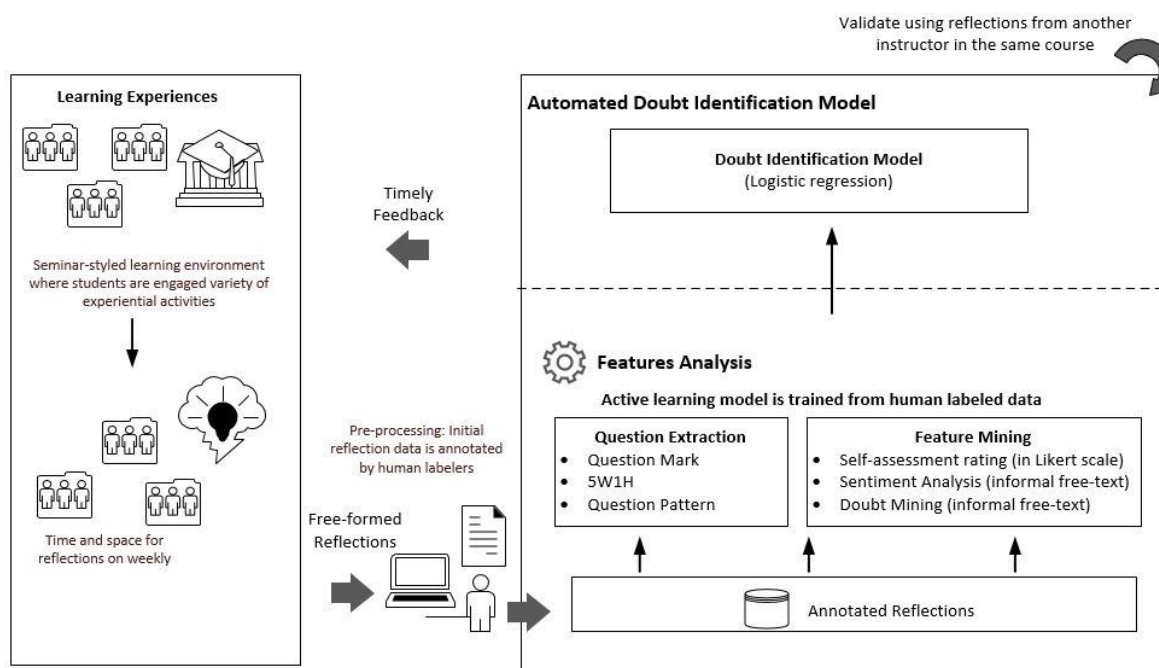


Figure 22 Automated Doubt Identification Approach

The reflections were collected from students in the same course across classes run by two different instructors. To understand if the students learnt from the experiential activities in a seminar-styled learning environment, reflections were collected at the end of each class (during or after the class via a survey). The data from various classes was combined and anonymized before the pre-processing stage. In the pre-processing stage, the initial reflection data was annotated by two human labelers based on presence and absence of doubt and different sentiments. Feature analysis was done on the annotated data, which included two components, namely Question Extraction and Feature Mining. The various features constructed were then assessed and used as the input of a Logistic Regression (LR) model. The LR based automated doubt identification model was validated with a second set of data from Instructor 2 so as to evaluate if the model can be applied to reflections of other students within the same course. Using our automated doubt identification model, instructors can take appropriate decisions and actions, such as timely feedback to help learners who may require more attention and devise additional activities for students who may be more advanced. The details of the various components are described in the following sections.

4.1 Data Preparation and Annotation

Since reflections are usually informal, it was important to clean the data prior to any data analysis. Each reflection was pre-processed to lower case with numbers and stop words removed. However, punctuation removal was selective because question mark was used as a feature in extracting potential questions.

In order to investigate the effect of using sentiment analysis to identify doubts raised by students, the data was annotated in two exercises. The first annotation exercise focused on identifying doubts by labelling via a 'yes' and 'no' label. The second was the annotation of sentiment with 'positive', 'negative' and 'neutral' labels. Since reflections can be an objective expression of the lesson learnt, neutral sentiment was commonly found in the data. In order to ensure the consistency and quality of the annotation, the following questions were derived to assist in determining the Doubt and Sentiment annotations.

For Doubt annotation, a reflection was labelled as 'yes' if one of the following conditions was fulfilled, otherwise, it would be 'no':

- Does the reflection ask for clarification on any of the topics?
- Does the reflection ask for additional information not previously covered in class?

For Sentiment annotation, each reflection was labelled based on the sentiment identified from one of the most relevant questions below. If none of the positive or negative sentiment was found, the reflection was annotated as 'neutral'.

- Does the reflection make remarks about positive/negative teaching (e.g., pace)?
- Does the reflection express any positive/negative feedback about the instructor (e.g., clarity)?
- Does the reflection express any positive/negative sentiment on the topic for the week (e.g., manageable, difficult to grasp)

Two annotators who were familiar with the reflection process annotated individually on the data. Additional review was done on all the records with inconsistent label before finalizing the two sets of annotations. These two sets of annotated data were used as ground-truth datasets in assessing the performance of the doubt identification models.

4.2 Question Features Extraction

Next, we evaluated if the common rule-based lexical and syntactic question features could be used to build an automated doubt identification model. The following features were considered:

- i. Question mark (QM)
- ii. 5W1H method
- iii. Rule-based question patterns (QP) (Efron & Winget, 2010)

The general rule of using QM and 5W1H is to detect questions by finding question marks at the end of sentences and 5W1H at the beginning of the sentence. However, we observed that this rule was not necessarily applicable in this study due to the informal nature of the reflections. Students used free-form expressions and questions which may not adhere to the format of a standard sentence or question structure. Thus, each QM and 5W1H was used as an individual feature regardless of their position in the sentence. On the other hand, the question patterns proposed by Efron & Winget, (2010) were expanded into phrases and each phrase became a feature, e.g., "I try to find" or "need to know". These phrases were generated based on the pattern "(pronoun)* [try, like, need] to [find, know]" where * sign is a wildcard, signaling 0 or more instances and verbs in brackets ([]) are treated as single words.

4.3 Features Mining

The two methods adopted to identify potential features for identifying doubts are student self-assessed rating and sentiment analysis. Since a Likert scale of 1 – 5 (with 1 being the lowest rating and 5 being the highest) were submitted by students together with their reflection to indicate their level of understanding. It was reasonable to assume that doubts are more likely found in reflections of the lower rating than the higher rating. Hence, this rating was extracted to analyze if doubts could be mined from the corresponding rating, with emphasis on the lower rating.

Sentiment analysis is commonly used in feedback analysis to extract feedback that is of value to improve course delivery and student experience. Since sentiments can provide insights to how well a student perceives the learning experience in class, we were interested in assessing if sentiment analysis could be leveraged for doubt identification. In this study, manual annotation of sentiment was done for a detailed analysis of the informal reflection data. In addition, an off-the-shelf sentiment analysis tool, TextBlob1, was used to assess if the sentiment identified by the software could be an identifier for uncovering the underlying doubts in the reflection. In particular, polarity score greater than zero indicated positive sentiment and polarity score lower than zero was considered as negative sentiment. The software was implemented, without adaption to the domain as we were interested in exploring if an off-the-shelf tool could be used to identify doubts.

4.4 Doubt Identification

In order to automate the doubt identification, machine learning algorithms built with various features and word representations were evaluated. Even though there are myriad of learning algorithms available, this study focused on the analysis and comparison of the features rather than the different types of learning algorithms. Logistic Regression (LR) was used in this study since preliminary study on Naïve Bayes did not yield better results from the data.

In LR models, the probability of observing a discrete label, y (1, 0) for a given input feature, x and the purpose of the binary classification is to find a parameter θ , such that $z = \theta^T x$, where $\theta \in \mathbb{R}^d$ and d is the dimensionality of input features x and z is a real number.

Logistic sigmoid function was then applied to z to fit to the range of [0,1] so that it can be interpreted as a probability of predicted output.

Given training data $\{x_i, y_i\}_{i=1}^n$, parameter θ can be found by minimizing the following where λ is a regularization parameter:

$$\min_{\theta} \sum_{i=1}^n \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|^2$$

The annotated data was split into 70-30 percent for training and testing purposes. Grid search was applied on a three-fold cross validation to find the best parameter of the training data. Result reported was based on the LR model run with the best parameter on the testing data. Two types of word representation methods were adopted. The first being the vector space model using unigram and bigram on both term frequency (TF) and Term frequency-inversed document frequency (TF-IDF) measures. The second was a neural embedding method using doc2vec's distributed bag of words model (Lau & Baldwin, 2016) with a dimension size of 100 and a minimum word frequency of 2. Doc2vec word representation has shown promising performance in various natural language processing (NLP) tasks so it is of interest to assess its performance against the commonly used vector space model in our context.

4.5 Evaluation Metric

Typical accuracy metrics used for statistical analysis of binary classification, which considers the true positive and true negative, have known issues in terms of reflecting the performance of a classifier (Sokolova, Japkowicz, & Szpakowicz, 2006). Therefore, we used F-measure or F1 score as the metric when assessing the performance of the various approaches proposed in addition to the correct assignment or accuracy percentage of positive and negative datasets. F1 score is the harmonic mean of both precision and recall where precision is defined as the ratio of true positive found from the predicted positive while recall is the ratio of true positive identified from the actual positive.

5. Experiments and Results

5.1 Analysis of annotated data

¹ <https://textblob.readthedocs.io/en/dev/>

Out of the 375 reflections extracted, 295 were annotated as ‘no’ or not containing doubts while 80 were labelled as ‘yes’. On the other hand, 100 were annotated as ‘positive’ or reflections with positive sentiment; 31 and 244 were labelled as ‘negative’ and ‘neutral’ respectively.

With the annotated data of doubt, Figure 2a shows that self-assessed rating can potentially be an indicator in identifying doubts since the percentage of reflections containing doubt decreases with increase in self-assessed rating. However, it does not necessarily imply that higher ratings (e.g., rating 4 and 5) do not contain doubts in the reflection. Similarly, in Figure 2b, the higher ratings have higher percentages of annotated positive sentiments. Conversely, lower ratings have higher percentage of reflections annotated with negative sentiment. Both ratings and sentiment analysis show their potential as features for identifying doubt in reflection statements.

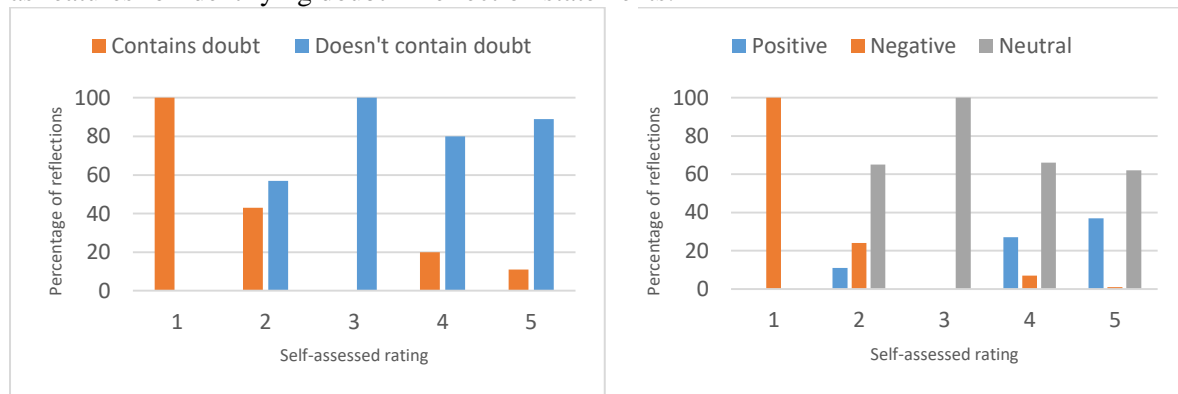


Figure 23. a) Percentage distribution of annotated doubt with respect to self-assessed rating and b) Percentage distribution of annotated sentiment with respect to self-assessed rating

We took further steps to assess if sentiment analysis could solely be used to identify doubts. Figure 3 shows that sentiment results should not be used directly to identify doubts as doubts can also be found in reflections with positive and neutral sentiments. These are two examples of actual reflections which students wrote (informally): (1) “Clustering interesting leh. HAHABA with the steps and powerpoint animations, very clear can understand. latent variable (g) and error -- g is unobserverd which is not reflecting in the SAS result or data so where does this infor appear ? but error is produced after the data is being analysed right?”; and (2) “I briefly learnt how K-means clustering work and how to interpret the results of a K-means clustering in SAS EG. I felt that I may need a brief revision on SSE.” The first reflection is an example of a reflection with positive sentiment containing doubt while the second reflection shows a neutral sentiment with doubt.

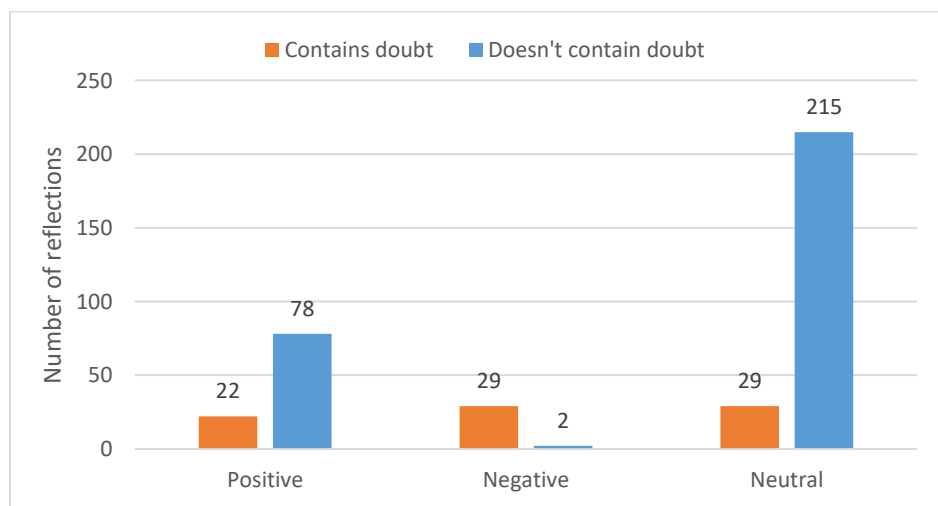


Figure 3. Number of reflections with doubts identified with respect to the three types of sentiments annotated

It is also reasonable to state an observation that reflections with negative sentiment are likely to contain doubts. However, statement of doubt can be found in all types of sentiments (refer to Figure 3). Therefore, it is not sufficient to identify doubts from sentiment analysis. It is essential to treat doubt identification problem separately from sentiment analysis.

5.2 Performance results of various features

With the annotated data, various LR models with different features were constructed for doubt identification. However, in view that the annotated data was an imbalanced dataset (with 295 reflections labelled as ‘no’ and 80 as ‘yes’), which might affect the accuracy of the model, random resampling via replacement of the smaller dataset was performed. The performance of the various features is presented in Table 2. The results (Model 1-4 in Table 2) shows that resampling of data achieved a higher F1 score. In other words, the imbalance sample size affects the performance of the LR model.

Next, we trained our model with a set of question features, that is, QM, 5W1H and QP (refer to Model 5 and 6 in Table 2) and similar results were found for both models. Further analysis was done on the question features and it was found that QM was detected in only 12% of the annotated reflections while 5W1H and QP patterns were found in 59% and 1% of the data respectively. Since QP was not commonly found, it is understandable why the model yielded the same result regardless of whether QP was used as a feature in the model. The question features, in fact was found to be one of the lowest performing features for identifying doubt. In short, reflection statements with questions do not necessarily reflect doubts.

In Model 7, polarity score assigned by TextBlob was used for detecting doubts. Specifically, a reflection with polarity score lower than zero is considered to contain doubt and a reflection with polarity score greater than zero does not. The result in Table 2 shows that the off-the-shelf sentiment analysis tool does not perform well in identifying doubts.

Table 2. Performance metric of LR models using various features

Model	Features*	Precision	Recall	F1 score
1	All data unigram features	0.70	0.29	0.41
2	All data unigram features with resampling	0.61	0.46	0.52
3	All data unigram, bigram features	0.86	0.25	0.39
4	All data unigram and bigram features with resampling	0.64	0.38	0.47
5	QM and 5W1H with resampling	0.29	0.38	0.33
6	QM, 5W1H and QP with resampling	0.29	0.38	0.33
7	TextBlob polarity score	0.24	0.51	0.33
8	Selected data with unigram features	0.70	0.67	0.68
9	Selected data with unigram, bigram features	0.83	0.62	0.71
10	Selected data with doc2vec embedding	0.76	0.75	0.75

*The result of Model 1-9 is based on the TF as the feature vector space. The result from TF-IDF is omitted since it is consistently lower than the above.

In view of the poorer-than-expected performance results of the features used in Model 1-7, it is plausible that none of the features can effectively separate the classes and thus we fine-tuned our feature selection method to select suitable features that can improve the model. Based on Figure 3, it is less likely for a reflection with positive sentiment to contain doubt, so a new dataset consisting of (a) all the annotated ‘yes’ reflections (80); (b) reflections with positive sentiment and ‘no’ doubt label (78) were extracted to be the selected training data. This training data was more distinctive in statements with doubt and those without doubts. The results (using unigram and both unigram and bigram as features) are listed as Model 8 and 9 in Table 2. The results show that selected data with unigram and bigram features achieve better F1 score compared to earlier models. One possible reason for the improvement of the result is due to the fact that reflections with positive sentiment contain features that can be used to clearly differentiate the identification of doubts. Indeed, with further analysis of the top features of the LR model, words such as “unsure, lost, confuse, don really” are extracted under the “yes” label (reflections with doubt identified) and “interesting, useful, clearer understanding, better” are found

under the features labelled as “no” (reflections that do not contain doubt). With the encouraging results from Model 8 and 9, doc2vec embedding was constructed using the selected data. The LR model with the embedding representation (Model 10) resulted in a highest F1 score of 0.75 among the tested models. Interestingly, both precision and recall values of Model 10 were of the same range, without being bias to any of the metrics. Therefore, we considered Model 10 as a better model compared to the best-performing vector space model (i.e., Model 9). We attribute the better performance of neural embedding methods to its ability to capture the semantic of words, which can be hard to represent using the vector space model.

5.3 Validation of model against qualitative survey reflection data without self-assessed rating

We validated our model using another qualitative reflection data collected by Instructor 2. This dataset had 71 reflections covering the topic from one lesson and there was no self-assessed rating to act as a reference. The three best models, namely Model 8, 9 and 10, from Table 2 were tasked to extract as many reflections containing doubts as possible. Based on the result stated in Table 3, LR model built using the doc2vec embedding performed the best with F1 score of 0.74. The other LR models (Model 8 and 9) built using TF vector space model resulted in lower F1 score of 0.57 and 0.59 respectively. This implies that neural embedding method is a better representation in identifying doubts from the informal reflections. The following are some of the reflections extracted that contains doubts:

- Need more help to determine the best fit line.
- I think the way p-value works should be revisited and explained more clearly because it was unclear during the class.

These reflections do not contain any known question features and hence model built purely with question features will not be able to identify these students requiring further attention.

Table 3. Result on validation data

Model	Features	Precision	Recall	F1 score
8	Selected data with unigram features	0.50	0.65	0.57
9	Selected data with unigram, bigram features	0.53	0.65	0.59
10	Selected data with doc2vec embedding	0.74	0.75	0.74

6. Discussions and Future Work

The LR model built using the selected data of annotated doubt and sentiment has shown a promising result in automating the identification of reflections containing doubt. The main reason that it was performing better than others is predominately the selection of the training data and the importance of identifying suitable features. One of the findings of this study shows that reflections with positive sentiment are less likely to contain doubts (but not the absence of it) and this feature can be used for constructing a better doubt identification model. In view of the importance of detecting positive sentiment, TextBlob was tasked with the evaluation. Out of the 100 reflections with positive sentiment, TextBlob only managed to identify 11 of them as positive. The rest were labelled as negative. With a F1 score of 0.2, it is of concern to use any of the off-the-shelf sentiment analysis tools to aid in identifying doubts. Since the accuracy of a sentiment analysis is heavily dependent on domain knowledge and data, it is essential to construct a sentiment analysis that is adapted to the domain for better accuracy.

We plan to expand the doubt identification study to more classes and courses to improve the model and analyze if it is feasible to identify phrases or doubt embedded patterns that are unique in the expression of informal reflections, for example, “I’m confused”, “need more help” etc. Besides that, various other machine learning algorithms such as Support Vector Machine, will be adopted on top of LR to assess the effect of different algorithms in doubt identification.

We noticed that student’s expressions from the reflections were more open, casual and in fact truthful. This is partly because the reflections were collected individually within a learning management platform where the statements were read only by the instructors. This behavior may not exhibit if reflections were collected in open discussion forums or formal feedback surveys such as course

evaluation collected by the institution where students may be more conscious about what they share and may be more reserved in expressing doubts or seeking help.

As shown in Table 2 and 3, LR model with doc2vec embedding performed the best based on the testing dataset of Instructor 1 and the qualitative reflection data of Instructor 2. One observation was the balanced values of precision and recall. This is encouraging since it showed that neural embedding is able to learn the intrinsic expression of informal reflections and can be used to improve the accuracy for doubt identification. With the promising results from various deep learning approaches in NLP, our future plan is to assess if pre-trained models coupled with transfer learning (Chronopoulou, Baziotis, & Potamianos, 2019) can be used to automate identification of relevant reflections with higher accuracy. The aim is to develop a generic model that can address doubt identification from other subjects and domains by allowing target task adaptation to the model. Besides that, the ultimate goal is to extract the topics or concepts from the reflections that required more attention. Once the doubts can be extracted, the next step is to perform topic modelling study to identify difficult or commonly misunderstood concepts for further clarification in class. We believe that timely feedback and doubt clarification play important roles in the success of learning.

7. Conclusion

In this study, we have analyzed the nature of informal reflections and investigated various features and word representation methods that can aid in identifying doubts from reflections. Our results show that selecting suitable features are important and reflections with positive sentiment do play a role in constructing a better model. In addition, using neural embedding as the word representation method has shown to achieve the best performance among our data sets. Analysis of the reflection data suggests that the self-assessed level of understanding Likert scale rating can be adopted together with the proposed automated approach to enable learner-centered methodology to improve students' understanding. With the ability to automate doubt identification, topics or concepts that may be difficult or misleading can be extracted more efficiently for providing timely feedback, doubt clarification and improve learning experience.

References

- Chronopoulou, A., Baziotis, C., & Potamianos, A. (2019, June). An embarrassingly simple approach for transfer learning from pretrained language models. *Proceedings of NAACL-HLT*, 2089-2095
- Cong, G., Wang, L., Lin, C. Y., Song, Y. I., & Sun, Y. (2008, July). Finding question-answer pairs from online forums. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval – SIGIR 08*, 467-474.
- Danielsiek, H., Paul, W., & Vahrenhold, J. (2012, February). Detecting and understanding students' misconceptions related to algorithms and data structures. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education – SIGCSE 12*, 21-26.
- Dhanalakshmi, V., Bino, D., & Saravanan, A. M. (2016, March). Opinion mining from student feedback data using supervised learning algorithms. *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, 1-5.
- Efron, M., & Winget, M. (2010, February). Questions are content: A taxonomy of questions in a microblogging environment. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-10.
- Gottipati, S., Shankararaman, V., & Gan, S. (2017, October). A conceptual framework for analyzing students' feedback. *2017 IEEE Frontiers in Education Conference (FIE)*, 1-8.
- Lau, J. H. & Baldwin, T. (2016, August). An empirical evaluation of doc2vec with practical insights into document embedding generation. *Proceedings of the 1st Workshop on Representation Learning for NLP*, 78-86

- Shankararaman, V., Gottipati, S., Lin, J. R. & Gan, S. (2017, December). Extracting implicit suggestions from students' comments – A text analytics approach. *Proceedings of 25th International Conference on Computers in Education*. 261-269.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006, December). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. *AI 2006: Advances in Artificial Intelligence*, 1015–1021.
- Wang, K., & Chua, T. S. (2010, August). Exploiting salient patterns for question detection and question retrieval in community-based question answering. *Proceedings of the 23rd International Conference on Computational Linguistics*, 1155-1163.

Towards Sustainable Learning Materials for MOOC in Poor Network Environments

Ahmed SHAMS^{a*} & Christoph MEINEL^b

^aRheinMain University of Applied Sciences, Germany

^bHasso-Plattner Institute, University of Potsdam, Germany

*ahmed.shams@hs-rm.de

Abstract: Videos are highly instrumental to the learning experience in many eLearning platforms. Lecture videos have become crucial to the success of any Massive Open Online Courses (MOOCs). However, conventional lecture videos are typically large computer files. Such videos are relatively enormous in file-size compared to other learning materials in MOOCs. With a slightly long runtime, a high-definition video file transfer comes with a high price. It is rather much expensive and time-consuming over poor or sparse networks. This commonly occurring situation significantly affects the learning experience. Online learners encounter intermittent connectivity or even offline more often during a learning session. In such situations, we claim that the prevalence of videos in MOOCs remains vague and subtle in poor network environments. In the file transfer process, it is not unusual for a video to face constant network disruptions. Therefore, this research work urges to review the provision of alternative, but more effective multimedia. Such multimedia are instead proposed to be much smaller in file-size. However, the multimedia should retain the essential criterion of a video-based format. The new materials are thus deemed less dependent on the bandwidth. This paper focuses on improvising video-based materials during a real-time learning session. Of course, be derived from their original lecture videos. We argue that this approach helps to improve the learning experience, especially for learners situated in low bandwidth areas. Network related issues such as long-wait buffering, low-quality streaming, or slow downloading are therefore tackled by using unconventional materials. The main benefit is to achieve and attain uninterrupted learning session even when the MOOC learner steps into offline or out of reach area. Also, the approach supplements images and audio to support sustainable eLearning – whenever wherever.

Keywords: Video, MOOC, low bandwidth, learning experience, materials, multimedia

1. Introduction

Videos are focal to the student's learning experience in the current generation of Massive Open Online Courses (MOOCs). Popular MOOC providers such as Coursera, edX, and openHPI intensively use videos as learning material. Videos are now an essential lecturing instrument. However, lecture videos used in MOOCs are typically large computer files, relatively big enough to encounter file transfer disruption. The file-size is usually a concerning issue when transmitting a video over a threshold of a network. A diminished bandwidth can prevent the transmission of the video or even corrupt the file. In poor performing networks such as 2G and public WiFi, the learning process might be affected due to slow video processing. These network related issues not only affecting online user experience, but also contribute student's dropout in MOOCs (Kim et al., 2014), (Mamgain, Sharma & Goyal, 2014).

In extended MOOCs; online courses follow the traditional university pedagogically. Online courses are mostly organized as sequences of instructor-produced videos infused with problem sets, reference materials, or other resources (Guo et al., 2014). Online learning is primarily inclined towards the use of lecture videos. Materials such as textbooks, PDFs slides, and reference web links are rather supplements. In MOOCs, videos have evolved to be a vital component in delivering educational content. Many MOOC platforms currently use their multimedia players to stream videos or provide a download mechanism (Meinel & Willems, 2013). The media player offers an interactive interface between the videos and the learners, an essential component of any MOOC. Despite the versatility of control functions, many media players present in MOOC platforms still lack dedicated features designed to sustain the learning process in poor network dynamically (Brinton & Chiang 2014).

Conventional approaches use data compression techniques for transmission of materials to curb broadband issues; however, this solution is limited to certain file-sizes which might not yield a significant gain in low bandwidth environments. A different, but traditional solution progressively reduces the display resolution, but compromises the viewing quality as well as discredits the auditory perception. Popular MOOC platforms, nevertheless fail to enact control over the choice of suitable learning materials based on the healthiness of the network or self-decision of the learner (Barba, 2015).

This research work applies a relatively new approach involving content extraction mechanisms follows by multimedia combination to provide video-based learning materials. The work focuses on the provision of images and audio as alternative multimedia to video. This approach significantly relaxes the necessity of network broadband for MOOCs. In our experiments, we managed to achieve relatively low data consumption using our alternative learning materials due to small file-size. By using suitable types of multimedia that avoid high consumption of network resources, we managed to reduce network dependency, sustain delivery of educational content in poor network environments.

This research work aims to improve the learning experience in MOOCs. In particular, the use of alternative materials in sparse networks. The rest of the paper is structured as follows: Section 2 reviews the challenges of using lecture videos at low bandwidth. Section 3 presents related work with state of the art tools and technologies in MOOCs. Section 4 executes performance experiments with results detailed in section 5. Section 6 concludes and gives a future outlook.

2. Challenges of Lecture Videos for MOOCs in Poor Networks

Video learning materials are more network-resource demanding while transferring from a server to a client –a learner. Contrarily to remaining materials such as PDF or text files due to their relatively small file-size. We set up a computer file transfer experiment to determine the success rate of transmitting a video over the network. The experiment used a real lecture video borrowed from the openHPI MOOC platform. However, the specimen video is roughly shorter in length, and of relatively small file-size compared to many videos in openHPI. Typically, most of the MOOC videos lecture from 8-20 minutes with typical file sizes of 50-140MB in Standard Definition (SD) quality (Stuchlíková & Kósa 2013). In this experiment, the specimen video is about 8 minutes long and 50MB in capacity. For clarity, we opt to use the same exemplary video throughout our experiments for the rest of this paper.

First, we measure the duration it takes to download the video with different download speed over the network. In a low bandwidth setup, the download time is set at less than a 512 kbit/s. A potential learner might fail to watch the video online as a file transfer would take more than 13 minutes to complete. Whenever the download speed drops, network bottleneck could delay or prevent the video streaming process; consequently affecting the student’s learning experience. In 2G connections, video streaming takes as long as twice the time of its duration. This situation is undesirable experience in MOOCs. Table 1 summarizes Test Status based on the Internet connection type. Results show 1 Mbit/s is a theoretical minimum download speed required to deliver a video of 50MB in less than 8 minutes to the learner efficiently, i.e., within the running time of the MOOC lecture.

Table 1

Theoretical minimum download times for the 50MB video file with different types of connections.

Connection Type	Download Speed	Download Time	Test Status
Dialup	28.8kbit/s	04:02:43	Fail
Modem	56.6kbit/s	02:04:49	Delay
ADSL	512kbit/s	00:13:39	Hold-up
ADSL Lite	1Mbit/s	00:06:59	Pass
Turbo 3G	7.2Mbit/s	00:00:58	Success
4G LTE	80Mbit/s	00:00:05	Success

In contrast, access to a persistent and high-speed Internet connection is still a continuing challenge to many existing and potential MOOC learners. In reality, many Internet users still live in regions with low coverage. Internet access in many regions is still limited to mobile wireless communications, which are usually more expensive and slow. The challenge is a far difficulty for

learners of MOOC in rural and remote areas. Therefore, a different approach should be explored that facilitates the transfer of learning materials from the side of MOOCs (Renz, Shams & Meinel, 2017).

2.1 Motivation, Approach, and Methods

Huge video file-size is a common problem facing MOOC students as it is expensive and time-consuming downloading course content. It is not unusual to receive comments from students concerning the file size problem as one below (Guo et al., 2014).

"First, I would like to share with you (MOOC provider) the trouble of downloading files but also access to your site (MOOC platform). It would be great if you could reduce the file size" and "Internet in my community is poor and streaming is difficult. Can I get an audio format of the lecture? Where can I download if possible?"

In response to the challenge above, we aim to provide relevant learning content suitable for narrowband environments. Nevertheless, we are motivated to improve the overall performance of MOOCs in intermittent connectivity. Our approach focuses on the development and provision of alternative learning materials for videos.

The approach primarily advocates the usage of blended multimedia, created from a combination of images and audio extracted from the original video. To achieve relevant content, we have opted to improvise a set of new learning materials from the original video itself. The experiment video would then provide images and audio assets, of course, in an automated digital process. The assets, in turn, would blend in much small file size, enough to sustain sparse networks. Nevertheless, the blended materials are supposed to retain the preferred video-based format. Figure 1 shows the potential benefits of using alternative and blended materials.

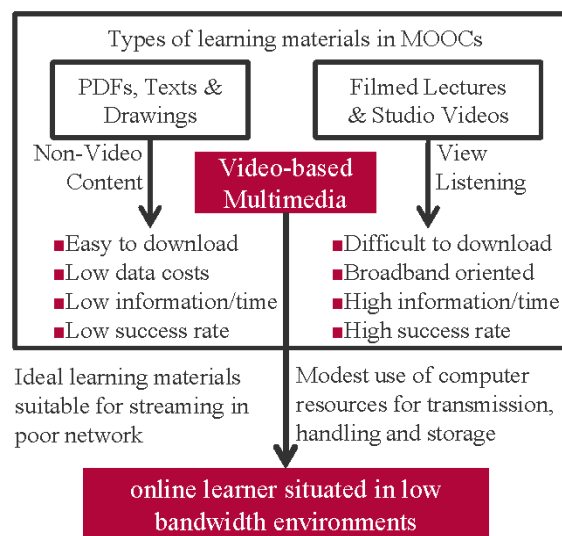


Figure 1. The materials shown on the left side are much easier and faster to download than on the right side.

3. Related Works

3.1 Usage of Videos as Learning Materials in MOOCs

Increasing use of smartphones, tablets, and similar mobile devices among students elevates enrollments in MOOCs. Operating systems of mobile devices, together with mobile applications, are now capable of processing varieties of videos. MOOCs are made possible because of readily availability and access to lecture videos via mobile devices (Jordan 2014, Bralić et al., 2015).

In contrast, mobile computing power is still constrained to battery power, display screen size, and wireless communications. All three factors are prone to poor performance whenever a video is used as prominent learning materials in MOOCs. Typically, many MOOCs try to address the network problem by reducing the video resolution, which affects the visual quality. Some MOOCs deploy compressed videos to tackle network issues. Nevertheless, in most cases, these approaches fail to deliver videos in poor network environments because the file size of the video is still not low enough to pass the threshold of the network (Huang N et al., 2017).

openHPI¹ MOOC by Hasso-Plattner Institute (Meinel & Willems 2013) aims to influence an offline mode in MOOCs. The first step involves the use of lecture slides, video transcripts, audio, and other alternative learning materials to supplement its High Definition (HD) quality videos. The material's file size is first supposed to be practical for local storage in cache memory for offline use. Manipulation of MOOC videos in previous research works is an interesting approach that provides distinctive multimedia files useful as alternative learning materials (Zhao H et al., 2018).

3.2 SMIL of Multimedia in MOOCs

Synchronized Multimedia Integration Language (SMIL) describes multimedia presentations. SMIL allows presenting media like images, audio, and other SMIL presentations (Télez 2010). Multimedia processing includes encoding, image extraction, and audio synchronization. To our video, we deployed an existing online tool² developed by the Tele-Task system to extract slides from the video (Haojin & Meinel 2014).

The tool applies automatic video segmentation and key-frame detection to offer a visual guideline for the video content navigation. Consequently, it detects the slide area from video frame and captures the content of the slide area to produce a thumbnail image. The margins surrounding the slide area are discarded, including the instructor. Figure 2 shows the extraction of content from the video.

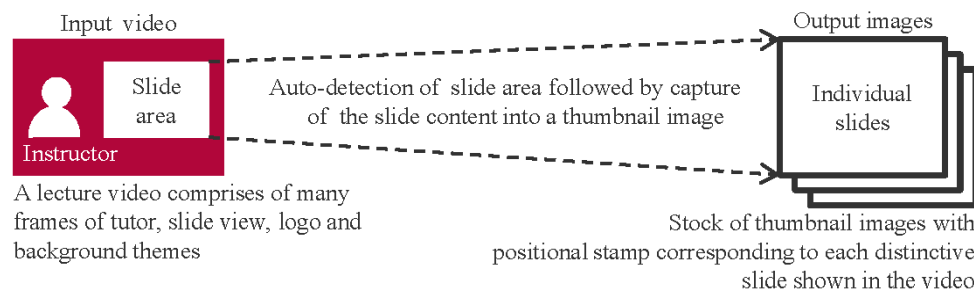


Figure 2. Automatic extraction of content from the slideshow. A stock of thumbnail images is thus produced over time that corresponds to the distinctive slides shown in the video.

For handling multimedia data, we use *FFmpeg*³, an open-source command line that provides codec library for digital manipulation of a video into an audio file. An FFmpeg library is integrated into the MOOC platform to enable automatic read of an input video file, transcode and produce an audio file in that sequence. By using related research works, we can acquire thumbnail images and audio from the conventional video (Lei, Jiang, & Wang 2013).

At this stage, we require another tool that stitches together images and audio files to create on-the-fly video-based multimedia. The multimedia should suitably play in the media player. For re-usability of web-based HTML elements, we use *Polymer*⁴, an open-source JavaScript library for building components to the media player that promptly combines images and audio files.

¹ <https://www.openhpi.de>

² <http://fb10mas02.hpi.uni-potsdam.de>

³ <https://www.ffmpeg.org>

⁴ <https://www.polymer-project.org/>

4. Performance Experiments

Lecture videos in MOOCs are typically recorded in a production studio with professional cameras before publication which is generally a costly process. Most of the videos are at least in standard definition quality with resolutions options of 240p, 360p, and 480p. Some MOOCs platforms provide 720p, 1080p, and original, which is the highest quality setting. As mentioned earlier, the video's running time usually varies from 8 to 20 minutes, with around 50-120MB of file-size.

First, we examine the properties of the specimen video from the MOOC course. Table 2 lists technical features of the approximately 50MB video file used in this representative experiment.

Table 2

Technical specifications of the 50MB specimen video

Duration ~8 min	File size 51MB	Bit rate 715kb/s	Codec H.264
No. of frames 11786	Frame rate 25fps	Audio mp4	Streams 2
Channel Stereo	Width 1280p	Height 720p	Display 16:9

Apart from the technical aspects, the video also contains a standing instructor (tutor) on the left side of the slide show display. The background is made up of MOOC logo, computer table, and any other background themes.

4.1 Multimedia and Design of Media Player Prototype

To validate our approach, we build a prototype to test our hypothesis. Our initial model consists of a media player that handles multimedia assets to produce video-based learning materials.

The early model of a media player is expected to render images while playing audio asset promptly to produce video-based learning content to the learners. The prototype comprises two main components which are slide extraction and video-audio transformation.

The first component intakes a chosen video as input and extracts its distinctive slides based on the number of pages in the wholly slide document. This process produces a stock of thumbnail images as an image asset. A single thumbnail image thus contains the content per page of the slideshow. In this case, a total of eight unique pages of slides are produced corresponding to eight thumbnail images. The second component employs an *FFmpeg* tool that transforms the video to an audio asset digitally; merely an audio output encoded in MP3 format.

In the next stage, The *Polymer* components combine the two assets promptly within the multimedia player. The elements supposedly display thumbnail images based on timing/positional information while playing audio. The combination of assets within the media player creates video-based learning content on-the-fly. The prototype makes use of play and pause function controls over the multimedia.

Besides, the player enacts control over slide seek and the time shift of the slides that offer added experience to learners. Figure 3 shows the design and workflow principal of the prototype. The two components work individually but promptly produce video-based content. The next subsection gives an in-depth detail of the two components.

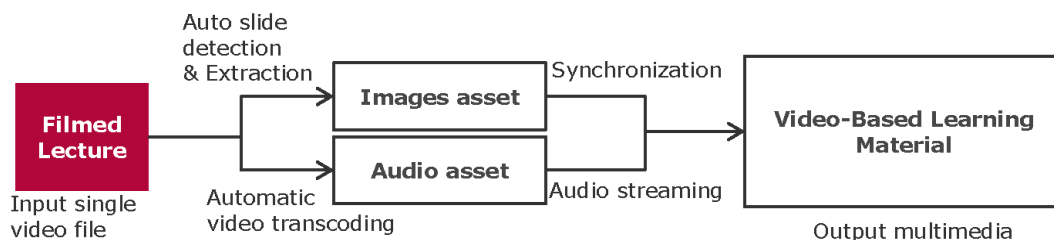


Figure 3. The architecture of the prototype for the media player outlining the production of new learning materials from a video input.

4.2 Slide detection and analysis of Thumbnail Images

A video is merely an integration of imageries. A burst of thumbnail images can be used to create a video when images are then displayed one after the other several times a second. The eye perceives this sequential display of images like a film - a video of moving objects. The quicker the images are played, the smoother and more fluid the movement looks.

A video file on computer stores all the frames together in a container and presents them in order, and the total frames stored for a typical MOOC video reach into tens of thousands. Our specimen video has 11,786 individual images, filmed at nearby 24 images per second. Each image is called a frame.

The image extraction tool is supposed to capture all distinctive frames based on the textual content of the slides. The tool thus produced only eight frames while discarded the rest (thousands of images with quite a similar slide content). The mechanism ensures extraction of unique frames from slides, i.e., eight images corresponding to the total page number of slides. The frames are stored as thumbnail images, entirely occupying as low as 593kb of disk space. Table 3 shows file-sizes as well as the position of each frame corresponding to the video.

Table 3

File size and position of each extracted individual frame

Thumbnail Image	1	2	3	4	5	6	7	8	Total
File size in KB	102	52	52	43	61	67	64	100	593
Position of a Frame	25	675	2000	3925	5675	8825	10400	11786	n/a

The actual position of such a video frame is presented in the last row in the table. This information is vital for a later stage of synchronization, which stitches the time of thumbnail and its corresponding starting position while audio is being played.

4.3 Video transcoding and Audio Analysis

Using FFmpeg library, a video file is fed as input to produce digital audio by the transcoding process. The library provides customized options to transcode video, as demonstrated in the following piece of code provided in the programming terminal.

```
ffmpeg -i input_video.mp4 -acodec libmp3lame -qscale:a 9 -ac 1 output_audio.mp3
```

This short, robust code encodes audio in MP3 format with a quality scale of 32 bit-rate as a scaled in 9. It is monophonic with a single channel layout, and the most important property in this context is its file-size of 1.88MB. The duration of the audio is 417 seconds, similar to the video length. The output audio is stored in local storage of the computer or mobile device.

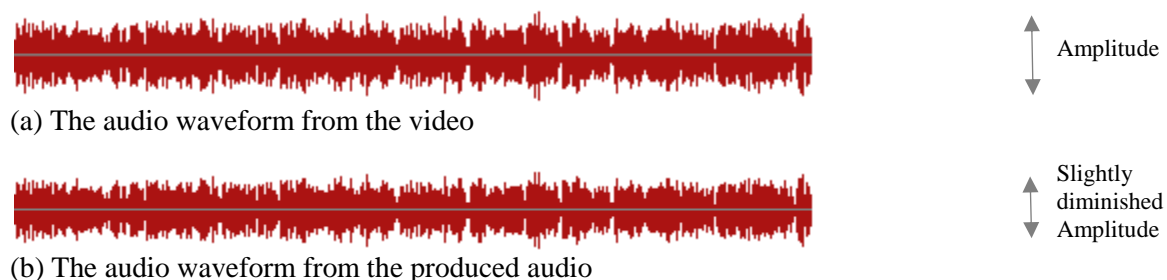


Figure 4. The audio waveform on top (a) shows a striking resemblance to the audio waveform on (b).

Figure 4 shows a time-domain display, a display of amplitude versus time waveform properties for both input video and output audio. The audio waveforms seem to be quite similar to each other, without major signs of distortions, implying that the sound quality offers a relatively similar listening experience as the video.

4.4 Combination and Synchronization of Multimedia

The media player finally combines images and audio in the browser and load a document as an HTML page. Regarding the file-size of the assets, thumbnail images altogether occupy 0.59MB while audio asset mounts to 1.88MB, yielding the multimedia with a total file-size of 2.47MB. An extreme reduction in file-size compared to the 50MB of the original video.

For a single video file, a page loading time of 2-3 seconds is registered of which a complete web page is rendered to the learner. The page loading time is much higher when using multiple files as it takes longer to assemble and load all the files. In our prototype, it takes up to 57 seconds to load all thumbnail images and audio files. Repetitive experiments with the help of cache memory show a decrease of page loading time, which improves the performance of the web page in poor networks.

5. Results and Discussion

Non-video learning materials such as PDFs, slides, and textbooks are easy to download and cost fewer data in low bandwidth environments, however, they are less prevalent in MOOCs, and their success rate is low compared to their corresponding videos. Video records of lectures and animations are the most preferred teaching materials in many MOOCs. Despite their prominence, transfer of videos over the network is broadband oriented and comes with high data costs that might fail to sustain access to MOOCs and learning events in poor network environments.

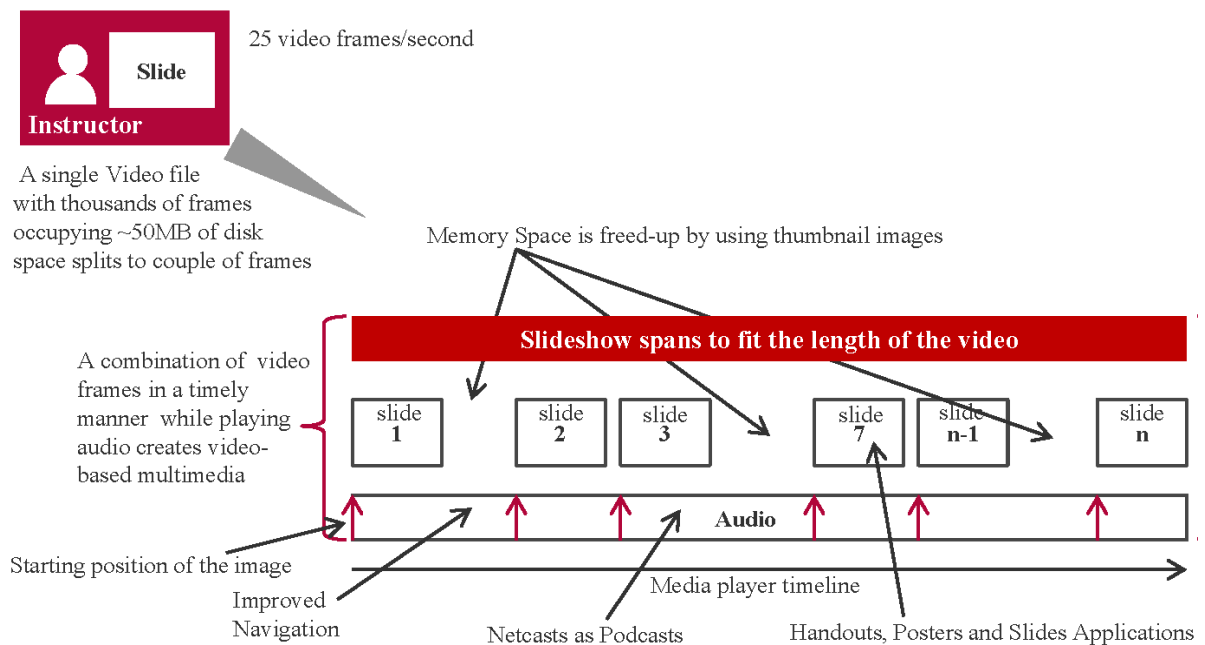


Figure 5. The structure and features of the video-based multimedia as alternative learning materials for MOOCs in a poor network environment.

To improve sustainability, we introduced an alternative, small-sized multimedia aimed to reduce network dependency as well as free-up disk spaces. Figure 5 illustrates the structure of lightweight learning content. The top left side of the image shows a single input entity as a video file, while the right side demonstrates the outcome derived from the prototype. Synchronous thumbnail images mount to an n integer which was 8 in the experiment. All thumbnail images are displayed within the playing time frame of the audio file.

Video-based multimedia presented in this paper fits in between the non-video and the conventional video learning materials in term of file-size, network transfer, handling, and storage. The said video-based multimedia consumes moderate bandwidth due to its file-size, which is suitable enough to sustain low bandwidth environments.

Regarding usage in mobile devices, a moderate computing power might be sufficient enough to process the learning materials due to the nature of the multimedia, i.e., a small number of thumbnail images rather than tens of thousands of video frames. Mobile devices only are expected to compute a number of images while playing audio to deliver educational content to the learner. This approach successfully renders the content, but also saves battery power in comparison to video computation.

The alternative video-based multimedia has improved access to learning materials in poor network environments. Multimedia in small file sizes use fewer network resources and incur low Internet data costs. In reducing network dependency, offline capabilities might be enabled when multimedia files can fit cache memories of the browsers. MOOCs can thus stream video-based learning materials directly from its local storage - cache memory even in offline mode. The offline-enabled MOOC is potentially in demand and useful to many students and professionals whose access to the Internet is only limited at universities and workplaces.

In this experiment, we digitally split a single video file into many images as well as audio files. The splitting process brings new types of learning materials, as benefits brought forward by different multimedia than a single video file. The resulting images, for example, are used for creating posters, printing hand-outs, thumb-nailing lecture courses and other educational purposes within and outside the realm of MOOCs. For instance, audios are useful in net casting, and a user can download and listen to as Podcasts. The formed multimedia also provides component re-usability and re-applicability potentials in MOOC.

A compromised learning mode surfaced as a drawback once lecture videos are removed from MOOCs since a video is more informative than combined still images and audio. Narration style and body language of the instructor affect the perception and understanding of the learner. In eLearning, the appearance of the tutor on the video is an added advantage to the learning process. However, in this case, the educational context depends on the slide's content and tutor's sounds alone. Self-adjustment in learners is, therefore, expected by carefully listening while paying concentration to the images.

The implementation of the new learning materials has one more shortcoming concerning system dependability and complexity. There is an increase in system dependability due to the usage of two or more components such as image extraction tools, and codec library that comes from third parties. This increases unreliability as well as elevates the complexity of the overall system. The website in a process also needs to load many images and audio at first place before rendering them to the user. A page loading time usually increases with the addition of individual multimedia files. A browser needs to process every additional multimedia file rather than a single video file. Consequently, the learner might wait for several seconds (might be more than a minute in poor networks) before the media player is completed loaded and ready to use.

Despite the potentials of MOOC such as free and open online education, only a handful of Internet users - potential students participate in eLearning. Participation is relatively less in disadvantaged regions where access to higher education remains a challenge. The reasons for shallow MOOC activities are usually associated with unreliable access to a persistent and high-speed Internet connection, thus hindering potential learners from streaming lecture videos to their mobile devices or laptops (Renz, Ahmed, & Meinel, 2017).

Recent advancement in mobile technologies and wireless communications produces a conducive digital environment enough to support eLearning at the pace and discrete of the learners despite mostly being in a broadband state. With Internet access even in low bandwidth, MOOC courses are expecting development anywhere the learner sees fit, for example, while commuting between cities, at home in rural areas or even in remote places (Renz, Shams & Meine 2017). The current state of the existing digital environment is thus a motive for the development of eLearning features sustainable in poor networks. Potential learners of MOOC expect the provision of course materials to be readily accessible to their systems regardless of the needy broadband. Figure 6 shows the expectation level of learner's participation in MOOCs provided that solutions to challenges such as access to alternative materials suitable for low bandwidth exist.

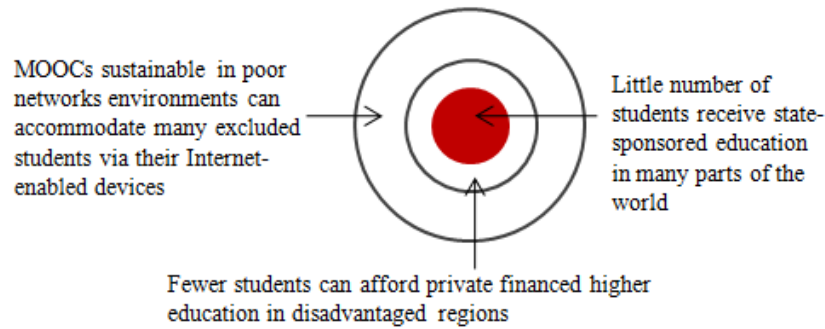


Figure 6. Conceptual Framework of MOOC Experiences based on General Expectation of Learner's Participation and Completion of Online Courses.

6. Conclusion

This research work presented and evaluated alternative multimedia that comprises images and audio assets. In an automatized process, the assets are extracts from existing lecture videos in MOOCs. The alternative multimedia are created on-the-fly and displayed in a video-based format to the learner. They have a significantly small file-size overall compared to the lecture video. The video-based display is a result of the timely ordering of images while playing audio. The educational content of the alternative multimedia seems to be suitable for MOOC in low bandwidth, intermittent Internet connection, and similar poor network environments.

In our opinion, the results presented here are of great interest to MOOC providers and learners, especially those potential learners living in regions where access to a persistent and high-speed Internet connection is still a challenge. The implementation of the alternative multimedia in MOOCs adds a choice value of learning materials while boosts learning experience. It is an important initial step towards the provision of educational content with versatile adaption to low bandwidth environments. Despite the potential observed success, this research is limited to a single MOOC platform with a small sample of research materials, the short duration of the study, and potentially biased choice of technologies involved. Different MOOCs might offer distinctive forms of learning materials from our experiment as there are no standard regulations that MOOCs must conform.

In addition, further research is needed to study the effects of learning materials in MOOCs in sparse networks. First, the conceptual framework for the selection and provision of learning materials to improve learners experience in this article required further application across diverse MOOC environments and learner populations to determine if it is a viable framework to consider when developing and implementing customizable learning materials. Second, there is a need to design technologies, such as versatile media player, which support different network environments and Internet connection types where learners can self-determine their personal preference and select an appropriate form of learning materials for success and completion of the online course. Third, future research should also examine MOOC design and conditions for engaging learners in offline. The offline capability in MOOC can bring new impulses and impact to high education.

Future Work: Despite the promising improvements of using on-the-fly, and relative small file-size multimedia to deliver video-based educational content to MOOC learners, our model is still at an infancy stage which requires further improvements to achieve efficient learning, particular to those situated regions prone to poor networks. Technical developments such as for network control feature that enables a user to toggle between different modes of the network depending on the Internet connection and bandwidth consumption. A toggle button is thus a useful feature to allow switching between different network modes. We also expect to deploy the media player online soon. Thus we can observe and study the usage of the prototype. The process, in turn, provides essential data which, upon analysis, might help to determine deciding factors to improve learner's experience in MOOC platforms.

Acknowledgements

We would like to thank all the people at the HPI who took part in this research work.

References

- Guo, P. J., Kim, J., & Rubin, R. How video production affect student engagement: An empirical study of MOOC videos. In Proceedings of the first ACM Conference on Learning @ Scale Conference, L@S '14, pages 41–50, New York, NY, USA, 2014. ACM.
- Meinel, C., & Willems, C. openHPI: The MOOC offer at Hasso Plattner Institute. Technical report, pages 1–17, Hasso-Plattner Institute, Potsdam, Germany, Sept. 2013.
- Renz, J., Shams, A., & Meinel, C. Offline-enabled web-based elearning for improved user experience in Africa. In IEEE AFRICON 2017 Conf. Proc., pages 5–13, Cape Town, South Africa, Sept. 2017. IEEE Explorer.
- Yang, H., & Meinel, C. Content based lecture video retrieval using speech and video text information. IEEE Transactions on Learning Technologies, 7(2):142–154, April 2014.
- Stuchlíková, L. & Kósa, A. "Massive open online courses - Challenges and solutions in engineering education," 2013 IEEE 11th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stara Lesna, 2013, pp. 359-364.
- Shams, A., Tareaf, R., Renz, J., & Meinel, C. Smart MOOC – social computing for learning and knowledge sharing. In Proceedings of the 10th International Conference on Computer Supported Education - Volume 2: CSEDU, pages 391–396. INSTICC, SciTePress, 2018.
- Barba, A., Lorena. Why so many MOOC videos are utterly forgettable. A post secondary learning article. May 11, 2015. <https://www.edsurge.com/news/2015-05-11-why-my-mooc-is-not-built-on-video>. Last visit on 20th May 2018.
- Mamgain, N, Sharma, A & Goyal, P. "Learner's perspective on video-viewing features offered by MOOC providers: Coursera and edX," 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE), Patiala, 2014, pp. 331-336.
- Lei, X., Jiang, X., and Wang, C. "Design and Implementation of a Real-Time Video Stream Analysis System Based on FFMPEG," 2013 Fourth World Congress on Software Engineering, Hong Kong, 2013, pp. 212-216.
- Bralić, A., Ćukušić M., & Jadrić M. "Comparing MOOCs in m-learning and e-learning settings," 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2015, pp. 979-985.
- Huang N., et al., "VideoMark: A video-based learning analytic technique for MOOCs," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 753-757.
- Zhao H., Li G., & Feng W. "Research on Application of Massive Open Online Course (MOOC) in Modern Medical Education Teaching," 2018 International Conference on Engineering Simulation and Intelligent Control (ESAIC), Changsha, 2018, pp. 389-391.
- Téllez A. G. "E-learning Authoring with Docbook and SMIL," 2010 10th IEEE International Conference on Advanced Learning Technologies, Sousse, 2010, pp. 246-248.
- Brinton C. G., & Chiang M. "Social learning networks: A brief survey", Proc. Conf. Informat. Systems Sci., pp. 1-6, 2014.
- Jordan K, "Initial trends in enrollment and completion of massive open online courses", Int. Rev. Res. Open Distrib. Learn., vol. 15, no. 1, pp. 1-28, 2014.
- Kim J., Guo P. J., Seaton D. T., Mitros P., Gajos K. Z., & Miller R., C. "Understanding in-video dropouts and interaction peaks in online lecture videos", Proc. 3rd ACM Conf. Learning@Scale, pp. 31-40, 2014

Propositional-Level Analysis of Collaborative Learning with Kit-Build Concept Map

Yusuke HAYASHI, Toshihiro NOMURA, and Tsukasa HIRASHIMA

Graduate school of Engineering, Hiroshima university

hayashi@lel.hiroshima-u.ac.jp

Abstract: Typically, classroom practices in the collaborative learning context center around three distinguishable levels of activity: individual activities, group work, and whole-class discussion. It is important to identify and analyze the correlation between these levels to enable teachers to understand and improve the dynamics of students' understanding in the context of collaborative learning. This study proposes a method based on the Kit-Build concept map (KBmap) to analyze the relation between individual activities and group work in the classroom. The KBmap is a type of close-ended concept map that provides decomposed concepts and links from the concept map made by a teacher. This mechanism enables teachers to assess students' understanding, and facilitate the coordination of learning in the classroom. To evaluate the proposed method, a junior high school in Japan is used as case study.

Keywords: concept map, kit-build, classroom orchestration, learning analytics

1. Introduction

One of the core roles of teachers in the classroom is that of the coordinator of classroom discussion. Therefore, it is necessary for them to monitor students' understanding or perspectives, and facilitate classroom discussion. Essentially, a teacher manages the classroom by orchestrating the integration of individual activities, teamwork, and whole-class activities (Dillenbourg, & Jermann, 2010; Dillenbourg, 2013).

There are several methods of monitoring collaborative learning; some of these methods include evaluating the students' understanding using rubrics, analyzing their learning activities step by step, and discourse analysis. Technology plays an important role in enabling teachers to monitor the activities of learners. Martinez-Maldonado et al. propose and develop a multi-tabletop classroom and dashboard to support collaborative learning (Martinez-Maldonado, Clayphan, Kay, & Yacef, 2014). Their study provides a special environment for learners to work collaboratively with concept maps, and for teachers to access their verbal and physical interactions. Matsuzawa et al. propose a tool for exploring the network structure of collaborative learning discourses (Matsuzawa, Oshima, Oshima, Niihara, Sakai, 2011). This tool visualizes the dynamics of the network structures of learners, discourse units, and words.

This study proposes the analysis of learners' understanding based on concept mapping and collaborative learning at the propositional level, that is, through the propositions they make. Similar to KBDeX, the analysis is content-oriented (Hoppe, 2017); however, the unit of analysis, words or propositions, is different. Concept mapping is a popular way of representing learners' state of understanding (Novak, & Cañas, 2006). A concept map is a generic name for the graphical representation of the process of organizing and representing knowledge or understanding. Novak's definition of the concept map is particularly well known. Generally, concept mapping is predicated on building concept maps based in an open system without restrictions using arbitrary nodes and links. In this method, simple components, nodes, and links represent concepts and their correlations are emphasized. Although concept maps can be a tool for individual learners to reflect on their understanding and thereafter communicate their understanding to others (Tergan, 2005), they are difficult to analyze (Herl, O'Neil, Chunga & Schacter, 1999).

The Kit-Build concept map (Hirashima, Yamasaki, Fukuda & Funaoi, 2011; Sugihara, Osada, Nakata, Funaoi & Hirashima, 2012), subsequently referred to as the KBmap, can automatically analyze concept maps, because learners create concept maps from components that are decomposed from the

concept maps created by teachers. Using these components, learners can organize their understanding in a comprehensible way in the form of concept maps, thus enabling teachers to assess their understanding immediately (Hirashima, Yamasaki, Fukuda, & Funaoi, 2015). The KBmap assessment method is automatic, and its validity for evaluating learners' understanding has been confirmed (Wunnasri, 2018). In addition to the one-on-one comparison, it is possible to merge the maps of individual learners to obtain the representation of their aggregated understanding. The teacher can also compare the overlapped map with the goal map. Using the overlapped map, the teacher can analyze the trend of learners' understanding in the class, thus making it possible to provide feedback to the learners (Pailai, Wunnasri, Yoshida, Hayashi & Hirashima, 2017; Yoshida, Sugihara, Nino, Shida & Hirashima, 2013).

In this paper, we analyze collaborative learning among learners from the point of individual activities through small-group task based on the KB method. The rest of this paper is organized as follows. Firstly, we present an overview of KBmap. Then, we present and discuss our case study and the data captured from it. We conclude with the results and future research directions.

2. Collaborative learning with KBmap in classroom

2.1 Goal of lesson

The goal of the lesson in this study is that, at the end of the course unit, the learners collaboratively organize what they have learned in the unit as shared knowledge in the class. Toward this end, the teacher designs the lesson as a sequence of individual activities, small-group task, and whole-class discussion. First, the students individually organize what they have learned, after which they share it to verify or compensate for a gap in understanding. Finally, the results of the small-group task are shared in the whole-class discussion, and understanding can be achieved based on the fusion of a variety of perspectives.

For coordinating this kind of lesson, it is necessary for the teacher to keenly monitor students' learning in each step, and provide suggestions accordingly. To meet this requirement, the teacher in this study uses the KBmap system to capture the collaborative learning process in the lessons. In addition to enabling teachers to efficiently evaluate students' learning in the form of the concept map, constructing the KBmap also enables learners to organize what they have learned.

Figure 1 shows the concept map made by the teacher, subsequently referred to as the goal map, in this study. The topic considered in the study is the characteristics of Latin America, with emphasis on the relationship among industries, economic development, and deforestation. The teacher assumes that the learners have no knowledge of all the relations, and have a preconceived notion the relationship between development and deforestation is negative. It is expected that after this lesson, the learners will know the correlation of all the industries with both economic development and deforestation, and be able to explain these correlations. Additionally, it is expected that beyond holding a simplistic view of development and deforestation being inharmonious, learners become aware of the dilemma of both concepts. The aim of this lesson is that the students learn by independently and collaboratively integrating the knowledge acquired in previous classes in addition to the teacher's complimentary instruction.

The goal map is decomposed into separate nodes and links, and provided to the students as a kit to compose a concept map representing their understanding. Figure 2 shows the kit provided to the learners in this study. In this lesson, the teacher decomposes only the right part of the goal map to clarify the tasks undertaken by the students. The left part of the goal map represents the geographical features of Latin America, and is not decomposed; it maintains its structure. The left part had been composed by the students in the previous lessons.

On the other hand, the right part represents the correlation between the features of economic growth (development) and deforestation (destruction). The main task of the students in this lesson is to explore the correlation and represent it on a concept map. In this lesson, students build a concept map representing their thoughts using the kit, and then, construct an accurate map following discussion.

As stated above, the teacher's expectation is that using this concept map, the learners learn the correlation among all the industries and development and deforestation. The KBmap editor becomes a learning material that aids the learners to represent their understanding, and the KBmap analyzer becomes a tool for teachers to capture learners' comprehension during and after class.

Although the KBmap is limited by its closed content, it sufficiently enables the shared understanding of the lesson content, and lays the ground for realizing collaborative knowledge. In this study, the aim is that the students share their understanding of what they learned in the previous class, as the preparation for creating collaborative knowledge (Stahl, 2000). In constructing the KBmap, learners use limited nodes and links provided by the teacher. Even if the learners only assemble a concept map instead of segmenting the source information, Funaoi et al. (2011) demonstrate that constructing the KBmap achieves the same learning effect in relation to the contents in the goal map. This suggests that the KBmap can help learners to express their understanding based on the provided components. Moreover, teachers can utilize the diagnostic results from the KBmap as a formative assessment tool for designing feedback (Yoshida, 2013)(Pailai, 2017).

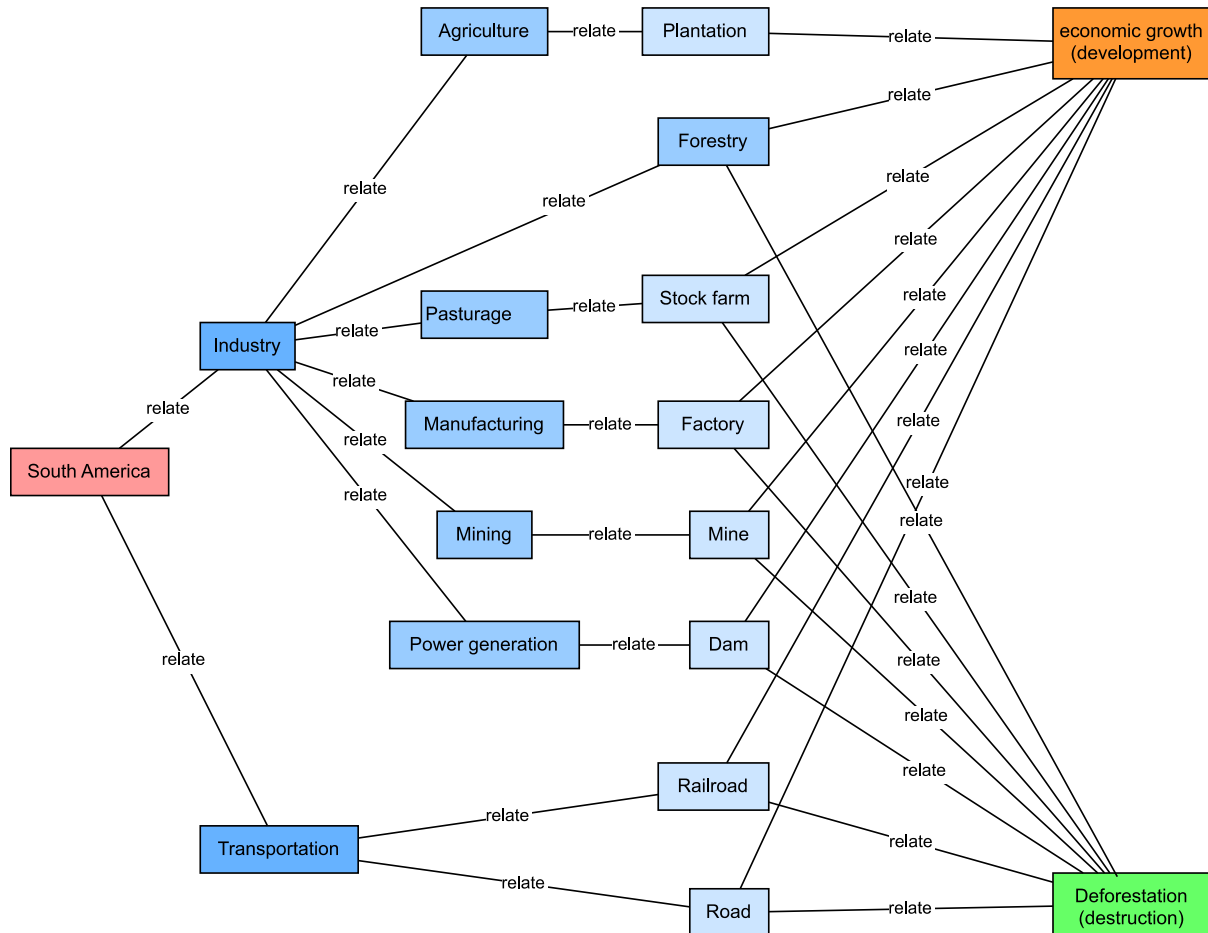


Figure 1. Goal map

2.2 KBmap system

The system for kit-build mapping is called the “KBmap system.” It is composed of two client systems, “KBmap editor” and “KBmap analyzer,” and the server system, “KBmap DB” [9]. Two types of KBmap editor work on desktop and tablet computers. The KBmap editor on desktop computers is for both teachers and learners. Teachers can create goal maps and kits for learners, and learners can make their map using a kit on the desktop version.

On the other hand, the tablet version is solely for the learners. Figure 2 shows the screenshot. The left side shows a kit that includes separate nodes and links. The right side is the KBmap made from the kit. A key characteristic of the tablet version is its portability. Thus, it can be used not only in computer rooms but also in normal classrooms. In this study, lessons were conducted in a normal classroom using a Wi-Fi network.

The KBmap analyzer works only on web browsers on PCs and tablet computers. The function of the KBmap analyzer is to show the group map that overlays the learner-maps. On this map, the more the number of learners that define a link between particular concepts, the more emphatically the link is displayed. On the window, links that many learners’ sets are displayed as thick and high-colored lines;

only few sets are displayed as narrow and light-colored lines. In addition to that, the KBAnalyzer shows each link made by the learners. With this information, the teacher can identify which links in the goal map are difficult for the learners to identify when they make their maps from a kit (Sugihara, Osada, Nakata, Funaoi & Hirashima, 2012).

The KBmap is applicable to a variety of subjects: science in elementary schools (Hirashima, Yamasaki, Fukuda & Funaoi, 2011; Hirashima, Yamasaki, Fukuda, & Funaoi, 2015), geography in junior high schools (Nomura, 2014), English as a second language (Alkhateeb, Hayashi, & Hirashima, 2015), and university-level social science and computer science (Hayashi & Hirashima, 2014)(Hayashi & Hirashima, 2015).

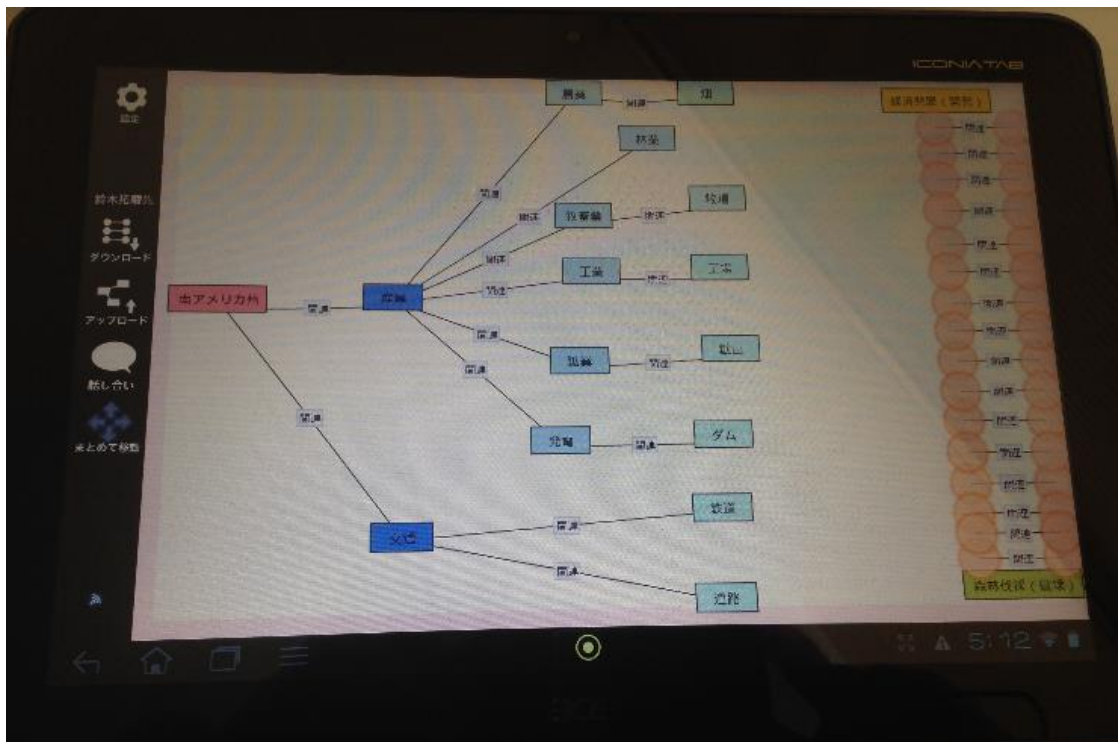


Figure 2. Kit made from the goal map

2.3 The procedure of the lesson

In the lessons, students performed the two activities below:

- organize the knowledge acquired from the previous lessons as a concept map from the kit, and
- compare and validate their understanding as represented on the KBmap editor through small-group and whole-class discussion.

Subsequently, the teacher explained the lesson he intended to impart based on the comparison of the goal map and the students' KBmaps. Through this, students were expected to resolve the lapses in comprehension following discussion.

The lessons were composed of the following steps:

Teacher reviewed the previous lessons with the students

Using some pictures, the teacher reminded the students of the features of Latin America that were learned in the previous classes. The teacher integrated all the separate lessons from each class in the form of a concept map. In addition to that, the teacher provided some pictures related to these features. This enables the students to have a concrete image of the features.

Individuals build concept maps (pre-map)

Using a tablet computer, each student individually completes the concept map representing the relation of the features to economic development and environmental degradation in Latin America within a set time. Here, individual students considered the relation based on what they learned in the previous

classes and the pictures on the tablet. Although the teacher had explained some of the relations, he had not emphasized them. Therefore, these were not mere recall tasks.

Groups build a concept map (collaborative map)

Just after building the pre-maps, without verifying the answers, the students went into small groups of four or five. In each group, the students discussed the correlation between the features of economic development and environmental degradation in Latin America. The goal of the group in this step was to reach a consensus, and build a concept map (collaborative map). Each group built a collaborative map representing the consensus reached from the personal maps of the members.

Individuals modify their pre-maps (post-map).

After the group map was built, and before it was reviewed by the teacher, students were allowed to revise their maps. The students could modify their thoughts to reflect the group discussion. The purpose of this step was to enable the learners be aware of their change of thought and record the change.

Whole-class instruction by the teacher (group map)

Finally, the teacher provided the correct answer based on the result of the groups. Although it is desirable for the teacher to explain the correct answer in detail, he simply provided an overview, because of classroom time constraint.

3. Data from a case of in-class collaborative learning

We conducted three lessons in three classes in a junior high school in Japan. The participants were 76 first-grade students (12-13 years old). The lessons were conducted in regular classrooms; the usual tasks with paper worksheets were replaced by tasks using the KBmap system on tablet computers. We made no comparison between paper worksheets and tablet computers, because the purpose of this study was not to measure the learning gain but to investigate the effectiveness of the KBmap system for detecting changes in learners' perspectives. The data captured during and after class using the KBmap system are presented in the next subsection.

3.1 Group map

During the classes, the teacher used only the aggregation of the collaborative maps, and was able to identify the commonly misunderstood area. Thus, the whole-class instruction was structured around it. According to our initial plan, the teacher was to use the group maps composed of the individual pre-maps. However, this was a challenge, because of time constraint .

Figure 3 shows the group maps, emphasizing the links over half of the group have failed to set in the collaborative-maps. In the goal map, all the industries are linked to both development and deforestation. Therefore, through missing links, the correlation many of the learners were not aware of could be identified. Based on the group maps, the teacher provided feedback to the students in the whole-class instruction.

As shown in Figure 3, the missing links varied from class to class. Before the class, the teacher assumed that majority of the students could not draw the correlation among the new concepts, railways, and roadways. However, according to the group maps, such prediction is not always accurate. Furthermore, the teacher did not foresee the lack of understanding of various propositions. A significant number of the students had no knowledge of some of the propositions. These propositions are as follows. In Class A, "Forestry is related to economic growth" and "Factory is related to deforestation"; in Class B, "Stock farm is related to economic development" and "Stock farm is related to deforestation"; in Class C, "forestry is related to economic developments" and "Mine is related to deforestation". Thus, after concept mapping, the teacher could focus solely on expatiating on these propositions. After the lessons, the teacher revealed that the approach made it possible to efficiently evaluate the students' understanding. Previous attempts to investigate the thoughts of individual students during group work had required the help of several other teachers, and organizing the result during lessons was a challenge. However, using the proposed method, the teacher was able to efficiently assess the learners' understanding, and structure the lesson around it.

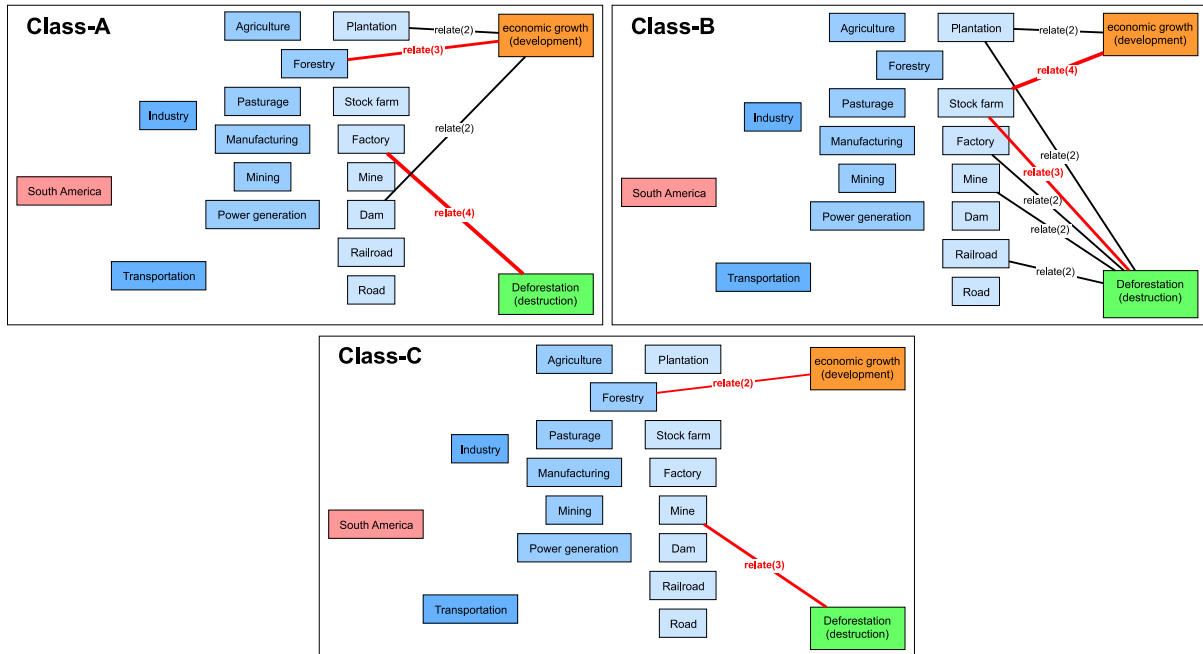


Figure 3. Group maps

3.2 Change of map score

Figure 4 and Table 1 show the comparison between the average scores of the pre-maps and the post-maps. The score of the learner's map reveals the degree of similarity between the learner's map and the goal map. It takes a value ranging from 0 to 1. If the score is 1, the learner's map is exactly the same as the goal map. The score is calculated using the following equation:

$$\text{mapscore of a learner's map} = \frac{\text{the number of the same propositions in the learner's map and the goal map}}{\text{the number of the propositions in the goal map}}$$

Table 1
Map Scores

	Pre-map	Collaborative-map	Post-map
Class-A	0.5817 (SD=0.2206)	0.8125 (SD=0.1936)	0.7764 (SD=0.1922)
Class-B	0.5885 (SD=0.2395)	0.7708 (SD=0.1407)	0.7839 (SD=0.1758)
Class-C	0.6400 (SD=0.2226)	0.8958 (SD=0.1164)	0.8950 (SD=0.0967)

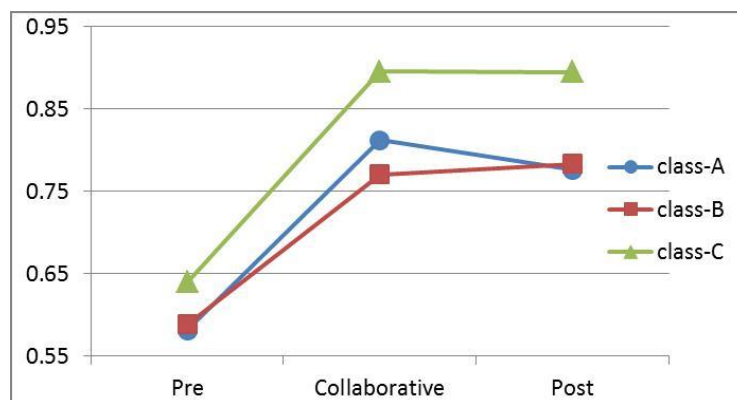


Figure 4. Average map scores in three classes

The average scores of the post-maps exceeded those of the pre-maps in every class. Based on the Wilcoxon signed-rank test, the difference is significant (Class A: $n = 26$, $V = 210$, $p < 0.01$; Class B: $n = 24$, $V = 148.5$, $p < 0.01$; Class C: $n = 25$, $V = 253$, $p < 0.01$). This result indicates that creating the KBmap in a group improved the students' understanding.

Furthermore, this study compared the score of the pre-map to the score of the collaborative map using the Wilcoxon rank-sum test. A significant difference was found between the pre-map score and the post-map score in every class (Class A: $n = 26$, $U = 31.5$, $p < 0.05$; Class B: $n = 24$, $U = 34$, $p < 0.01$; Class C: $n = 25$, $U = 34.5$, $p < 0.01$). The scores of the collaborative maps exceeded those of the pre-maps. This finding suggests that the knowledge attained collaboratively was more accurate than the average knowledge attained individually.

Consequently, the closer the pre-map was to the collaborative map, the higher the degree of the similarity among the group members' maps, and because the score of the collaborative map exceeded that of the pre-map, the students' knowledge improved.

3.3 Types of map change following small-group discussion

Figure 5 shows the classification of the group according to hierarchical cluster analysis. There are four types of group: advanced and improvement, follow-up and raise, collective and improved, and collective and raise.

Cluster 1 is "advanced and improved". Here, the scores of the collaborative maps exceeded that of the maximum score of the pre-maps in the group. This implies that the collaborative map was not a mere fusion of the pre-maps. Rather, the students, as a group, successfully integrated their pre-maps, and possibly found a new relation. Furthermore, the average, maximum, and minimum score of the post-maps exceed those of the pre-maps, thus demonstrating that the members' understanding is generally improved through the small-group discussion.

Cluster 2 is "follow-up and raised". Here, the scores of the collaborative maps are the same as the maximum score of the pre-maps. Each group in this cluster adopts the maximum score of the pre-maps as their collaborative map. Finally, although the average and minimum scores of the post-maps of the groups increased, the maximum score remained the same. Specifically, in Groups B1, B4, and C3, all the members' scores following the post-maps are equal to the collaborative map. In this cluster, it was possible for some students to improve their understanding through the small-group discussion.

Cluster 3 is "collective and improved". This cluster is similar to Cluster 1. However, the scores of the collaborative maps were lower than the maximum score of the pre-maps. Although it was not possible to fully collate the acquired knowledge of the group members, they could all achieve improved comprehension.

Cluster 4 is "collective and raised". This cluster is similar to Cluster 3, as the scores of the collaborative maps were not the same as the maximum score of the pre-maps in the group. Although the average scores of these groups' post-maps increased, the maximum score and even the minimum scores remained consistent in several groups. In this cluster, some students were able to improve their understanding through the small-group discussion.

The characteristics of these clusters as stated above depend on the membership of the group. For example, Groups B1, B2, C3, and C6 consisted of individual learners that earned full marks in the pre-map scoring. Thus, these scores could not be exceeded in the post-test. It was however possible for learners to obtain a lower score at the post-map. Their score is not changed, and the others adopt some answer to him/her.

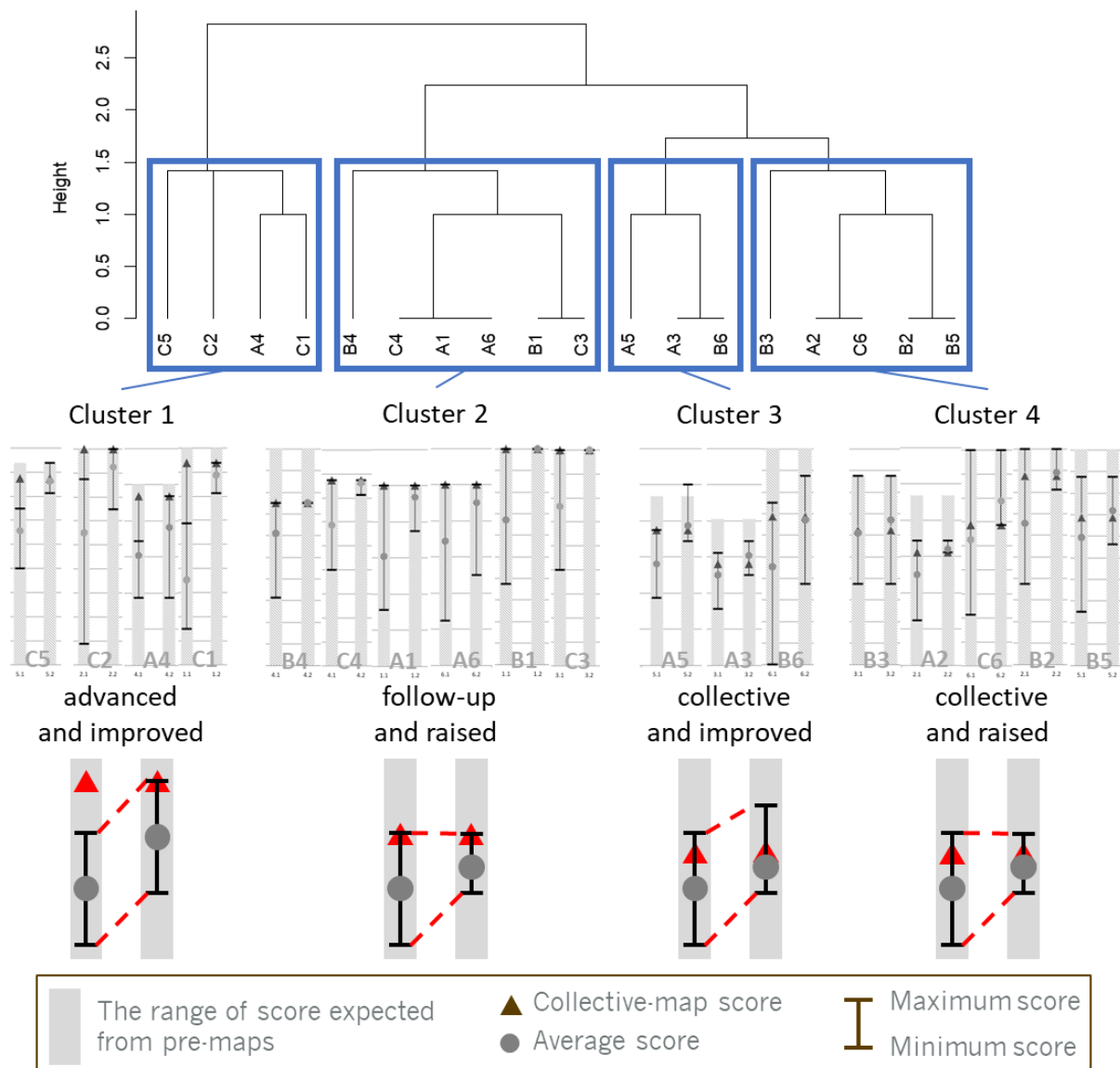


Figure 5. Clusters of groups

4. Conclusion

This paper presents the result of a case study of the implementation and data analysis of in-class collaborative learning using the KBmap. The KBmap automatically evaluates the concept maps of individual learners based on the teacher's goal map. It is a potential solution to the challenges of implementing concept mapping in the classroom. The data captured using the KBmap system, such as correctness and similarity, can be used during and after classes.

Such data can be used for formative and summative evaluation. The teacher in this study was able to qualitatively assess the students' understanding, rather than quantitatively, and provide feedback accordingly. Typically, it is difficult for teachers to monitor the learning and comprehension process during collaborative learning. The conventional methods of acquiring this knowledge include allowing the students to give a presentation, or carefully scrutinizing their conversation. Although the KBmap provides a kit for concept mapping and delineates the bounds of the map, the small-group and whole-class discussion is not limited to the boundary. In other words, the KBmap itself is a close-ended learning environment. However, depending on implementation and the activities, a class using the KBmap can be open-ended. For example, in this study, there are many reasons to identify the correlation between industry and development or deforestation. In such a case, the KBmap can assess students' understandings at a minimum level where the teacher wants to share with them. Naturally, the teacher allows for the diversity of perspectives depending on the topics; such can be covered through

discussion or other methods. This study demonstrates the possibility of capturing and analyzing data during in-class collaborative learning, the consequent realization of formative and summative evaluation, and the corresponding feedback during and after class. In the future, we aim to clarify the range of capture and analysis of collaborative learning with the KBmap, and to develop functions for evaluation and feedback by teachers.

References

- Alkhateeb, M., Hayashi, Y., and Hirashima, T. (2015) Comparison between Kit-Build and Scratch-Build concept mapping methods in supporting EFL reading comprehension. *Journal of Information and Systems in Education*, 14(1), 13–27.
- Dillenbourg, P. and Jermann, P. (2010) *Technology for Classroom Orchestration*, New Science of Learning, pp. 525–552, Springer, New York.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485–492
- Hayashi, Y. and Hirashima, T. (2014) Kit-Build concept mapping for being aware of the gap of exchanged information in collaborative reading of the literature. *Proc. Int. Conf. on Human Interface and the Management of Information*, 31–41.
- Hayashi, Y. and Hirashima, T. (2015) Analysis of the relationship between metacognitive ability and learning activity with Kit-Build concept map. *Proc. Int. Conf. on Human Interface and the Management of Information*, 304–312.
- Herl, H. E., O'Neil, H. F. Jr., Chunga, G.K.W.K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Comput. Hum. Behav.*, 15, 315–333.
- Hirashima, T., Yamasaki, K., Fukuda, H., and Funaoi, H. (2011) Kit-Build concept map for automatic diagnosis. *Proc. of 15th Conf. on Artificial Intelligence in Education*, 466–468.
- Hirashima, T., Yamasaki, K., Fukuda, H., and Funaoi, H. (2015) Framework of Kit-Build concept map for automatic diagnosis and its preliminary use. *RPTEL*, 10(1), 1–21.
- Hoppe, H.U. (2017) Computational methods for the analysis of learning and knowledge building communities, *Handbook of Learning Analytics First Edition*, <https://solaresearch.org/hla-17/hla17-chapter2/>
- Martinez-Maldonado, R., Clayphan, A., Kay, J. and Yacef, K. (2014) Towards providing notifications to enhance teacher's awareness in the classroom. *ITS*, 510–515.
- Matsuzawa, Y., Oshima, J., Oshima, R., Niihara, Y., Sakai, S., (2011) KBDeX: A Platform for Exploring Discourse in Collaborative Learning *Procedia - Social and Behavioral Sciences*, Vol. 26, pp. 198–207.
- Nomura, T., Hayashi, Y., Suzuki, T. & Hirashima, T. (2014) Knowledge propagation in practical use of Kit-Build concept map system in classroom group work for knowledge sharing. *ICCE2014 Workshop Proceedings*. pp. 463–472.
- Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *IHMC*, 1.
- Oshima, J., Oshima, R., Matsuzawa, Y. (2012) Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse, *Educational Technology Research and Development*, Volume 60, Issue 5, pp. 903–921, 2012.
- Pailai, J., Wunnasri, W., Yoshida, K., Hayashi, Y., and Hirashima, T. (2017) The practical use of Kit-Build concept map on formative assessment. *RPTEL*, 12:20. <https://doi.org/10.1186/s41039-017-0060-x>.
- Ruiz-Primo, M.A. (2004). Examining Concept Maps as an Assessment Tool. *Proc. of the First Int. Conference on Concept Mapping*, Pamplona, Spain. Available: <http://cmc.ihmc.us/papers/cmc2004-036.pdf>
- Stahl, G. (2000). A model of collaborative knowledge-building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences*, pp. 70–77), Mahwah, NJ: Erlbaum.
- Sugihara, K., Osada, T., Nakata, S., Funaoi, H., and Hirashima, T. (2012) Experimental evaluation of Kit-Build concept map for science classes in an elementary school. *Proc. ICCE 2012*, 17–24.
- Tergan S-O. (2005). Digital concept maps for managing knowledge and information. In *Knowledge and Information Visualization*, Sigmar-Olaf Tergan and Tanja Keller (Eds.). Springer-Verlag, Berlin, Heidelberg 185–204.
- Yamasaki, K., Fukuda, H., Hirashima, T., & Funaoi, H. (2010). Kit-build concept map and its preliminary evaluation. In *Proc. of ICCE*, 290–294.
- Yoshida, K., Sugihara, K., Nino, Y., Shida, M., and Hirashima, T. (2013) Practical use of Kit-Build concept map system for formative assessment of learners' comprehension in a lecture. *Proc. ICCE 2013*, 906–915.

An Exercise Recommendation Method for K-12 Students Based on the Syllabus

Zhizhuang Li, Zhengzhou Zhu, Teng Yan,
Peking University, China
lizhizhuang@pku.edu.cn, zhuzz@pku.edu.cn

Abstract: For each subject, the syllabus specifies what knowledge points students should master and how well they should master them. Whether students can acquire the ability required by the syllabus is an important evaluation criterion for a student's learning achievements. However, most of the exercises for k-12 students at present contain knowledge points in the examination syllabus, but the difficulty of the examination of knowledge points is inconsistent with the teaching syllabus. Solving and studying such exercises does not help students acquire the skills required by the syllabus quickly. Therefore, students should choose exercises that fit well with the syllabus when practicing. However, at present, there is no way to recommend exercises that fit well with the syllabus for students, so as to improve learning efficiency. In order to solve the above problems, this paper proposes an exercise recommendation method for k-12 students based on the syllabus. It is proved by experiment that this method can recommend exercises that fit well with the syllabus for students, so as to improve their learning efficiency. The use of this method can help students achieve better learning outcomes and achieve higher scores in the final examination than ordinary students under the condition of learning and mastering the same number of exercises with the same difficulty.

Keywords: Exercise recommendation, Machine learning, Linear regression, EM algorithm

1. Introduction

At present, the problem of overloaded homework for k-12 students has become an important problem to be solved. In order to reduce the learning burden of students without reducing the learning effect of students, it is necessary to improve the fit between exercises and the teaching syllabus on the degree of mastery of various knowledge points, so as to help students avoid inefficient learning.

The syllabus for any course sets the requirements for each knowledge point. For different knowledge points, the syllabus has different requirements for students to master. If students can practice and learn by using exercises with the same difficulty in each knowledge point as in the syllabus, their learning efficiency will be improved and their learning burden will be reduced. However, the exercises in the current question bank often only show the knowledge points that the exercises examine, but do not show the depth of the exercises examine to the knowledge points. There are a lot of exercises that examine the knowledge points that are included in the syllabus but the depth of the examination of the knowledge points is different from the exercises that are required by the syllabus. For example, the syllabus requires that the assessment of knowledge point N1 is proficiency, and the assessment of knowledge point N2 is understanding. Now we have exercises Q1, Q2, Q3, Q4. Q1 requires proficiency in both knowledge points. Q2 requires understanding of both knowledge points. In question Q3, the assessment requirement for knowledge point N1 is understanding, and the assessment requirement for knowledge point N2 is proficiency. In question Q4, the assessment requirement for knowledge point N1 is proficiency, and the assessment requirement for knowledge point N2 is understanding. Of the 4 questions, only Q3 is consistent with the syllabus.

The existing exercises recommendation method cannot help teachers and students to screen out the exercises that fit the syllabus. In order to solve this problem, this paper proposes a method based on the syllabus to recommend exercises for k-12 students.

The contributions of this paper are as follows :(1) we propose RAE algorithm for cognitive diagnosis of students, which combines linear regression and EM algorithm to calculate the degree of each student's mastery of each knowledge point. (2) we put forward the KPLL algorithm, which can

calculate the inspection degree of each exercise to each knowledge point. (3) this paper proposes a method to recommend exercises that fit well with the syllabus for students. It has been proved by experiments that using this method can help students achieve better learning results and get higher scores in the final examination when they learn and master the same number of exercises with the same difficulty.

2. Related Works

The current exercise recommendation methods are mainly based on two aspects: to recommend exercises for target students that contain weak knowledge points for target students, and to recommend exercises that are moderately difficult for target students. Y. Huo et al. proposed an exercise recommendation method based on collaborative filtering. This method can help students find weak knowledge points and recommend exercises containing weak knowledge points [1]. L. Fang proposed an exercise recommendation method that can recommend English exercises containing specified grammatical structures for students [2]. Changmeng J et al. proposed to recommend exercises of medium difficulty for target students according to the scope of knowledge points selected by students [3-6].

3. Method Design

The purpose of this method is to recommend the exercises that are consistent with the syllabus and contain the knowledge points designated by the target students to the target students. Since the k-12 final exam is formulated by the proposition expert strictly according to the syllabus, the requirements of the final exam on the mastery of various knowledge points can be approximately regarded as corresponding to the syllabus. For each question, the difficult questions on the final exam will be no more difficult than the syllabus requires, and the simple questions on the final exam will be less difficult than the syllabus requires. Therefore, for a specified knowledge point, we regard the requirement of the knowledge point on the final exam with the highest level of requirement as the requirement of the syllabus. First of all, we judge how well the syllabus requires students to master various knowledge points by their answers to the final examination questions. Then we choose the exercises that fit into the syllabus as candidate exercises. According to the range of knowledge points assigned by students, we recommend candidate exercises that meet the standards to target students. The specific steps of this method are as follows:

(1) We select a group of students who use the same syllabus, and extract the data of these students in solving daily exercises and solving final exams. Taking the data of these students' daily exercises as input, we use RAE method (the algorithm is described in detail in 3.1) to calculate the degree of each student's mastery of each knowledge point.

(2) We divide students' mastery of each knowledge point into five levels from low to high. Level 5 corresponds to the degree of mastery of the knowledge in the top 20% of the students, level 4 corresponds to the mastery of the knowledge degree ranking 20% ~ 40% of the students, level 3 corresponds to the mastery of the knowledge degree ranking 40% ~ 60% of the students, level 2 corresponds to the mastery of the knowledge degree ranking 60% ~ 80% of the students, level 1 corresponds to the mastery of the knowledge degree of 80% ~ 100% of the students.

(3) With each student's mastery of each knowledge point and the data of all students' answers to all the final examination questions as input, using the KPLL algorithm (which is described in detail in 3.2), we calculate the mastery level required by each final examination question for each knowledge point contained in it. For example, the level of knowledge required for knowledge point N in question Q is L, which means that only students with a level of knowledge point N no lower than L are likely to correctly solve question Q.

(4) For each knowledge point, we find out the final exam questions with the highest requirements for the mastery level of the knowledge point. We regard the mastery level required by this question for this knowledge point as the mastery level required by the syllabus for this knowledge point.

(5) With each student's mastery of each knowledge point and the data of all students' solutions to all exercises as input, we use KPLL algorithm to calculate the required mastery level of each exercise for each knowledge point contained in it.

(6) Choose the exercises that meet the following conditions as candidate exercises: this exercise requires the same level of mastery of all knowledge points as the syllabus requires for this knowledge point.

(7) According to the knowledge points selected by the students, we recommend the exercises containing the specified knowledge points from the candidate exercises.

3.1 RAE algorithm

The input of this algorithm is the historical answer data of all students, and each data includes student number, exercise number, the number of knowledge points contained in exercise, exercise score, and actual score. The output of this algorithm is each student's mastery of each knowledge point. First, define the concept of knowledge points mastery index. The concept of knowledge point mastery index is used to measure the degree of students' mastery of a certain knowledge point above or below the average level.

The formulas for calculating the knowledge points mastery index are defined respectively when the weight of knowledge point is unknown or known.

The formula for calculating knowledge point mastery Index G_{SK} of knowledge point without assigning the weight of knowledge point is as follows:

$$G_{SK} = \frac{\sum_{i=1}^n (s_i - m_i)}{\sum_{i=1}^n t_i} \quad (1)$$

Among them, s_i is the score of student S for exercise i, an exercise involving knowledge point K that he/she has done, m_i is the average score of exercise i, and t_i is the total score of exercise i.

The formulas for calculating knowledge point mastery Index G'_{SK} of knowledge point under the condition that the weight of knowledge point has been assigned are as follows:

$$G'_{SK} = \frac{\sum_{i=1}^n (s_i - m_i) k_i}{\sum_{i=1}^n t_i k_i} \quad (1)$$

Among them, s_i is the score of student S for exercise i, an exercise involving knowledge point K that he/she has done, m_i is the average score of exercise i, k_i is the weight of knowledge point K in exercise i, and t_i is the total score of exercise i.

Based on a pre-selected set of student answer data, the steps for cognitive diagnosis are as follows:

First, we choose questions with only a single knowledge point, use formula (1) to roughly calculate and store the knowledge point mastery Index of each student's all knowledge point.

We use the linear regression model to calculate the weight of each knowledge point in each question. We illustrate the use of the linear regression model with example question e and student s. We assume that there are n knowledge points in question e, student s's score prediction formula for question e is as follows:

$$h_i = \sum_{j=1}^n k_j x_j \quad (3)$$

h_i is the predicted score of the assigned question e for student s, x_j is the knowledge point mastery Index of the student s for the jth knowledge point of question e, and k_j is the weight of x_j . We use all the student history answer data to train the model, then we can get the weight of each knowledge point in each question.

We use formula (2) to calculate the more accurate knowledge point mastery Index of each student for each knowledge point when the weight of each knowledge point is known.

Using the EM algorithm, the previously calculated value of knowledge point mastery Index is taken as the current value of students' knowledge point mastery Index. Each time, the current value of knowledge point mastery Index is used to calculate the weight of knowledge point in a new round of exercises by linear regression. Then, the weight of knowledge point in the new round of exercises is used to calculate the knowledge point mastery Index of the new round of students through formula (2). After repeated iteration for 10 rounds, the final result of knowledge point mastery Index was obtained.

3.2 KPLL algorithm

The input of this algorithm is each student's mastery of each knowledge point and the historical answer data of all students. Each data includes student number, exercise number, the number of

knowledge points included in the exercise, exercise score, and actual score. The output of this algorithm is the requirement of each question for each knowledge point.

This algorithm calculates the requirement of each question to each knowledge point. We illustrate this algorithm by using exercise Q and knowledge point K. First, we find out who has a level 5 command of all knowledge points except K contained in Q. (if there are no students meeting the above conditions, the requirement for the target students will be reduced to those who have mastered at least level 4 in all knowledge points except K included in the exercise Q) According to the mastery level of knowledge point K_i , all students are divided into 5 categories. Calculate the average score of each type of students for exercise Q. Considering the factors such as carelessness, we select categories which scored average of more than 80% of exercise Q's total score. We mark the mastery level of knowledge point K required by the category with the lowest mastery level of knowledge point K among all selected categories as the mastery level required by exercise Q for knowledge point K.

4. Experiment

4.1 Experiment Environment

The data used in this experiment were provided by an online education company. Considering that China implemented a nationwide textbook update for primary and secondary schools in September 2017, the teaching syllabus also changed after the textbook update. If we use the answer data after September 2017, we will not be able to obtain enough final exams that fit into the syllabus. Therefore, the data we used is the online answer data of all the fifth-grade primary school students in a Chinese city from March 1 to April 1, 2017 for a certain chapter of mathematics. These data include the responses of 8,092 students to 1,149 questions, totaling 240,000 pieces of data. Each piece of data includes student id, exercise id, exercise type, id of knowledge points included in exercise, total score of exercise, actual score, starting time of exercise and ending time of exercise. The data for the experiment included two types of exercises, one for the final exam from 2012 to 2016 in each district of the city, and the other for general exercises. The data for this experiment includes 197 final exams and 952 general exercises.

4.2 Experiment Process

First of all, we divide all the answer data into two groups: the first group is the data of March 1, 2017 solstice, March 15, 2017, and the second group is the data of March 16, 2017 solstice, April 1, 2017. We use the method proposed in this paper, take the first group of data as input data, calculate the degree of each student's mastery of each knowledge point, and select the candidate exercises that fit the syllabus from the general exercises.

We selected students who correctly answered a certain number of candidate exercises in the first set of data and students who correctly answered the same number of general exercises and compared their performance on the final exam questions in the second set of data. By comparing the gap between the two, we can judge whether the candidate exercises selected by the method proposed in this paper have a higher fit with the syllabus. The specific steps are as follows:

(1) We randomly selected three knowledge points and recorded them as experimental knowledge points. In the first set of data, we found all the exercises that contained only one or several of the experimental knowledge points and recorded them as experimental exercises. We randomly selected five questions from the final exam from the second data set and recorded them as experimental examination questions. These five questions cover all the experimental knowledge points and are used to examine the recommendation effect.

(2) According to the classification criteria of whether the exercise is a candidate exercise, we divide all the exercise into two groups, which are respectively denoted as exercise group 1 and exercise group 2. The exercises in exercise group 1 are candidate exercises, while the exercises in exercise group 2 are not.

(3) We first select the students who have done all the experimental examination questions during March 16, 2017 solstice April 1, 2017. Among the selected students, we found out all the students who had done 3 exercises belonging to exercise group 1 and all of them answered correctly, all students who had done 4 exercises belonging to exercise group 1 and all had answered correctly, all

students who had done 5 exercises belonging to exercise group 1 and all had answered correctly, and mark them as student group A1, student group A2 and student group A3 respectively.

(4) Among the selected students, we found out all the students who had done 3 exercises belonging to exercise group 2 and all of them answered correctly, all students who had done 4 exercises belonging to exercise group 2 and all had answered correctly, all students who had done 5 exercises belonging to exercise group 2 and all had answered correctly, and mark them as student group B1, student group B2 and student group B3 respectively.

(5) We found out the experimental exercises done by students in the above six groups, and calculated the average score rate of all other students for these exercises (average score/total score of exercises). We measure the difficulty of a problem by the average score rate of all the other students on the problem. If the average score rate of all other students for problem Q is s , the difficulty of problem Q is defined as $1-s$. We randomly remove some students in student group B1, student group B2 and student group B3 who have done the experiment exercises with low difficulty. In this way, the average difficulty of experimental exercises done by students in student group B1, student group B2 and student group B3 is not lower than that of student group A1, student group A2 and student group A3, respectively.

(6) Calculate the average scores of 6 groups for all experimental examination questions.

4.3 Experiment Results

As shown in figure 1, we made statistics on the number of students in each group, the average difficulty of the experimental exercises done by all members in each group, and the average score of all members in each group on the experimental examination questions (converted into a full score of 100 points). We marked the average difficulty of all experimental exercises done by all members of the group as the average difficulty, and the average score of all members of the group for the experimental examination questions as the average score.

table 1

Statistical table of experimental data

Student group	Number of students	Average difficulty	average score
A1	17	0.21	89.3
B1	33	0.23	82.5
A2	11	0.19	92.7
B2	25	0.21	86.5
A3	9	0.22	94.1
B3	14	0.24	88.0

According to table 1, the average difficulty of the experimental exercises done by group A1, A2 and A3 is lower than that of group B1, B2 and B3, respectively. The average scores of group A1, A2 and A3 for the experimental examination questions were higher than those of group B1, B2 and B3, respectively. Therefore, students who do well on the same number of problem sets of the same difficulty are more likely to do well on the final exam than those who do well on the ordinary problem sets. As a result, students who did well on the same number of candidate exercises with same difficulty, which were recommended by this method, were more likely to do well on the final exam than those who did well on the regular exercises. It can be inferred that mastering the exercises recommended by this method is more likely to help students achieve good results in the final examination than mastering the common exercises with the same number and difficulty. Therefore, we can say that using this method can help students achieve better learning results than ordinary students and achieve higher scores in the final examination under the condition of learning and mastering the same number of exercises with the same difficulty. Since the requirement of mastering various knowledge points in the final examination

can be approximately regarded as a fit with the examination syllabus, we can infer that this method can recommend exercises that fit with the teaching syllabus for students.

5. Conclusions

This paper proposes an exercise recommendation method for k-12 students based on the syllabus. It is proved by experiment that this method can recommend exercises that fit well with the syllabus for students, so as to improve their learning efficiency. The use of this method can help students achieve better learning outcomes and achieve higher scores in the final examination than ordinary students under the condition of learning and mastering the same number of exercises with the same difficulty.

Acknowledgements

This paper was supported by National Key Research and Development Program of China (Grant No. 2017YFB1402400), Ministry of Education "Tiancheng Huizhi" Innovation Promotes Education Fund (Grant No. 2018B01004), National Natural Science Foundation of China (Grant No. 61402020, 61573356), and CERNET Innovation Project (Grant No. NGII20170501).

References

- [1] J. Xia, G. Li and Z. Cao (2018). "Personalized exercise recommendation algorithm combining learning objective and assignment feedback," *Journal of Intelligent and Fuzzy Systems*, vol. 35, (3), pp. 2965-2973.
- [2] Huo, Yujia, Jiang Xiao, and Lionel M. Ni (2018). "Towards Personalized Learning through Class Contextual Factors-Based Exercise Recommendation", 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), vol. 2018-, pp. 85-92.
- [3] Yanxia P, Wenyan L, Yuanyuan J, et al. (2018). Adaptive recommendation for MOOC with collaborative filtering and time series[J]. *Computer Applications in Engineering Education*, 2018.
- [4] Martin Labaj. (2014). Utilization of Exercise Difficulty Rating by Students for Recommendation. *International Conference on Web-Based Learning*, 2014 :13-22.
- [5] Changmeng J, Jun F, Xia S, et al. (2018). Personalized exercises Recommendation algorithm based on Knowledge Hierarchical Graph, ReKHG[J]. *Computer Engineering and Applications*.
- [6] L. Fang et al (2018). "Personalized question recommendation for English grammar learning," *Expert Systems*, vol. 35, (2), pp. e12244-n/a. Conference Name: ACM Woodstock conference

An Analysis of Learning Behavior Patterns with Different Devices and Weights

Chengjiu YIN^{a*}, Juan ZHOU^b, Lingyu LI^a, Koudai Yamaguchi^a, Noriko UOSAKI^c, & Hiroaki OGATA^d

^aKobe University, Japan

^bRitsumeikan University, Japan

^bOsaka University, Japan

^bKyoto University, Japan

*yin@lion.kobe-u.ac.jp

Abstract: With e-learning systems gradually being implemented, researchers worldwide have started devoting increasing attention to Learning Analytics. At Kobe university, a digital textbook reading system has been developed to collect learning logs in the face-to-face classroom. In a previous study, *k*-means clustering was implemented to analyze learning behavior patterns; however, there were problems such as few variables for clustering and a failure to consider weighting of the learning elements. Therefore, in this study we applied clustering by increasing the number of learning elements and assigned weights to the learning elements, then analyzed the learning behavioral patterns. We found some behavioral patterns of students who can save learning time if they effectively write memos and add markers.

Keywords: Learning behaviors, *k*-means clustering, learning log, digital textbooks

1. Introduction

In recent years, owing to the development and spread of PCs and mobile devices, many educational institutions have been investigating and implementing e-learning systems. In fact, 33.8% of the universities in Japan had instituted Learning Management Systems (LMS) in preliminary learning and post-learning in 2010. This value had risen to 44.9% in 2013. On the other hand, 36.3% of the universities had implemented blended learning (a kind of learning method that combines real courses with e-learning in which learners can use a real course and web group course together) in 2010, a figure that had risen to 44.0% in 2013.

With e-learning systems gradually being implemented, researchers worldwide have started devoting increasing attention to Learning Analytics (LA) (Yin & Hwang, 2018). Researchers have reported that LA can be used for learning support and class improvement (Mostow, 2004; Yin et al., 2014; Yin et al., 2015), and LA is positively related to student efforts (Campbell, DeBlois, & Oblinger, 2007), performances (Macfadyen & Dawson, 2012), and outcomes (Archer, Chetty, & Prinsloo, 2014; Hrastinski, 2009; Yin & Hwang, 2018). Yamada (2017) pointed out that in LMSs, administrators and teachers usually pay attention to the logs used to detected abnormal activities, such as unauthorized access and system failure. The system-use logs are frequently neglected. However, LA pays attention to the system-use logs, which are used to analyze learners' behavior to find ways to improve teaching and support learning. For example, some researchers used learning logs of online courses from MOOCs (Massive Open Online Courses) and OERs (Open Educational Resources) to perform learning analytics.

Many recent studies have focused on collecting data from online performance for learning analytics; in contrast, the application of learning analytics in face-to-face classes is rare worldwide (Yin et al., 2018). Therefore, to collect learning logs in face-to-face classes, Yin et al. (2017) developed an electronic teaching material system, DITeL (Digital textbook for Improving Teaching and Learning), and then, using learning logs collected by DITeL, *k*-means clustering was used to divide learners into clusters and analyze the relationship between learners' learning behavioral patterns and learning achievement (Yin et al., 2018).

However, few learning elements were used for clustering in the previous study, and weighting of learning elements was not considered. To solve these two problems, in this study we increased the number of learning elements, such as markers and devices, to apply learning analytics more effectively. As different learning elements have different learning effects on academic performance, we also assigned a weight to each element and divided learners into clusters according their learning logs. Cluster analysis was then performed with the statistical analysis software package SPSS (developed by IBM), and we analyzed and discussed the relationship between learning elements and learning achievement. We found some behavioral patterns for such groups as students who can save learning time if they effectively write memos and add markers.

2. Literature Review

Data analysis is an internal step in the process of data collection (Yin, Hirokawa, et al., 2013; Yin, Sung, et al., 2013). In this study, the system collects data automatically while the learner is using the system, therefore the data are objective (Yin et al., 2018).

2.1 Clustering

There are many other data mining methods used in educational research, such as Apriori and SVM. In this study, k -means clustering was adopted to group large and diverse data into groups. The variables have similar values in the same group significantly different from the values for the other groups; therefore, k -means clustering is better than other methods (Yin et al., 2018).

In a previous study, four learning elements (reading time, the number of pages viewed, page backtracking rate, and the times of preview lessons) were used to divide the data into clusters (Yin et al., 2018). However, only a few learning elements were used in the clustering, although there are other learning elements that also affect learning achievement. Therefore, in this paper, besides the reading time, the number of pages viewed, and page backtracking rate, we added the number of Markers and the number of Memos for k -means clustering.

Liu, Feng, Shi, and Guo (2014) insisted that weighting should be performed when dividing data into clusters. This study assigned a weight to each learning element before clustering. A previous study (Yin et al., 2018) did not consider the weighting of different variables (learning behaviors) even though different variables have different impacts on grades. Since it was confirmed that the learning achievement has a different correlation with each learning element, it is necessary to assign a weight to each element.

3. E-Book System DITeL

As motioned above, we developed an e-book system, DITeL, to collect data in face-to-face classes. DITeL is supported on a variety of devices such as PCs, Mobiles, and Tablets, because different learners learn on different devices. In addition, DITeL has various functions to make learning more efficient.

Every user action (such as learning material name, page number, action time, and device) is recorded as a learning log. Table 1 shows an example of the learning logs from which the data are aggregated. The variables for aggregating are shown in Table 2.

Table 8
An Example of the Learning Logs Recorded on the Server

Userid	Action name	Learning material	Page Number	Action time	Device
Student1	Next	Law Course	16	2017/5/22 8:40	PC
Student1	Prev	Law Course	15	2017/5/22 8:42	Mobile
Student2	Add UnderLine	Law Course	15	2017/5/22 8:42	Tablet
Student3	Add Memo	Law Course	15	2017/5/22 8:42	Mobile

Table 2

The Variables of the Data Recorded on the Server

Elements	Define
PC Read Pages (PRP)	The number of pages the student read on a PC
Mobile Read Pages (MRP)	The number of pages the student read on a Mobile device
Table Read Pages (TRP)	The number of pages the student read on a Tablet PC
BookMark(BM)	The number of Bookmarks created by students
Memo	The number of Memos created by students
Highlight	The number of Highlights created by students
UnderLine	The number of Underlines created by students
Prev	The number of times the student returned to the previous page
Next	The number of times the student advanced to the next page
Reading Time(RT)	The number of seconds the student spent reading the teaching material
Reading Page(RP)	The number of pages the students read of the teaching material

4. Analysis of Learning Logs and Analytical Results

4.1 Problems in Analysis

In previous studies analyzing learning behavior patterns with learning logs, *k*-means clustering was applied. However, there were some problems:

1. The number of variables collected was few, and the relationships between variables were not examined at all.
2. The differences in devices and the students' tendencies to learn had not been considered.
3. The different variables have difference influences on learning achievement, so it is necessary to assign weights to the variables.

In order to solve the above problems, the relationship between learning behavior patterns and learning achievement was analyzed after assigning weights to the variables. The analytical steps are described below.

- Step 1: Data Structuring; summarizing learning logs and formatting the data.
- Step 2: Data Correlation; in order to examine the relationship between learning achievement and other variables, correlation analysis is used.
- Step 3: Assigning weights; the variables are assigned weights based on the correlation coefficients.
- Step 4: Clustering; students were clustered into groups by these weighted variables.
- Step 5: Multiple comparison test to find and objectively consider differences between groups.

4.2 Variables

Based on the variables in Table 2, new variables were also generated and unnecessary variables deleted as described below.

- Due to the small number of users, “Tablet ReadPage” and “BookMark” were not used because their influence on the results would not be accurate.
- Since “HighLight” and “UnderLine” are similar in usage, they were combined to create a new element “Marker,” and “HighLight” and “UnderLine” were not used.
-

$$Marker = HighLight + UnderLine$$

- In order to investigate the difference in consciousness between individuals, the variables “Prev” and “Next” were converted to the new variables “PrevPer” (PP) and “NextPer” (NP).

$$PP = \frac{Prev}{Prev+Next}, \quad NP = \frac{Next}{Prev+Next}$$

- “GPA” was used to represent the final grade of students.

4.3 Correlation Analysis

In this study, correlation analysis was performed to determine the degree of influence of each factor on GPA as measured by the correlation coefficient. Statistical analysis software (IBM SPSS Statistics) was used for the correlation analysis.

In this study, p-values were considered significant at the 5% level. Also, negative values were not considered, as they had negative correlations with GPA. The general results are as follows.

- 1) Since “RP” and “PRP,” “MRP” and “Prev,” and “MRP” and “Next” were strongly correlated with each other, their effects on GPA overlapped. It was found that these variables were extracted from the same element of the number of pages. In this paper, in order to focus on devices, “RP,” “Prev,” and “Next” were not considered, but “RT” was retained because it was strongly correlated with GPA, and because it represents the time of extracting variables differently.
- 2) Since “NP” had negative correlations with all variables, it was not considered in this research.

4.4 How to Assign Weights to Variables?

Correlation analysis found that “GPA” could be modeled by six variables: “Memo,” “Marker,” “RT,” “PP,” “PRP,” and “MRP.” Considering the relationships between them, these six variables were roughly divided into three groups. We defined learning with “Memo” and “Marker” as plus- α learning, “RT” (reading time) and “PP” (rate of return to the previous page) were learning traces, and “PRP” (read pages using PC) and “MRP” (read pages using mobile devices) were devices. These same groups did not affect each other. We then analyzed the relationship between those groups and found two sufficient conditions:

1. If the value of device is larger, then learning trace and plus- α learning are larger. For example, if the student reads many pages of content, then he will return to the previous page many times and write memos or add markers many times.
2. If the value of learning trace is larger, then positive plus- α learning is larger. For example, if the student reads the content for a long time, then he will write memos or add markers many times.

The weights were calculated based on these two sufficient conditions. For example, “GPA” was correlated with “Memo” with a strength of 0.294. Also, “Memo” was correlated with “RT,” “PRP,” and “MRP,” respectively, with strengths of 0.146, 0.36, and 0.427. That is, “RT,” “PRP,” and “MRP,” respectively, were correlated with “GPA” with strengths of $(0.294 * 0.146)$, $(0.294 * 0.36)$, and $(0.294 * 0.427)$. It was possible to obtain weights by this calculation method and sum them all. The values of the weights are listed in Table 3.

Table 3

The Values of the Weights

Memo	Marker	RT	PP	PRP	MRP
0.294	0.339	0.601	0.498	0.816	0.984

5. Discussion

To further understand students' possible behavior patterns, cluster analysis was employed. Students were clustered into five groups according to the similarities in their learning behaviors. We used *k*-means clustering to group learners.

Table 4 presents the center, mean values, and standard deviations of each cluster as well as comparisons with a post-hoc test (Scheffe). Clusters 1–5 (C1, C2, C3, C4) included 65, 54, 43, 51, and 22 students, respectively. We then analyzed the learning behavior features in each group.

Table 4

Assigned Weighted k-Means Clustering Results and Analysis

Learning Behavior Clusters							
	Cluster 1 (<i>n</i> = 65) (mean/ <i>SD</i>)	Cluster 2 (<i>n</i> = 54) (mean/ <i>SD</i>)	Cluster 3 (<i>n</i> = 43) (mean/ <i>SD</i>)	Cluster 4 (<i>n</i> = 51) (mean/ <i>SD</i>)	Cluster 5 (<i>n</i> = 22) (mean/ <i>SD</i>)	<i>F</i> -value (ANOVA)	Post-hoc (Scheffe) tests
PRP	0.034/0.034	0.047/0.04	0.155/0.093	0.222/0.126	0.123/0.093	50.776**	4>3.5>1.2
MRP	0.03/0.034	0.186/0.082	0.04/0.047	0.094/0.076	0.443/0.163	140.882**	5>2>4>1,2>3
Memo	0.01/0.024	0.04/0.047	0.104/0.054	0.018/0.034	0.116/0.064	52.14**	3.5>2>1,3.5>4
Marker	0.01/0.018	0.034/0.03	0.09/0.054	0.023/0.033	0.141/0.082	60.241**	5>3>2>1,5.3>4
PP	0.167/0.099	0.189/0.072	0.207/0.074	0.278/0.054	0.251/0.062	17.697**	4.5>1.2,4>3
RT	0.035/0.024	0.12/0.056	0.107/0.047	0.159/0.086	0.253/0.088	65.171**	5>4>2.3>1
GPA	5.357/2.355	6.548/2.364	6.842/1.508	7.125/2.017	7.782/1.036	8.642**	2.3.4.5>1

***p* < 0.001.

PRP: PC Read Pages; MRP: Mobile Read Pages; RT: Reading Time; PP: The rate of returns to previous page

Cluster 3: The reading time (RT) of Cluster 3 is lower than Clusters 4 and 5; however, "Memo" and "Marker" are higher than in Clusters 1, 2 and 4, and "Memo" was almost the same as in Cluster 5, while "MRP" was as low as in Cluster 1, meaning that the students in Cluster 3 were more likely to use a PC device and were actively involved in the learning process, as they actively wrote memos and added markers. They were thus characterized by plus- α learning.

There was no significant difference between Clusters 2, 3, 4, and 5 in "GPA" through a ceiling effect: The upper limit of "GPA" was fixed, so even extensive study could not greatly increase the GPA; any such growth would be small. In other words, for example, the members of Cluster 5 could acquire knowledge with less learning time.

Cluster 3 should be noted. Despite the short learning time, this cluster effectively used the plus- α learning method to good effect. Clusters 2 and 4 would have been able to reduce their learning time if they had effectively handled "Memo" and "Marker." Also, if the students in Cluster 1 also had a bottleneck when learning, it would be best for them to try the plus- α learning method first.

6. Conclusion

In this paper, we have described the need to weight the variables, which was confirmed by the experimental results.

In this paper, we particularly focused on the learning effects of "Memo" and "Marker." The use of these functions led to reduced learning times and made it possible to intensively learn some contents or other subjects. Also, it is good for learners who do not like learning, because it has been verified as reducing the learning time. Thus, students can try this learning method using "Memo" and "Marker." In

addition, if students want to use “Memo” or “Marker,” it is suggested that they use a PC with a large screen, whereas if they want to use this system in their free time or in a conveniently carried form, they can use mobile devices.

Acknowledgements

Part of this research work was supported by the Grants-in-Aid for Scientific Research Nos. 16H03078, 16H06304, and 18H04125 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. We would like to thank all the people who prepared and revised previous versions of this document.

References

- Archer, E., Chetty, Y. B., & Prinsloo, P. (2014). Benchmarking the habits and behaviours of successful students: A case study of academic-business collaboration. *The International Review of Research in Open and Distance Learning*, 15(1), 62-83 .
- Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review*, 42(4), 40-57.
- Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, 52(1), 78-82.
- Macfadyen, L., & Dawson, S. (2012). Numbers are not enough: Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3), 149-163.
- Liu, R., Feng, S., Shi, R., & Guo, W. (2014). Weighted graph clustering for community detection of large social networks. *Procedia Computer Science*, 31, 85-94.
- Mostow, J. (2004). Some useful design tactics for mining its data. *Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes*. In: Lester J.C., Vicari R.M., Paraguaçu F. (eds) Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science, 3220. Springer, Berlin, Heidelberg (pp. 20-28).
- Yamada, M. (2017). The recent trends of learning analytics research and future directions (in Japanese). *Educational Technology Research*, 41(3), 189-197.
- Yin, C., Hirokawa, S., Yau, J., Nakatoh, T., Hashimoto, K., & Tabata, Y. (2013). Analyzing research trends with cross tabulation search engine. *International Journal of Distance Education Technologies*, 11(1), 31-44. doi:10.4018/jdet.2013010103.
- Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning*, 10(4), 455-468.
- Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M...Ogata, H. (2015). Identifying and analyzing the learning behaviors of students using e-books. In: Ogata, H. et al. (Eds.) *Proc. of 23rd International Conference on Computers in Education*. Hangzhou, China: Asia-Pacific Society for Computers in Education (pp. 118-120).
- Yin, C., Okubo, F., Shimada, K., Yamada, M., Fujimura, N., & Ogata, H. (2014). Mobile devices based data collecting system for analyzing learning behaviors. In C.-C. Liu et al. (Eds.), *Proc. International Conference of Computers on Education*, Nara, Japan: Asia-Pacific Society for Computers in Education (pp. 575-577).
- Yin, C., Sung, H. Y., Hwang, G. J., Hirokawa, S., Chu, H. C., Flanagan, B., & Tabata, Y. (2013). Learning by searching: A learning approach that provides searching and analysis facilities to support research trend surveys. *Journal of Educational Technology & Society*, 16(3), 286-300.
- Yin, C., Uosaki, N., Chu, H., Hwang, G., Hwang, J., Hatono, I., Kumamoto, E., Tabata, Y. (2017). Learning behavioral pattern analysis based on students' logs in reading digital books. In : Chen, W. et al. (Eds.) *Proceedings of the 25th International Conference on Computers in Education*, (2017). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education (pp. 549-557).
- Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2018). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human Factor Approach. *International Journal of Human-Computer Interaction*, 35(4-5), 313-322. DOI: 10.1080/10447318.2018.1543077

Research on the Development of a Personalized Learning Assessment Model: Building Connections Between Knowledge Components and Cognitive Levels

Xiaoling Peng^{*}, Bian Wu

Department of Educational Information Technology, East China Normal University, China

**pengxiaoling_7@163.com*

Abstract: Assignment and examination are typical formative and summative assessment strategies in K-12 education. A large number of assessment data generated by learners offers an opportunity for personalized assessment. The research on assessment data has centered on large-scale reporting on aggregate level results, fewer studies have focused on student-level features. In this study, we tried to align Bloom's taxonomy of educational objectives with learning assessment, and construct a personalized assessment model using the assignment and examination data based on the cognitive diagnostic assessment approach. The model includes three assessment dimensions including the achievement of educational objectives, the mastery level of knowledge components and risk detection. The model was validated using 2,600 online learning data from 50 senior high school students. The testing content includes one topic from algebra and another one from trigonometry. The results indicate that the model can help students make timely and targeted remedies of their learning gaps. There is a positive correlation between students' cognitive level and their mastery of knowledge components, and students with the same scores have different cognitive structures and knowledge structures, although they are at the same level in the traditional sense, they can find out the complementary intervals and increase the effective interaction. Assessment data is an explicit form of students' internal cognitive level, compared with a total score, teachers are more concerned about students' cognitive level and their mastery of specific knowledge, especially knowledge components with risks.

Keywords: Personalized assessment, assessment data, taxonomy of educational objectives, knowledge components

1. Introduction

Learning analytics (LA) is the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gasevic, 2012). Presently educational institutions compile and store huge volumes of data, such as student attendance records and behavioral data, as well as their examination results. Mining such data yields better understanding of student performance. However, traditional learning assessment is limited to the statistical analysis of students' scores, class average scores, percentile ranks, etc., and ignore the details of test content and answering data (Liu, You, Wang, Ding, & Chang, 2013; Roberts & Gierl, 2010), resulting in data information cannot be effectively recorded, mined and used. While data-driven learning evaluation gradually becomes intelligent (Dutti, Ismaili, & Herawani, 2017), and the research of assessment model tends to be personalized, comprehensive and self-adaptive. There are still some deficiencies, including lack of personalized assessment from the perspective of knowledge and cognition, and basic education practice is difficult to achieve to promote learning by assessment. In view of the above problems, this study constructed the personalized assessment model from the perspective of evaluation process based on student learning assessment data. This model analyzes the learning assessment data of students from the micro and fundamental

level, and evaluates learners' performance by connecting cognition and knowledge, and detects students' learning gaps and learning risks.

2. Conceptual Framework

2.1 Bloom's Taxonomy of Educational Objectives

The taxonomy of educational objectives is a framework for classifying statements of what we expect or intend students to learn as a result of instruction. In the revision of Bloom's taxonomy, Anderson et al. (2005) used the research results of cognitive psychology for reference, and distinguished the objectives from the two-dimensions of knowledge and cognitive process. The knowledge dimension includes *Factual Knowledge*, *Conceptual Knowledge*, *Procedural Knowledge* and *Metacognitive Knowledge*, mainly to help teachers distinguish what to teach. The cognitive process dimension is divided into six categories: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate* and *Create*, which help teachers to clearly promote the stage process of students' mastering and applying knowledge. Any objective would be presented in two-dimensions table, which termed the *Taxonomy Table*. Using the *Taxonomy Table* to classify objectives, activities, and assessments provides a concise, visual representation. In this study, the guiding significance of Bloom's taxonomy of educational objectives includes item classification, attribute classification and result elaboration.

2.2 Cognitively Diagnostic Theory

The increasing demand of researchers and educational stakeholders for more formative information from educational test has fueled research efforts in CDT (Nichols & Joldersma, 2008). In contrast to classical test theory and item response theory conceptualize learners' competence as a unidimensional latent construct, cognitive diagnosis models (CDMs) assume multiple, discrete skills or attributes, thus allowing CDMs to provide a finer-grained assessment of learners' tests performance. It is designed to measure students' specific knowledge structures and processing skills so as to provide information about their cognitive strengths and weakness (Leighton & Gierl, 2007). This study uses the attribute mastery probability cognitive diagnostic model based on Q matrix (Zhu, Zhang, & Xin, 2009). By specifying a number of skills/attributes required to solve the test items, attribute profiles are reported for any specific response pattern (Rupp, Templin, & Henson, 2010). Q-matrix is a common component of CDMs for specifying the attributes required for each item (Tatsuoka, 1983). This theory, by determining the non-observable cognitive attributes, and transforming them into observable question answering modes, links the non-observable cognitive structures with observable answering responses on items, and provides a basis for understanding students' cognitive structures.

3. Personalized Assessment Modeling Based on Learning Assessment Data

3.1 Classification of Test Questions

The analysis of mainstream online test software, such as *Onion Math*, *Zuoyebang*, *Geek Big data*, shows that software covers multiple types of data. In this study, test questions and test result data are the core data. Test questions data includes test number, content, educational objectives involved, knowledge components covered, etc. Test result data includes right or wrong answers, problem solving process, etc. Bloom's taxonomy of educational objectives is used in this study to support learners' internal cognitive dimension, which provides a basis for the classification of cognitive objective attributes.

The classification of test questions is the basis of composing test papers, analyzing test papers and evaluating students. In this study, the test questions were classified from Bloom's objectives classification and knowledge components classification. The specific contents are as follows: (1) classification of teaching objectives. When using Bloom's taxonomy of educational objectives, we only need to correspond the relationship between nouns and verbs in objectives and each level in two

dimensions. For example, in the objective of "using sine theorem solve an authentic problem", the verb "using" corresponds to the "Apply" in the cognitive process category of the Taxonomy Table, and the noun "sine theorem" corresponds to the "Conceptual Knowledge" in the knowledge category, which belongs to "apply conceptual knowledge". (2) classification of knowledge components. Analyzed the content of the textbook, and then the question data is divided into each interrelated knowledge component.

3.2 Data analysis process based on Q matrix theory

In this study, the attribute mastery probability model based on Q matrix is used to calculate the learning assessment data. The feasibility and effectiveness of the model have been verified to meet the practical needs of teachers in the transformation process from assessment data to effective evaluation. The calculation steps are as follows:

1. Suppose that in a certain test, there are m questions and n students, and the correct answer is marked as 1 and the wrong answer is marked as 0. The project response matrix, i.e., R matrix, for all students to answer right or wrong on all questions can be obtained.

2. Assume that all the test questions only involve l attributes. Through the analysis of the test questions, if the test questions involve this attribute, it will be denoted as 1; if not, it will be denoted as 0. Thus, a Q matrix describing the relationship between test questions and measured attributes is formed:

$$R_{n \times m} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix}, Q_{m \times l} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1l} \\ q_{21} & q_{22} & \cdots & q_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{ml} \end{pmatrix}, r_{ij}, q_{jk} \in \{0,1\}$$

3. According to the Q matrix and R matrix obtained above, using matrix multiplication $N_{n \times l} = R_{n \times m} Q_{m \times l}$, the number of correct responses of each student on each attribute can be obtained by N_{ik} , that is, the number of correct responses of student i to the test questions involving attribute k :

$$N_{n \times l} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1l} \\ q_{21} & q_{22} & \cdots & q_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{ml} \end{pmatrix} = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1l} \\ n_{21} & n_{22} & \cdots & n_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ n_{n1} & n_{n2} & \cdots & n_{nl} \end{pmatrix}$$

4. The probability that student i correctly answered question j is estimated as the product of the correct frequency of all the attributes involved in the question. If question 1 involves attribute 2 and attribute 3, the probability that student 1 correctly answered question 1 is $g_{-11} = f_{-12} * f_{-13}$. Thus, the probability that student i correctly answered question j can be obtained:

$$g_{ij} = \prod_{k=1}^l (f_{ik} \vee (1 - q_{jk})), x \vee y = \max(x, y)$$

5. Finally, the probability of student i 's mastery of attribute k = the sum of the correct answer probabilities of all items involving attribute k and correctly answered by student i / the sum of the correct answer probabilities of all items involving attribute k . At this point, the estimated probability of all students' mastery of all the attributes involved in this exam can be obtained:

$$p_{ik} = \frac{\sum_{j=1}^m \min(r_{ij}, q_{jk}) \times g_{ij}}{\sum_{j=1}^m q_{jk} \times g_{ij}}, x \wedge y = \min(x, y), \text{ if } \sum_{j=1}^m q_{jk} \times g_{ij} = 0, \text{ then } p_{ik} = 0.$$

On the basis of determining data sources and research methods, we have sorted out and divided the dimensions and contents of personalized assessment, and built a personalized assessment model (Figure 1). In the model, we use bloom's two-dimensional educational objectives to evaluate the students' internal cognitive level. It takes the mastery of attributes as the quantitative form, the probability method of attribute mastery as the algorithm support. The left round wheel of the personalized assessment model takes iPadagogy wheel(Cochrane, Narayan, & Oldfield, 2010) as the source of the design idea, so as to show the data basis and the division of the assessment dimension. With the accumulation of assessment data, personalized assessment and the authentic learning situation of individual students tend to be consistent, and gradually play the role of personalized assessment, problem diagnosis, prediction and warning.

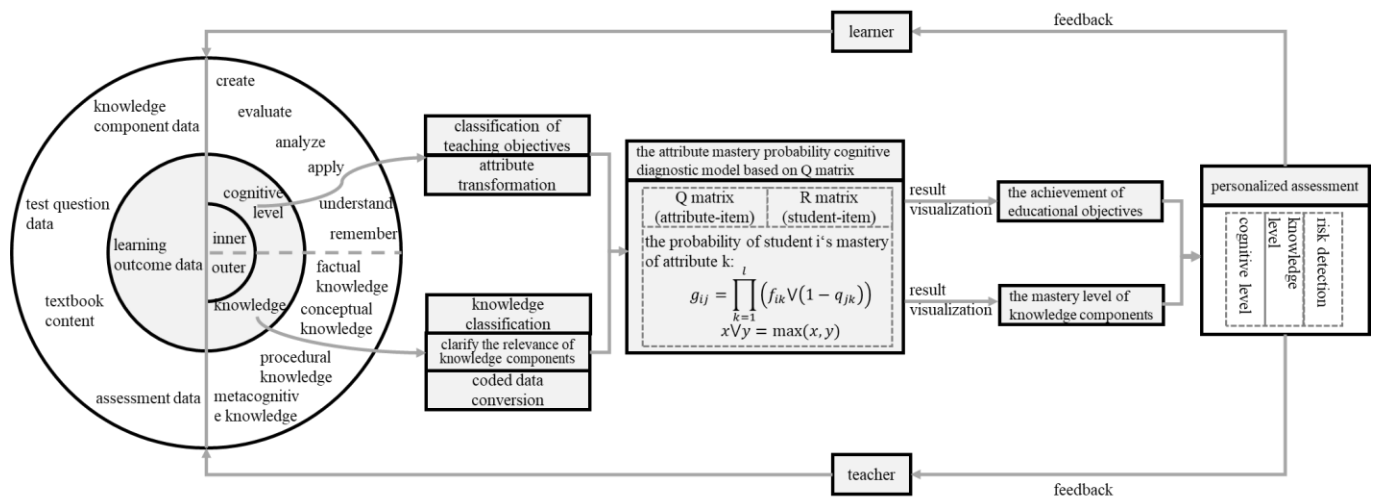


Figure 1. The personalized assessment model based on learning assessment data

4. Data Validation of the Personalized Assessment Model

4.1 Participants and Data Collection

In this study, 50 senior high school students in a class of a high school in Jiangsu province, China were selected as research participants, among which 28 were male students and 22 were female students. Three examination information of the participants on a data analysis platform was taken as data source to test the model. After collaborative screening with the teacher, 21 questions that were not included in the teaching content and incomplete data of students' answers were excluded. A total of 52 test questions were selected as data sources. The types of questions involved include choice question, fill-in-the-blank questions and calculation questions, among which the teaching contents involved in the test questions are the compulsory high school mathematics course 5 chapter 11 "solving triangles" and chapter 12 "sequence of numbers". On this basis, the tested knowledge components can be divided into six knowledge components: the sine theorem, the cosine theorem, the application of the sine theorem and the cosine theorem, the arithmetic sequence, the geometric sequence, and the comprehensive application of the sequence.

4.2 Data Analysis

Firstly, the data are collected and the item response matrix $R_{50 \times 52}$ of students and test questions is obtained. Secondly, by analyzing the test questions, we get the correlation matrix $Q_{52 \times 12}$ of 52 questions and 12 Bloom educational goal attributes involved. Finally, the probability estimates of students' mastery of target attributes are calculated, and the cognitive level is explicit with specific numerical values. Similarly, through the above data analysis method, we calculated the probability estimates of 50 students' mastery of knowledge attributes in the class.

4.3 Results

This paper makes statistical analysis from two aspects: average situation of class and individual situation of students. The results show that the average achievement of Bloom's educational objectives in the whole class decrease as the increase of the objective level. Students generally have the best command of "remember" and "factual knowledge". As the cognitive process from lower order thinking skills to higher order thinking skills and knowledge ranging from concrete to abstract, students' objective achievement degree decreases successively. In terms of the individual situation of the students, student A is similar to the average level of the class in terms of remembrance and understanding, but lower than the average level of the class in terms of application, evaluation and analysis. Therefore, the student should focus on strengthening the training of high-level cognition. In

terms of knowledge dimension, this student has not reached the average level of the class, especially the factual knowledge and procedural knowledge, which are far from the average level of the class (figure 2a). The reason may be that this student is not good at mastering a lot of factual knowledge and thus falls behind other students. To sum up, the student should strengthen the learning of factual knowledge and gradually improve his high-level cognitive level. In addition, by comparing students with the same score in different grades, it is found that students with the same score also have different cognitive distribution, as shown in figure 2 (b). Student B has the same score as student C, and both of them are behind the class average. Student C is superior to student B in low level cognition, factual knowledge and conceptual knowledge. Therefore, it is essential to quantify the cognitive attribute level, find out the complementary interval and increase the effective interaction between students.

The results knowledge components are as follows: (1) The analysis of the average level of knowledge components in the class can help teachers identify weak knowledge components. The results show that students' overall mastery of knowledge components from high to low is Cosine Law, Sine Law, arithmetical progression, geometric progression, application of the sine-cosine law, application of the sequence. (2) In terms of the individual situation of students, we conducted a horizontal comparison of the mastery of each knowledge component of individual students, and found that the knowledge components with learning risks. For example, Student x has a good command of the concepts and principles of the sine theorem and the cosine theorem, but his application ability is lower than the average level of the class (Figure 2c), so as to put forward targeted guidance and suggestions. Secondly, student y and student z with the same score have different knowledge structure (Figure 2d), a longitudinal comparison of students' mastery of various knowledge components is conducted to find complementary components and find suitable learning partners.

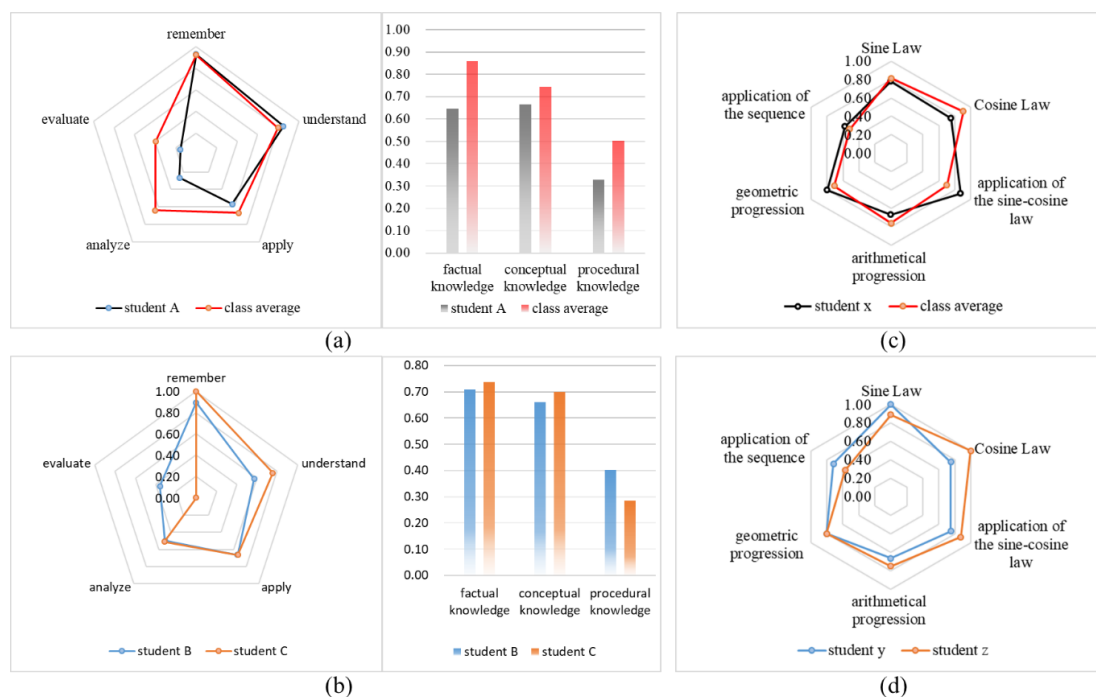


Figure 2. The achievement of educational objectives and the mastery level of knowledge components

5. Discussion and Conclusion

5.1 Discussion

Assessment analysis of cognitive goals links test scores provides an explanatory framework. Based on an examinee's observed response pattern, detailed feedback about an examinee's cognitive strengths and weaknesses can be provided through a score report. This diagnostic information can then be used to inform instruction tailored to the examinee, with the goals of improving or remediating specific cognitive skills. In the calculation results, the knowledge component with poor mastery level, namely

the learning risk problem component, has attracted the common attention of teachers and students. Some teachers in the school believe that with the increase and improvement of assessment data, it is possible to discover learning risk knowledge components, solve learning problems, reduce teaching load and realize individualized teaching. This paper mainly reports the diagnosis results of students' knowledge structure and cognitive structure level, and is limited to the analysis of test data. Another limitation is that this study does not explore how the personalized assessment report of this study is used by teachers, parents and students to help teaching and learning, and does not track the evaluation data to explore the learning trajectory of students, which is also one of the author's future research directions.

5.2 Conclusion

The personalized assessment based on student assessment data provides the possibility for teachers to teach students in accordance with their aptitude, and satisfies the teachers' needs for understanding various mathematical and statistical knowledge in the educational measurement theory. This study analyzes the process of personalized evaluation and calculation, and constructs the personalized assessment model from two dimensions of cognition and knowledge, based on Bloom's taxonomy of educational objectives and cognitively diagnostic theory, and supported by the method of attribute mastery probability. At the same time, the assessment data of 50 senior high school students on a data analysis platform were used to verify the model. The model was iteratively optimized through effective feedback from teachers and students to improve its accuracy, and we investigate teachers' and students' view on the results of data analysis to ensure the feasibility and effectiveness of the model. In the follow-up research, the author hopes to design and develop a personalized assessment tool, which can be integrate into the existing learning platform, so as to conduct personalized process analysis and assessment for learners, help teachers to teach students in accordance with their aptitude, and ultimately help students to improve the learning effects.

References

- Anderson, L. W., & Krathwohl, D. R. (2005). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. *European Legacy*, 114(458), 1013-1014.
- Cochrane, T., Narayan, V., & Oldfield, J. (2010). iPadagogy: Appropriating the ipad within pedagogical contexts. *International Journal of Mobile Learning & Organisation*, 7(1), 146-154.
- Dutti, A., Ismaili, M. A., & Herawani, T. (2017). A Systematic Review on Educational Data Mining. *Ieee Access*, 5, 15991-16005. doi:10.1109/Access.2017.2654247
- Leighton, J. P., & Gierl, M. J. J. L. A. P. (2007). Why cognitive diagnostic assessment? *Lap Lambert Academic Publishing*.
- Liu, H. Y., You, X. F., Wang, W. Y., Ding, S. L., & Chang, H. H. (2013). The Development of Computerized Adaptive Testing with Cognitive Diagnosis for an English Achievement Test in China. *Journal of Classification*, 30(2), 152-172. doi:10.1007/s00357-013-9128-5
- Nichols, P. D., & Joldersma, K. (2008). Cognitive Diagnostic Assessment for Education: Theory and Applications. *Journal of Educational Measurement*, 45(4).
- Roberts, M. R., & Gierl, M. J. (2010). Developing Score Reports for Cognitive Diagnostic Assessments. *Educational Measurement: Issues and Practice*, 29(3).
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement: Theory, methods, and applications.
- Siemens, G., & Gasevic, D. (2012). Guest Editorial - Learning and Knowledge Analytics. *Educational Technology & Society*, 15(3), 1-2.
- Tatsuoka, K. K. J. J. o. E. M. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, 20(4), 345-354.
- Zhu, J., Zhang, M., & Xin, T. (2009). A Model of Cognitively Diagnostic base on Q Matrix: Classifying Model of the Probability of Attributes' Mastery. 45(2), 117-122.

Modeling Self-Planning and Promoting Planning Skills in a Data-Rich Context

Huiyong LI^{*}, Yuanyuan YANG^a, Rwitajit MAJUMDAR^b, Gökhan AKÇAPINAR^{b,c},
Brendan FLANAGAN^b & Hiroaki OGATA^b

^aGraduate School of Informatics, Kyoto University, Japan

^bAcademic Center for Computing and Media Studies, Kyoto University, Japan

^cDepartment of Computer Education and Instructional Technology, Hacettepe University, Turkey

*lihuiyong123@gmail.com

Abstract: Students' learning behaviors in an online learning environment can be automatically recorded by learning systems. Such learning records provide new opportunities to model students' learning process. On the other hand, it has become more common to see students having wearable devices that assist in tracking their personal physical activities. These activity tracking can be integrated into a data-rich context for training students for developing their data-informed self-direction skills. We are building the GOAL (Goal Oriented Active Learner) system to support the development of self-direction skills using learning and health activity data. A key phase in any self-directed activity is goal setting and planning. This paper will introduce how to build a new model for self-planning and support the acquisition of planning skills in the GOAL system. We combine learners' data from the self-directed activity and their interaction trace to build the model in the GOAL system. The modeling involves computing of trend value and degree of plan difficulty, then diagnosis of planning skills using a 5-point scoring criteria. An adaptive support is selected based on the computed score. The contribution of this work is modeling planning and promoting planning skills in a data-driven manner. Our approach grounds the theory of self-direction skills and enables learners to develop the skills in everyday life.

Keywords: Planning skills, self-direction, learning analytics, activity tracking, ubiquitous and mobile computing

1. Introduction

Self-direction Skills (SDS) are acquired through experience, training, and effort. The benefits of experience and training will depend on the degree to which people engage through volitionally initiated thought processes. Since it is a cognitively and behaviorally complex task to execute SDS, diagnosing learners in underdeveloped skills and subsequently designing support for promoting the skills are essential.

Although there are multiple approaches to capture data on learner's self-direction or self-regulation, self-report measures have still stayed dominant so far. The recent availability of large and fine-grained datasets has led to investigations of self-regulation by applying learning analytics (LA). The assessment of frequencies and sequences of regulatory activities in learning environments provides a novel perspective on self-regulation that complements and potentially supersedes traditional self-report measures (Bannert & Sonnenberg, 2014; Li et al., 2018). On the other hand, the increased availability of the activity tracking data gives individuals more opportunities for establishing benchmarks in objective metrics and improving achievements through the experience of reality (Swan, 2013). The research and design of data quantification have grown as an interest area in information and learning sciences (Lee, 2019).

This leads us to build the GOAL (Goal Oriented Active Learner) system to support the development of SDS with integrating the learning records and self-tracked data (Majumdar et al., 2018). The GOAL system not only leverages the rapidly increasing activity data but also creates data-driven feedback loops for the acquisition of SDS. In this paper, we will introduce how to build a new model for self-planning and support the acquisition of planning skills in that data rich context.

2. Related Work

2.1 Self-Direction Skills

According to P21 (Partnership for 21st Century Skills, 2016) framework, *Initiative and Self-Direction* requires monitoring one's understanding and learning needs, demonstrating initiative to advance professional skill levels, defining, prioritizing and completing tasks without direct oversight and demonstrating commitment to lifelong learning. It requires learners to handle multiple environments, goals, and tasks while understanding and adhering to organizational or technological constraints of time, resources, and systems. The framework gives a general criterion for a self-directed learner.

Self-directed learning (SDL) and self-regulated learning (SRL) are two most frequently used of the terms in today's educational discourse on learning process (Brockett & Hiemstra, 2018; Candy, 1991; Winne et al., 2006; Zimmerman, 2008). Literature highlights their commonality and differences (Saks & Leijen, 2014). Both SDL and SRL have 4 key phases: Task definition – Setting goals and Planning – Enacting strategies – Monitoring and Reflecting. SDL due to its adult education roots is mostly used for describing the learning activities outside traditional school environment. SRL, on the other hand, is mostly studied in the school environment.

Technological innovation in the field of data logging and rapidly increasing digital world have expanded the intersection of SDL and SRL. The processes of executing and developing SDL and SRL can be captured. For our work, we proposed a five-phase process model, DAPER which synthesizes the SDL and SRL models to conceptualize data-driven self-direction skill execution and acquisition.

2.2 Planning Skills in Self-Direction

A key phase in self-direction and self-regulation is goal setting and planning (Brockett & Hiemstra, 2018; Winne et al., 2006). Previous studies of self-direction and self-regulation has highlighted learner agency regarding how they learn and the superiority of autonomous motivation for learning (Stockdale & Brockett, 2011). Here we follow that paradigm and let students choose their own goal and direct their own plan.

Self-directed learners are expected to actively and autonomously engage in goal setting and planning (Wang, Shannon & Ross, 2013). They demand a high level of goal setting and planning skills during the self-direction process. Compared with assigned goals and plans, personal goals and plans produce higher goal commitment since the learners who are aware of their goals have high learning and achievement expectations. Therefore, self-planning skills required taking full responsibility for personal plans. They should check current status of any activity before planning and then create plans to change that status to a more desirable one when they are ready.

2.3 Support Planning in Self-Direction

In comparison to executing assigned plans, setting and striving personal plans provide individuals with less support on how to define personal plans and continuously improve them. Learners may lose their directions if without reliable, revealing and relevant data that support decision-making for planning. Following the learning analytics process model learners need to translate awareness into action (Bodily et al., 2018). They need a 'representative reference frame' to interpret the data (Wise, 2014). Both the context data and process trace data can be valuable ways to create such a reference frame. We have proposed a measurement rubric as a basis of adaptive support (Majumdar et al., 2019).

Therefore, this paper explores how to design a data rich context for self-planning, how to model self-planning, and how to provide support for promoting planning skills.

3. Our Context: Activity Data, DAPER Model, GOAL System

3.1 Activity Data in the Context of Learning and Physical Activities

Activity tracking involves some process or system to collect data generated by an individual during their execution. A variety of everyday life activities can be tracked by the currently available behavior

sensors. We synthesize activity data in learning and physical activity context to involve individuals in self-directed practices. For learning activities, data is from learning logs in our learning platform which include digitized reading logs, status of course assignments, and answers of quizzes (Flanagan & Ogata, 2017). For physical activities, records are collected through native mobile health applications like Apple Health or Google Fit containing data regarding runs, workouts, sleep, steps taken, weight, heart rate, and calories burned.

3.2 DAPER Model

The DAPER (data collection-analyze-plan-execution monitoring-reflect) model conceptualizes the process of data informed SDS execution and acquisition (see Fig. 1). It has five phases, the initial phase of data collection which gives learners the initiative, followed by other four phases: data analysis, planning, execution monitoring, and reflection.

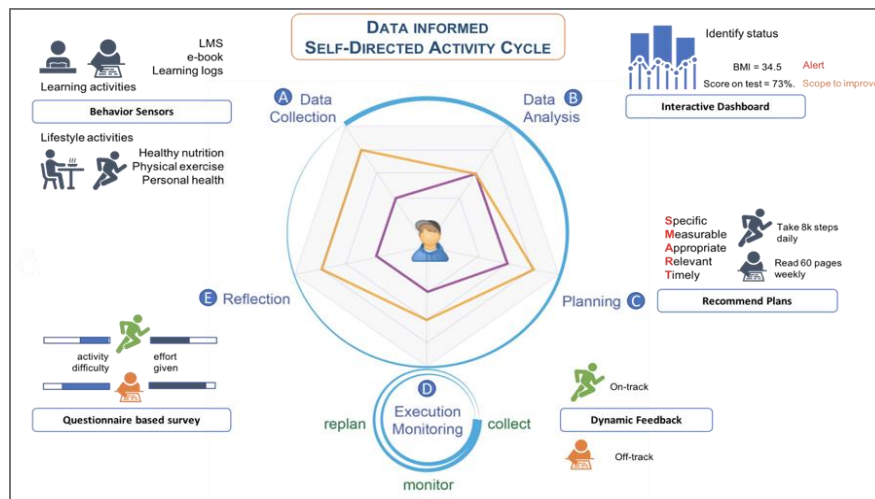


Figure 1. DAPER Model of Self-Direction Skills Execution (Majumdar et al., 2018)

3.3 GOAL System

The Goal Oriented Active Learner (GOAL) system integrates data during learners' learning and physical activities, and implements DAPER model with the functionalities required in each phase. The system further records individual's interaction log data. The GOAL architecture is given in Figure 2.

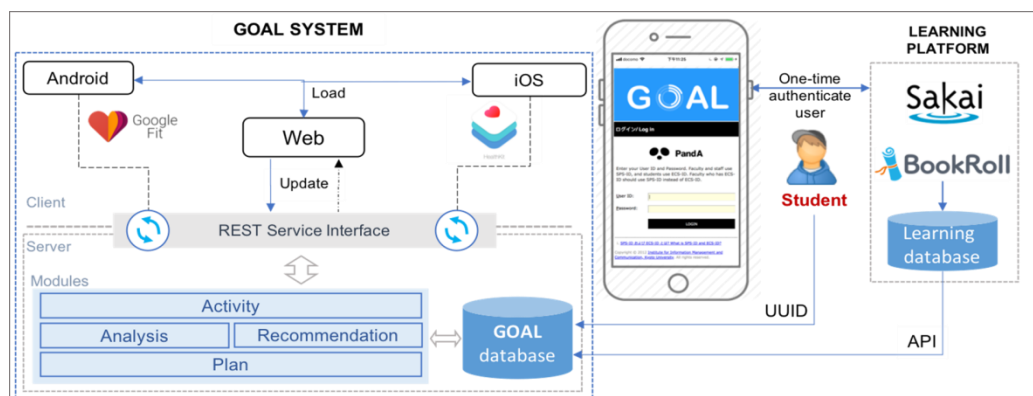


Figure 2. Architecture of the GOAL System (Majumdar et al., 2018)

The GOAL system includes cross platform applications and an analysis server. Learners can link automatically their learning activity data from the LMS and other linked e-learning tools. For physical activity data, students authenticate to synchronize that data directly from native mobile health apps like Apple Health or Google Fit. Furthermore, we log interactions between learners and the GOAL system as eXperience API (xAPI) statements in the analysis server of GOAL system (Li et al., 2019).

4. Modeling Self-Planning

Our proposed self-planning model is shown in Figure 3. First, the trend value of activity is computed from the activity data using autoregressive (AR) model. Second, the degree of plan difficulty is calculated from the trend value and planned value. Third, the planning skills are diagnosed using a 5-point scoring criteria. Finally, an adaptive support is generated based on the score of planning skills.

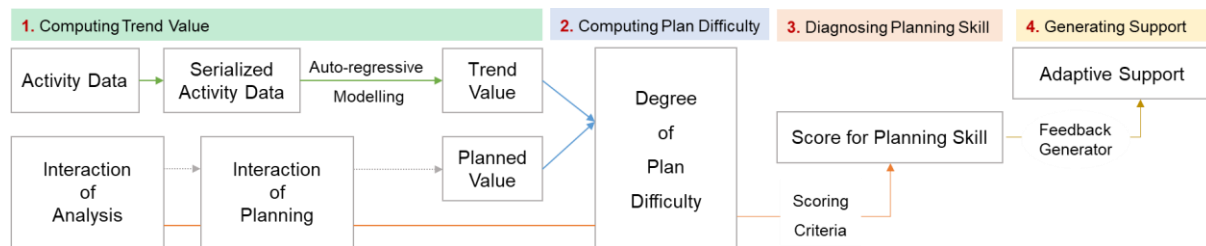


Figure 3. Modeling Self-Planning

The action type, variable, definition, and example for modeling are shown in Table 1. Activity data is data logged during an activity session, which synchronized from learning platforms or health apps. Serialized activity data is a time series data, which accumulated from activity data on the same unit over a fixed period (hour, day, week, etc.). It's used for further computing the trend value as input data. GOAL interactions mean interactions between learners and the GOAL system. They are tracked by the eXperience API, including actor, verb, object, and timestamp.

Table 1: Action Type, Variable, Definition, and Example

Type	Variable	Definition	Example
Activity	Activity data	Data logged during an activity session	"Reading 10 pages in e-book from 2:00 pm to 2:25 pm"
	Serialized activity data	Activity data accumulated over periods in a time series	"0, 10, 0, 0, 20, 0, 0 (pages read/day)"
GOAL interactions	Interaction of planning	Interaction of creating a plan in the system	"John created a plan that taking 8,000 steps per day for the next week at 8:00 am"
	Interaction of analysis	Interaction of executing preparatory analysis in the system	"John checked his activity data for last 7 days at 7:00 am"
	Planned value	Value set in the interaction of planning	"8,000 (steps)"

4.1 Computing trend value

The auto-regressive (AR) model is adopted to predict the trend value of activity from previous serialized activity data at the last p time value:

$$A_t = \beta_0 + \beta_1 A_{t-1} + \beta_2 A_{t-2} + \dots + \beta_p A_{t-p} + u_t$$

Where A_t is the value of A in period t , data set of $A_{t-1}, A_{t-2}, \dots, A_{t-p}$ are time series value in periods $t-1, t-2, \dots, t-p$. $\beta_0, \beta_1, \dots, \beta_p$ are coefficients found by optimizing the model. For example, the trend value of steps taken for the next day could be predicted given the previous serialized activity data in daily scale for 7 days. For each activity, the trend value is computed with daily, weekly, monthly scales. The trend value in scale is a baseline value for self-planning in daily, weekly, or monthly activities.

4.2 Computing Degree of Plan Difficulty

The degree of plan difficulty (DPD) is calculated by comparing the planned value to the trend value. If the planned value is less than the trend value, DPD is zero. If the planned value is more than the trend value, DPD is the relative increase with respect to the trend value. For instance, if the trend value is 1,000 steps per day and the planned value is set as 1,500 steps per day, then the DPD is set to 0.5.

4.3 Diagnosis of Planning Skills using 5-point Scoring Criteria

Three parameters are considered for the diagnosis of planning skills: interaction of planning (IP), interaction of analysis (IA), and DPD. The planning skills are measured as a 5-point scoring. Table 2 indicates the scoring criteria for planning skills with action description and scoring parameters.

Table 2: Scoring Criteria for Planning Skills

Score	Action Description	Parameters
4	Set appropriately challenging plan after analysis	IP=1, IA=1, $0 < \text{DPD} \leq \mu$
3	Set too difficult plan after analysis	IP=1, IA=1, $\text{DPD} > \mu$
2	Set too easy plan after analysis	IP=1, IA=1, $\text{DPD} \leq 0$
1	Set plan without analysis	IP=1, IA=0
0	No plan is set	IP=0

IP: Interaction of planning, IA: Interaction of analysis, DPD: Degree of plan difficulty, μ : Threshold $\mu = 0.2$ in the initial plan, which will be adjusted by individual's achievement

4.4 Adaptive Support based on Skill Scores

Learners are classified into 5 groups based on the scoring criteria for planning skills. They are given adaptive support using a feedback generator (see Table 3). The feedback contains a suggested action and a brief description of current skill level for learners.

Table 3: Adaptive Support for Learners based on Planning Skills

Stage	Planning Skill	Feedback
S4	Set challenging plan after analysis	You have set appropriately challenging plan You already master the skill of planning
S3	Set too difficult plan after analysis	Please decrease the difficulty level of the plan to achieve timely You have partly mastered the skill of planning
S2	Set too easy plan after analysis	Please increase the difficulty level of the plan to challenge yourself You have partly mastered the skill of planning
S1	Set plan without analysis	Please analysis activity data before plan for it You have initiated to acquire the skill of planning
S0	No plan is set	Please try to create a plan You have not shown the skill of planning yet

5. Conclusion and Discussion

In this paper we proposed a novel model for self-planning and support the acquisition of planning skills in a data rich context. The diagnosis of planning skills helps learners to further understand how they engage in planning, it's important to better understand which feedback could be provided. Benefits from using the adaptive support are to facilitate the transfer of control between system and learners. The learner who is in underdeveloped skills is partially guided by system and then exerts more control over the direction as a fully self-directed learner.

The contribution of this paper is modeling planning and promoting planning skills in a data-driven manner. 1) *The rationality of modeling*. The data for modeling is the activity tracking data in learning and health, which gives learners the initiative to advance themselves. We confirm the fundamental assumption of self-direction that learner's agency is central in autonomous learning (Stockdale & Brockett, 2011). Our model could be applied in everyday activities across the contexts. For instance, potential activities could be vocabulary remembering in language learning, gym training in health, piano practice in music learning. 2) *The flexibility of adaptive support*. Because learners are divided according to the relative value of activity data and interaction trace data, the adaptive support is

provided based on dynamic groups. The model could help learners who are ready to develop planning skills, and also help learners who just want to identify current status.

For future work, we will conduct the study in K-12 and higher education settings to evaluate the effects of the skill diagnosis and adaptive support on the improvement of engagement and planning skills. The differences of effects across different settings and activities will also be examined. The research project aims to explore a data-driven paradigm to develop SDS, and support learners while they become more autonomous in learning and in life.

Acknowledgements

This work was partly supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304, JSPS KAKENHI Research Activity Start-up Grant Number 18H05746 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning*, 9(2), 161-185.
- Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: a systematic review. *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41-50). ACM.
- Brockett, R. G., & Hiemstra, R. (2018). *Self-direction in adult learning: Perspectives on theory, research and practice*. New York: Routledge.
- Candy, P. C. (1991). *Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice*. Jossey-Bass, 350 Sansome Street, San Francisco, CA 94104-1310.
- Flanagan, B., & Ogata, H. (2017). Integration of learning analytics research and production systems while protecting privacy. *Proceedings of the 25th International Conference on Computers in Education (ICCE2017)*, pp. 333-338.
- Lee, V. R. (2019). On researching activity tracking to support learning: a retrospective. *Information and Learning Sciences*, 120(1/2), 133-154.
- Li, H., Flanagan, B., Konomi, S. I., & Ogata, H. (2018). Measuring Behaviors and Identifying Indicators of Self-Regulation in Computer-Assisted Language Learning Courses. *Research and Practice in Technology Enhanced Learning*, 13(1), 19.
- Li, H., Majumdar, R., Yang, Y. Y., Flanagan, B., & Ogata, H. (2019). Extracting Self-Direction Strategies and Representing Practices in GOAL System. *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, Tempe, USA, 2019.
- Majumdar R., Yang Y.Y., Li H., Akçapınar G., Flanagan B., & Ogata H. (2018). GOAL: Supporting Learner's Development of Self-Direction Skills using Health and Learning Data. *Proceedings of the 26th International Conference on Computers in Education (ICCE2018)*, pp. 406-415.
- Majumdar, R., Yang, Y. Y., Li, H., Akçapınar, G., Flanagan, B., & Ogata, H. (2019). Adaptive Support for Acquisition of Self-direction Skills using Learning and Health Data. *Proceedings of the 19th IEEE International Conference on Advanced Learning Technologies (ICALT2019)*, pp. 54-56.
- Partnership for 21st Century Skills. (2016). *A state leader's action guide to 21st century skills: A new vision for education*. Tucson, AZ: Partnership for 21st Century Skills.
- Saks, K., & Leijen, Å. (2014). Distinguishing self-directed and self-regulated learning and measuring them in the e-learning context. *Procedia-Social and Behavioral Sciences*, 112, 190-198.
- Stockdale, S. L., & Brockett, R. G. (2011). Development of the PRO-SDLS: A measure of self-direction in learning based on the personal responsibility orientation model. *Adult Education Quarterly*, 61(2), 161-180.
- Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big data*, 1(2), 85-99.
- Wang, C. H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, 34(3), 302-323.
- Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R., & Perry, N. E. (2006). Supporting self-regulated learning with gstudy software: The learning kit project. *Technology Instruction Cognition and Learning*, 3(1/2), 105.
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 203-211). ACM.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1), 166-183.

Learning Support System for Software Component Design based on Testability

Yasuhiro NOGUCHI^{a*}, Daiki IHARA^b, Satoru KOGURE^a, Koichi YAMASHITA^b,
Tatsuhiko KONISHI^a & Yukihiro ITOH^c

^a*Faculty of Informatics, Shizuoka University, Japan*

^a*Graduate School of Integrated Science and Technology, Shizuoka University, Japan*

^b*Faculty of Business Administration, Tokoha University, Japan*

^c*Shizuoka University, Japan*

*noguchi@inf.shizuoka.ac.jp

Abstract: In this paper, we describe a learning support system for software component design for learners at the early programming learning stage. In general, the source code used in instructional programming exercises does not have a testing code and is not designed to rapidly verify behaviors. We think one of the difficulties for students in designing their source code to be testable is that most cannot analyze their component design themselves. Therefore, we summarized the testability features applicable to such learners and designed our learning support system to improve instructional exercises in analyzing source codes and creating improved component design through feedback. The learning outcomes using the system were evaluated based on the subjects' improved scores between pre- and post-tests on analyzing source codes and creating improved component design. The evaluation results indicated the effectiveness of the program with some testability features. We implemented the system in three different programming classes to evaluate the applicability of the system.

Keywords: Programming education, learning support environment, software component design, unit test, testability

1. Introduction

In recent years, in addition to increasing the size and complexity of developed software, the importance of software testing has accordingly increased. Feathers (2004) indicated that a source code without a testing code is a legacy code. However, few studies on software testing and software component design for testing at the early programming learning stage have been conducted. In reality, source codes developed by learners in their exercises do not typically include testing codes for their components. However, there are two problems with learners not writing testing codes and/or not paying attention to the component design for testing. First, because they cannot edit the source code with rapid verification, they tend to take more time for debugging in their exercises. Second, while they continue their programming exercises without testing codes and/or testing designs, their programming style (debug later programming) will be fixed and their programming experience will not be sufficiently varied.

One of the software development processes that enforces programmers to write testing code is test-driven development (TDD). The process makes programmers follow the test-first approach, the programmer cannot write any component before preparing their test codes. Riley and Goucher (2009) reported that TDD can improve the testability of the software components that are being developed. However, the process requires the programmers to have enough knowledge and skills to write test codes with adequate component design for the testing. It is difficult for learners in the early programming learning stage to focus on learning software testing and component design for the testing in the process.

Funabiki et al. (2013) suggested a learning assistant system based on the TDD method. The learning assistant system lets learners follow the TDD process by supporting their testing code creation using a testing framework. Matsuoka (2013) suggested a support system to visualize the learners' testing progress and their testing coverage. However, these studies were focused on the support the learners need to write their testing code. They did not assume that the learners need to improve the design for the tested components before they write their testing codes.

Ohashi (2015) proposed a support system for automated testing by suggesting its weakness based on the following aspects related with tested components:

- **Reproducibility:** the component always returns to the same value for the same parameters.
- **Simplicity:** one method in a component should be simplified, as only one test can verify it.

However, the system focused on improving the tested components for stably maintaining their testing codes; the tested components were prepared their testing code templates. As for learners in the early programming learning stage, we think that the system does not sufficiently improve the learners' component design whether the component can be tested or not.

In this paper, we focus on a learning component design in which the learner can write testing code based on testability. For learners at the early programming learning stage, we discussed effective testability features for software component design and related learning strategies. We implemented a web-based learning support system based on the discussion. Our evaluation supported some learning effects of the proposed system for some testability features.

2. Learning Design for Testability-Based Software Component Design

2.1 Testability Related to Early Programming Learning Stage

Boehm (1976) introduced the concept of testability. Treon et al. (1999) suggested self-testable components. Bach (1999) categorized intrinsic testability related to the software component itself according the following features: observability, controllability, algorithmic simplicity, unbugginess, smallness, decomposability, and similarity. Binder (1999) claimed that controllability and observability are important features according to the object-oriented design perspective. We considered that decomposability is also an important feature for learners to understand at the early programming learning stage. A component whose decomposability is high indicates that the component encloses multiple functions. Thus, before testing individual function, the learners should divorce these individual functions from the component. Based on the above discussion, our learning support system focused on observability, controllability, and decomposability, as shown in Figure 24.

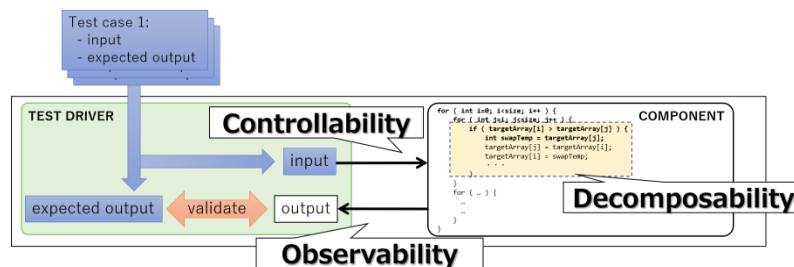


Figure 24. Controllability, observability, and decomposability in the testing model.

2.2 Learning Process for the Testability-Based Software Component Design

We assumed that the learners had already learned basic test driver implementation using a testing framework. We organized the learning items for component design into the following two points.

(I) *Understanding the Concept of Testability*

(II) *Analysis and Improvement for Concrete Code Based on Testability*

As for (I), it is necessary for learners to understand the concept of testability. Especially for controllability, observability, and decomposability, the learners should learn what kind of component design reduces testability. We expected that learners would acquire their knowledge from the lecture.

As for (II), the learners should be able to analyze the testability of concrete components based on the concept of the testability. In looking for the reason why we can write test code for the components, they should acquire the skills to analyze the components and identify their lack of testability. Furthermore, to improve the testable components, they should acquire the skills to create their component design by fixing the lack of testability and modifying the code based on the plan. We think

that exercises where learners analyze/modify various patterns of concrete components are required for learners to develop the needed skills. In such exercises, we know that many learners cannot evaluate their analysis and/or their component design themselves, and they do not have confidence in their answers even when they can complete the modification of the components following their design. Thus, to support their analysis, we should provide the learners with feedback on whether their analysis and component design are adequate based on the testability features.

Based on the above, we developed a learning process for the testability-based software component design, which is displayed in Figure 25. Our proposed system supports steps (2) – (5).

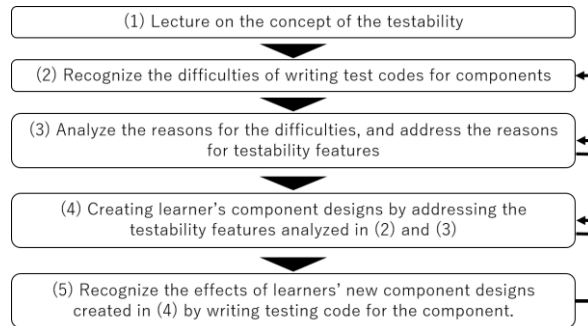


Figure 25. Learning process for the testability-based software component design.

3. Development of the Learning Support System

We implemented our learning support system as a web application. In this section, we explain the learners' usage of the system and the system's feedback (A) – (D) for their inputs in each step.

3.1 System Usage in steps (2) and (3)

Figure 26 shows the interface of the system in step (2). The source code area shows sample codes of the exercise chosen by the learner. The testing code area shows the testing code template for the function the learner should verify. The instruction area shows the system's message to inform the learner how to proceed in the next step. The bottom of the instruction area shows the learner's input field.

First, the exercise explains the learner about the function they should verify via the instruction area. The learner should choose the adequate code area for the function by selecting a rectangle. The system provides feedback (A) on whether the code area chosen by the learner is adequate or inadequate for the function set by the exercise.

In the testing code area, the learner should try to write their testing code for their chosen codes. They should also recognize the difficulties of writing their testing code in the current component design. Finally, they should externalize the reason why the component is difficult to test into the text input field on the bottom of the instruction area.

After submitting, the system gives feedback (B) that highlights words related to the testability features in the explanation. These words the system should highlight are registered by lack testability features of each prepared problem. Then, the system asks the learner to choose an appropriate reason by comparing their own externalized reason. The system suggests that the learner selects the reason related to each testability feature as follows. When the learners choose an answer, the system gives feedback (C) on whether the answer is correct or incorrect in the sample code with the explanation.

- In the testing code, I cannot change the input for testing.
- There is a useless input parameter in the component.
- The component is too restricted to be called from the testing code.
- In the testing code, I cannot observe the output from the component.
- Any components that are mismatched to the code area I should write testing code for.

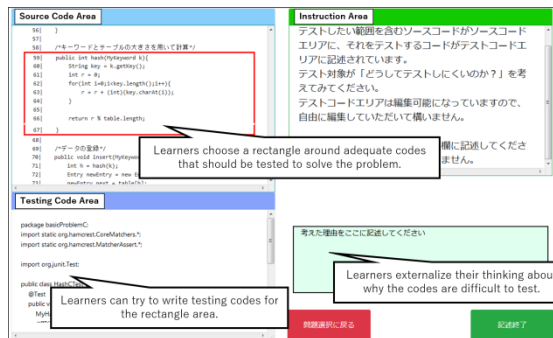


Figure 26. Externalizing their analysis in step (2).

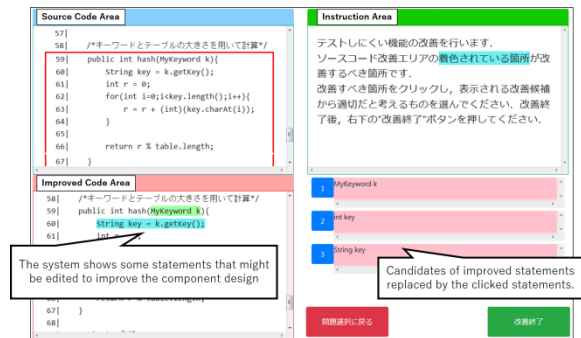


Figure 27. Creating their design in step (4).

3.2 System Usage in step (4)

Figure 27 shows the interface of the proposed system in step (4). The source code area shows the original sample codes. The improved code area shows the sample codes with some highlighted statements. When the learner clicks the highlighted statement, the system suggests some candidates at the bottom of the instruction area. When the learner selects a candidate, the testing code area is updated by the learner's selection. By changing the combinations of statements, the learner changes the sample code into a new component design the learner can write testing code for. When the learner submits the improved code, the system provides feedback (D) based on the learner's improved code. If the learner's improved code has syntax problems, the system sends the learner syntax error messages generated by an appropriate compiler of the programming language. If there are no syntax problems in the learner's improved code, the system decides whether the improved code enables the learner to write their testing code for the function indicated by the exercise. The system provides feedback messages on why their improved code has not fixed the problem the learner analyzed in steps (2) and (3). We defined these messages for every possible combination of candidates when we created the exercise.

3.3 System Usage in step (5)

The system provides the learners with their improved source codes as a project archive for Eclipse IDE. The learners can evaluate their new component design by executing their testing codes in the project.

4. Evaluation

4.1 Hypotheses

We compiled the following three hypotheses on learning the testability-based software module design with the proposed system.

- Hypothesis 1. Learners who completed the exercises with the system can better analyze the lack of testability in the source code than those with only IDE support.
- Hypothesis 2. Learners who completed exercises with the system can create their component design to improve testability better than those with only IDE support.
- Hypothesis 3. The learners who finished exercises with the system will not rely on the system when they apply their acquired skills to other problems.

4.2 Evaluation Process

Figure 28 shows the evaluation process. 15 subjects were college students with basic syntax knowledge in the Java programming language. The pre-test confirmed their basic knowledge of algorithms, basic syntax of the Java programming language, and that they did not have skills for analyzing/improving

codes that lack testability. We divided the subjects into the experimental group with 8 subjects and the control group with 7 subjects to equalize the results of the pre-test.

As for the exercises, we gave both groups same four problems analyzing/improving codes that lack testability. In the post-test, there were three problems where the subjects analyzed the sample codes and improved them. The subjects were asked about (a) the area of source code supporting the function to be tested, (b) why the function is difficult to test, (c) where the area of code to be edited is, and (d) how to improve the code to solve the problem. As for question (c), we evaluated the answer using a point-deduction scoring system.

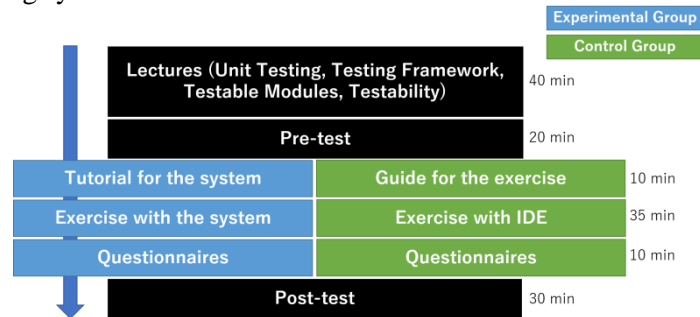


Figure 28. Evaluation process.

4.3 Results

Table 9 shows the average score for answers (a) and (b) in the post-test. The difference was calculated by subtracting the score of the control group from the score of the experimental group. The t-value was calculated using Welch’s one-sided t-test. Table 10 shows the average scores for answers (c) and (d) in the post-test.

Table 9

Average Scores for Analyzing Lack of Testability in the Post-Test

Lack of Testability	Controllability		Observability		Decomposability	
	(a)	(b)	(a)	(b)	(a)	(b)
Avg. of Experimental Group	0.750	0.625	1.000	0.813	0.250	0.250
Avg. of Control Group	0.571	0.429	1.000	0.071	0.357	0.286
Difference	0.179	0.196	0.000	0.741	-0.107	-0.036
T-value	0.252	0.242	N/A	0.000	0.320	0.443

Table 10

Average Scores for Creating Improved Component Design in the Post-Test

Lack of Testability	Controllability		Observability		Decomposability	
	(c)	(d)	(c)	(d)	(c)	(d)
Avg. of Experimental Group	-0.167	0.833	-0.125	0.375	-2.667	1.333
Avg. of Control Group	-1.000	0.500	-2.000	0.000	-3.000	0.667
Difference	0.833	0.333	1.875	0.375	0.333	0.667
T-value	0.024	0.107	0.000	0.010	0.607	0.464

As for Hypothesis 1, Table 9 shows the experimental group’s scores of controllability and observability are the same or better than the control group’s scores. In Figure 29, the subjects’ feelings of understanding in the exercise of the experimental group are better than those of the control group. Although the t-value only supports the score of question (b) for observability, the system does not seem to have bad effects based on the subjects’ exercise and questionnaire results. Regarding decomposability, the average scores of the control group were better than those of the experimental group. There were many subjects in the experimental group who felt less understanding compared with the feelings of controllability and observability. In both groups, the subjects who could answer correctly were divided. In the observations of the subjects’ post-test results, we found that some subjects in both groups were unable to answer the question. The system may have some educational effect for learners

to analyze the lack of controllability and observability in source code. However, the system is not well designed for decomposability.

As for Hypothesis 2, Table 10 shows that all scores of the experimental group were better than those of the control group. The t-value supports the scores in question (c) on controllability and observability, and the score of question (d) on observability. We think that the system has some educational effect for learners' component design against the lack of controllability or observability.

As for Hypothesis 3, even if the subjects of the experimental group answered the post-test without the system's support, more subjects answered correctly compared with the control group. Therefore, we think that the learning effect with the system is negligible.

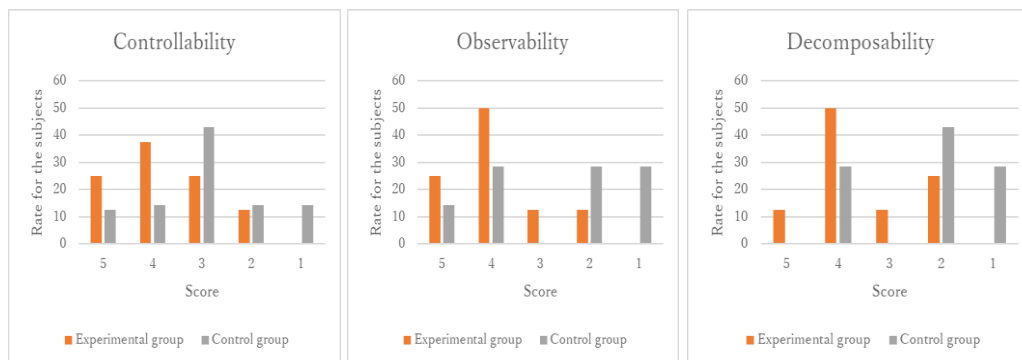


Figure 29. Subjects' feelings of understanding in their exercises.

5. Conclusion

In this study, we built a learning support system for software component design based on testability. The system can support learners' analysis of the lack of testability in the source codes and create their improvement plan for the component design against the lack of testability. The system covered three types of testability: controllability, observability, and decomposability. Our evaluation indicated some learning effects of the proposed system in relation to controllability and observability. We succeeded in introducing the system to three programming classes to evaluate the applicability of the system. After the exercises with the system, the participants tried to analyze larger sample source codes without the system, most participants analyzed the sample codes based on the testability features and suggested their refactoring plans, although all prepared functions with the lack of testability were not analyzed due to the time limitations of the classes.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP18K11566.

References

- Bach, J. (1999). Heuristics of software testability. Retrieved from <http://www.satisfice.com/tools/testable.pdf>
- Binder, R. V. (1994). Design for testability with object-oriented systems. *Communications ACM*, 37, 87–101.
- Boehm, B. W. (1976). Quantitative evaluation of software quality. *Proceedings of the 2nd of ICSE* (pp. 592–605).
- Feathers, M. C. (2004). *Working effectively with legacy code*. Upper Saddle River, NJ: Prentice Hall.
- Funabiki, N., Matsushima, Y., Nakanishi, K., Amano, N. (2013). A Java programming learning assistant system using test-driven development method. *IAENG International Journal of Computer Science*, 40(1).
- Le Treon, Y., Deveaux, D., & Jézéquel, J. M. (1999). Self-testable components: From pragmatic tests to design-for-testability methodology. In editor name(s) (Ed(s).), *Proceedings of Technology of Object-Oriented Languages and Systems* (pp. 96–107). Nancy, France.
- Matsuoka, S., & Katayama, T. (2013). Development of unit testing visualization tool to implement real-time visualization of testing progress. *IEICE Technical Report*, 112(373), 37–42. (in Japanese)
- Ohashi, A., Kaminaga, H., Nakamura, S., Morimoto, Y., & Miyadera, Y. (2015). A support system of automated tests in software development exercises. *IEICE Technical Report*, 114(513), 13–18. (in Japanese)
- Riley, T., & Goucher, A. (2009). *Beautiful testing: Leading professionals reveal how they improve software*. Sebastopol, CA: O'Reilly Media.

Identifying Reading Styles from E-book Log Data

Ivica BOTICKI^{a*}, Hiroaki OGATA^b, Karla TOMIEK^a, Gokhan AKCAPINAR^b, Brendan FLANAGAN^b, Rwitajit MAJUMDAR^b & Nehal HASNINE^c

^a*Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*

^b*Kyoto University, Japan*

^c*Tokyo University of Agriculture and Technology, Japan*

*ivica.boticki@fer.hr

Abstract: In this paper, a model for identifying e-book reading style is proposed and applied onto a learning log dataset. Learning log data available as non-structured data source is processed to identify patterns of reading exhibited by users using three main structures: reading sessions, reads and passages. These structures are used to extract information on users' reading style to be used as part of user modeling process. The proposed model is applied on a set of log data generated by university students during one semester of digital resource use. The findings show students adopt predominantly receptive reading style, while responsive style occurs rarely. Further analysis revealed no significant relationships between reading style variables and student academic success for the Architecture course indicating the variables of responsive and receptive reading bring new information as part of user modeling.

Keywords: Reading styles, log data, user modelling, e-books

1. Introduction

E-books have become a popular medium for content delivery and are widely being accepted as an alternative to traditional paper-based sources. There is a number of advantages to e-books such as the convenience of use, greater interactivity etc. that make them valuable educational tools and resources today (Jagušt, Botički, & So, 2018). Nevertheless, more research is needed to ascertain how they affect users, and how can their interactive features be used in supporting learners.

There are two main log data subsets to be used as part of this study: the data on user reading and navigation within e-books, and the data on user interaction with e-books, such as making bookmarks, notes or memos. In order to prepare non-structured learning log data for the analysis, a process model with structures called sessions, reads and passages is utilized. Such data structures are used to model specific properties of e-book reading usage, facilitating further high-level e-book reading style analysis. It is to be noted that reading styles, as adopted in this paper, will be tied to contextual nature of e-book reading process. Such an approach is necessary to distinguish between potentially different conditions that arise from the course within an e-book content was being read, particular students using the e-book, or specific e-book resource being used. The contextual nature of the model is to allow for deeper understanding of the specifics of e-book use and its relationship with the identified user reading style.

2. Theoretical Background

Recognizing reading styles has been of great interest even before the appearance of cutting-edge technologies such as e-books and learning log mechanisms. Researchers had been trying to understand how humans read and thereby often relied on the use of methods such as the observation of reader physiognomic behavior, surveys and questionnaires, interview and diary studies (Freeman & Saunders, 2016; Marshall, 2009). Nevertheless, the issue of interfering with so called silent reading had presented a big methodological obstacle (Pugh, 1979), as reading is a personal and a reflective process.

Learning data logs are quantitative educational data, and they are used to meet the following objectives in the learning and teaching domains (Ogata et al., 2017): (1) Learning: Analyzing the details of behavior of “active learners” to make the students more active; and based on the relationships between log patterns and academic achievements, detecting the students who may drop out and those who will perform excellently; and (2) Teaching: Based on the logs made during a class session, improving course designs, which include collaborative learning and flipped classroom approaches and are based on the students’ patterns of viewing e-books (e.g., understanding which page was frequently viewed), improving teaching materials and the structure of the e-books.

We adopt the approach developed in (Freeman & Saunders, 2016), where reading style is generated from learning log data. The authors build on work by Thayer et al. (Thayer et al., 2011) to further interpret reading styles originally defined in (Pugh, 1979) and devise the following classification:

- *Receptive reading* - reading sequentially from beginning to end with little variation in pace, to find out what an author has to say
- *Responsive reading* - active engagement with arguments in the text, with frequent changes of pace, pauses, rereading

In (Thayer et al., 2011) specific kind of e-books used as part of the study “were well suited to receptive reading, searching, and scanning, but did not support responsive reading and skimming well at all”. We further explore the notion of receptive and responsive reading by using an e-book platform developed at Kyoto university (Flanagan & Ogata, 2018) with added interactive features, such as adding bookmarks, memos and markers.

3. Learning Log Processing Model

The process is based on the work presented in (Freeman & Saunders, 2016) and has several steps used to extract user sessions which represent contiguous block of e-book usage. Sessions are then used to create higher-level structures such as reads, passages and passage pairs. These are input into the algorithms for reading style identification, to be subsequently used as part of user modeling.

3.1 Sessions

In order to allow for high-level analysis of learning log data, sessions, reads, passages and paired passages are used (Freeman & Saunders, 2016). Sessions occur naturally as users progress in reading an e-book. Every session presents one time-limited usage of an e-book in which its reader traverses e-book pages or interacts with the e-book by adding bookmarks, notes, memos and similar. Figure 2 gives examples sessions generated from e-book usage log data.

As illustrated in Figure 2, sessions are composed of continuous events that last for a limited period of time and include any possible sequence of page changes: they can represent reading page by page, include jumps forward or backwards etc. Session is a useful structure since it organizes learning log data according to user interaction with a specific e-book resource in a limited amount of time and usually within the same reading context (i.e. covering 10 pages in a mathematics book on Thursday evening).

3.2 Reads and Paired Passages

Reads are used to merge sessions data coming from one user reading one resource. They are used to model the overall reading of a single resource done by one user and indicate reading jumps rather than concrete resource pages. Figure 3 shows reads structure created upon sessions laid out in Figure 2.

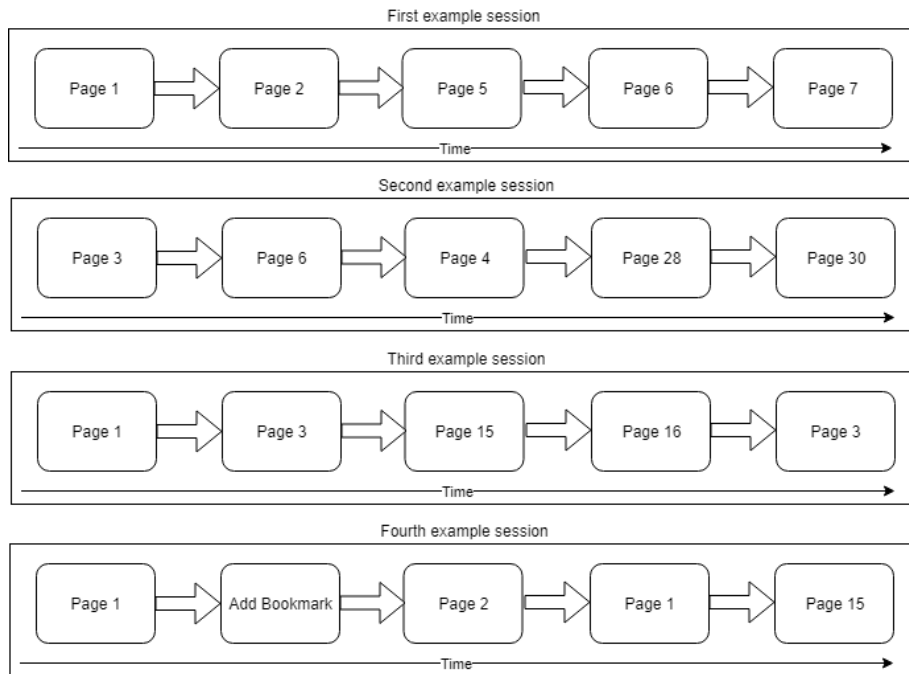


Figure 2. Example sessions extracted from learning logs.

In order to interpret the reads data in terms of reading dynamics (consecutive reading or jumps), interpretation notations names passages and paired passages are used (Freeman & Saunders, 2016). This notation focuses on the length and orientation of jumps differentiating between short, long, forward and backward jumps. Table 1 lists passages and paired passages created from reads in Figure 3.

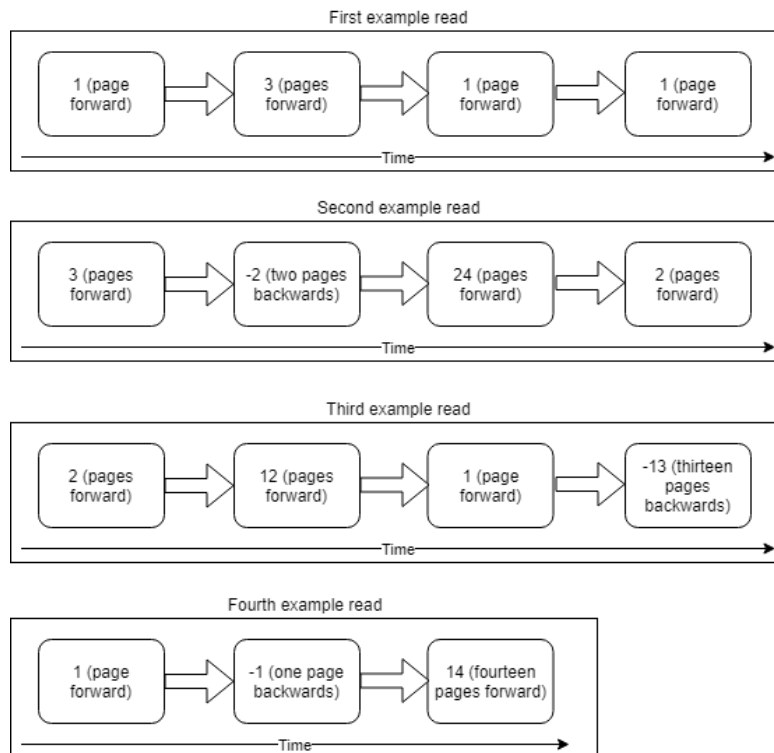


Figure 3: Reads created from sessions from Figure 2.

Passages are shortened representation of reads and they model five main jump categories: FOR (one or two pages skipped), SJF (small jump forward), BJF (big jump forward), SJB (small jump backwards) and BJB (big jump backwards). The first number of each passage is the number of pages

jumped. The second is the number of continuous pages read after the jump. Paired passages combine passages into pairs whereby immediate change in reading direction can be identified.

Table 1

Passages and paired passages created from reads shown in Figure 3

Example	Passages	Paired passages
First example	FOR(1,1), SJF (3,3)	(FOR-SJF)
Second example	SJF(3,1), SJB(-2,1), BJF(24,2)	(SJF-SJB), (SJB-BJF)
Third example	FOR(2,1), BJF(12,2), BJB(-13,1)	(FOR-BJF), (BJF-BJB)
Fourth example	FOR(1,2), BJF(14,1)	(FOR-BJF)

3.3 Extracting Reading Styles

By following the proposed classification of reading styles, two variables are being modeled: the receptive reading variable and the responsive reading variable. Paired passages are primarily used to identify the level of receptive reading of students by calculating the proportion of forward-oriented paired passages in reading an e-book resource, which are generated from all recorded sessions in which user interact with the resource.

Receptive reading (x) = number of forward oriented paired passages (x) / total number of paired passages (x)
where x stands for a specific time period.

Responsive reading is extracted from e-book session data by using information on interactive events which appear during e-book session (i.e. creating memos, bookmarks or makers). It is modeled using the following formula:

Responsive reading (x) = (number of memo operations (x) + number bookmark operations (x) + number of marker operations (x)) / total number of pages read (x)
where x stands for a specific time period.

4. Case Study

In this study, data sample coming from the BookRoll e-book learning system developed at Kyoto University in Japan [1] is analyzed. The presented process model is applied onto one semester of usage log data generated by 407 students from 2017-09-09 to 2018-02-05. There are in total 6546 sessions generated from NoSQL log entries recorded while the students engaged in educational activities with 48 resources (e-books). The identified sessions were used to generate 1486 reads structures, whereby the number of associated users was reduced to 331 and the number of resources to 33 (the reduction took place due to elimination of trivial sessions by some users who did not engage significantly in e-book usage).

Table 2 shows the detailed analysis data generated during the process model steps execution. The results indicate that the average student read period (time from the first to last access to a resource) was about 4 days, with the average resource read period being around 3 days. It is to be noted that reads are continuous structures that map to a specific resource and that one resource is usually taught during one week of lectures. All read periods lengths come with high standard deviations indicating high between-student differences in total period spent on a resource.

In terms of the receptive reading style variable, values are uniform across general, user and resource categories (means range from 0.83 to 0.86) with observably smaller standard deviation values. This indicates reading is done in fairly linear fashion with relatively small proportion of large jumps across the reading resource. On the other hand, the responsive reading variable values indicate there is

about one interactive event per 10 pages of student reading, with large observed between-student differences. Correlation analysis was done to test for possible relationships between the variables of read period time, receptive, responsive reading, and the final course grade. Results are shown in Table 3 and Table 4.

Table 2

The detailed analysis results while applying the process model, including user reading style data

	AVG	STDEV
Reads per student	4.49	2.75
Reads per resource	45.00	50.48
Reads per student per resource	1	0
Read period (s)	340,367	970,262
Read period per student (s)	400,334	820,189
Read period per resource (s)	236,437	391,596
Receptive reading	0.85	0.16
Receptive reading per student	0.83	0.13
Receptive reading per resource	0.86	0.10
Responsive reading	0.01	0.09
Responsive reading per student	0.01	0.1
Responsive reading per resource	0.02	0.04

Table 3

Results of correlation analysis for resource read period time, and receptive and responsive reading style variables

	Receptive reading	Responsive reading
Resource read period	-0.054*	0.023
Receptive reading	1	-0.033

*p<0.05

The correlation analysis results indicate all correlation values are low and that there is only one significant correlation coefficient (between receptive reading and resource read time), however of weak power $r=-0.054$.

For the purposes of in-depth analysis of reading style at a specific course level (Botički, Budiščak, & Hoić-Božić, 2008), a course taught in the logged time period was chosen – the Architecture course. Although the final course grade in the Architecture course correlated with the overall read period time, which is in line with prior research on usage time of e-books (Oi, Okubo, Shimada, Yin, & Ogata, 2015), no significant and strong correlations were identified between the final grade and the receptive and responsive reading variables.

Table 4

Results of the correlation analysis: final grades in the Architecture course, and receptive and responsive variables

	Final grade
Resource read time	0.237**/0.352**
Receptive reading	0.077/-0.077
Responsive reading	-0.034/-0.234

**p<0.001

There was observable non-significant negative correlation between the final course grades and the responsive reading variable possibly indicating students with such reading style perform poorly on

the final exam, but more research is needed to confirm this (units of are analysis one resource per one student in the Architecture course/all resources per one student in the Architecture course).

5. Conclusions

In this paper learning log data was analyzed via the adopted process model in order to extract usage sessions, reads, read passages and read passage pairs. Both sessions and read passage pairs were then used as means of defining reading style variables, with the final aim of including these variables in user modeling.

The results of the analysis indicate that students predominantly utilize linear, forward-oriented reading of e-book materials and have relatively low interaction with the material, although interactive features of e-books are readily available for their use. Further correlation analysis indicates that the receptive and responsive reading variables do not correlate with the student final course grades.

Limitations of the study stem from the characteristics of the data sample being used, which covers only undergraduate students in academic environments. Future research will focus on alternative approaches to reading style variables extraction including the use of sub-symbolic classifiers (Pessa, n.d.) and on identifying possible educational activities with e-books which rely on the usage of reading style variables to enhance some elements of the educational process.

Acknowledgements

This work is supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304.

References

- Botički, I., Budiščak, I., & Hoić-Božić, N. (2008). Module for online assessment in AHyCo learning management system. *Novi Sad J. Math*, 38(2), 115–131.
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management and E-Learning*, 10(4), 469–484.
- Freeman, R. S., & Saunders, E. S. (2016). E-Book Reading Practices in Different Subject Areas: An Exploratory Log Analysis. In *In Academic E-Books: Publishers, Librarians, and Users* (pp. 223–248). Purdue University Press.
- Jagušt, T., Botički, I., & So, H. J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers and Education*, 125, 444–457. <https://doi.org/10.1016/j.compedu.2018.06.022>
- Marshall, C. C. (2009). *Reading and Writing the Electronic Book. Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 1). <https://doi.org/10.2200/S00215ED1V01Y200907ICR009>
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., ... Hirokawa, S. (2017). Learning analytics for E-book-based educational big data in higher education. In *Smart Sensors at the IoT Frontier* (pp. 327–350). https://doi.org/10.1007/978-3-319-55345-0_13
- Oi, M., Okubo, F., Shimada, A., Yin, C., & Ogata, H. (2015). Analysis of preview and review patterns in undergraduates' e-book logs. *Proceedings of the 23rd International Conference on Computers in Education, ICCE 2015*, 166–171. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84979681388&partnerID=40&md5=a0cb568ab33b35c83c91ad19e635aa28>
- Pessa, E. (n.d.). Symbolic and subsymbolic models, and their use in systems research. *Systems Research*, 11(3), 23–41. <https://doi.org/10.1002/sres.3850110303>
- Pugh, A. K. (1979). Styles and Strategies in Silent Reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language* (pp. 431–443). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-0994-9_27
- Thayer, A., Lee, C. P., Hwang, L. H., Sales, H., Sen, P., & Dalal, N. (2011). The Imposition and Superimposition of Digital Reading Technology: The Academic Potential of E-readers. *CHI 2011, Session: Reading and Writing*, 2917–2926. <https://doi.org/10.1145/1978942.1979375>

Exploring the Relationships between Students' Engagement and Academic Performance in the Digital Textbook System

Gökhan AKÇ APINAR^{a,b*}, Mohammad Nehal HASNINE^c,

Rwitajit MAJUMDAR^a, Brendan FLANAGAN^a & Hiroaki OGATA^a

^a *Academic Center for Computing and Media Studies, Kyoto University, Japan*

^b *Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*

^c *Dept. of Computer and Information Sciences, Tokyo University of Agriculture & Technology, Japan*

*akcapinar.gokhan.2m@kyoto-u.ac.jp

Abstract: In this paper, we analyzed the relationships between students' engagement and academic performance in the digital textbook system. To measure students' engagement, we first extracted features from the students' digital textbook reading logs (click-streams) that represent their engagement with the contents. Then, we used percentile rank transformation to create normalized engagement scores and an overall engagement score. In the analysis, we first investigated the correlation between engagement scores' and the students' final scores. Second, we modeled students' transition patterns from the engagement to academic performance by using Markov Chains. Third, we analyzed engagement patterns of the students with different academic performance levels. Our results showed that there is a positive moderate correlation between students' academic performance and their engagement with digital textbooks. Our results also revealed that a single engagement score can be used to measure students' engagement with the system, which is easy to understand by non-expert users. We also introduced our dashboard interventions that are developed based on this engagement score.

Keywords: Engagement, academic performance, digital textbook, reading pattern, engagement pattern, markov model

1. Introduction

Students' engagement with the learning environment is closely related to learning outcomes (Hu & Li, 2017; Lu, Huang, Huang, & Yang, 2017). Therefore, regular monitoring of students' engagement is crucial for timely interventions, particularly for at-risk students. However, measuring various dimensions of students' engagement from their learning traces is still challenging task for researchers. In particular, creating engagement scores which are easy to interpret by students and instructors.

In this paper, we used students' digital textbook reading data to create metrics for measuring their engagement with the digital textbook system. We investigated relationships between students' academic performance and their engagement score. We also compared weekly engagement patterns of students with different academic performance levels. Finally, we introduced the dashboard interventions that we developed based on this engagement score for assisting at-risk students.

1.1 Digital Textbook Data Analysis

Digital textbook systems made possible to collect data related to student reading and note-taking behavior that are not possible to capture with paper books (Abaci, Morrone, & Dennis, February 22, 2015). This data previously used to predict students' learning outcomes (Akçapınar, Hasnine, Majumdar, Flanagan, & Ogata, 2019; Hasnine et al., 2018), to analyze students' weekly reading patterns (Akçapınar, Majumdar, Flanagan, & Ogata, 2018), and for the visualization purposes. Junco and Clem (2015) also found that digital textbooks-based interaction data is a stronger predictor of students' learning outcome. Their study also highlighted that students, those who spent longer time in reading textbooks earned higher grades in the course over to those who spent less time.

1.2 Engagement and Learning

Fredricks, Blumenfeld, and Paris (2004) have proposed that engagement is studied as a multifaceted construct, with behavioral, affective, and cognitive dimensions. *Behavioral engagement* refers to students' involvement in academic activities. *Emotional engagement* includes positive and negative reactions to teachers, classmates, academics, and school. *Cognitive engagement* refers to the effort required to master difficult skills and understand complex ideas.

Behavioral engagement is considered crucial for achieving positive academic outcomes and preventing dropping out (Fredricks et al., 2004). In this study, we focused on students' behavioral and cognitive engagement with the digital textbook system and we extracted features from the interaction logs related to students' reading behaviors. In other words, we used learning logs as a proxy to measure students' engagement with the system and analyzed relationships between their engagement and academic performance.

2. Method

2.1 Data Collection

In this study, data was collected from a digital textbook system called BookRoll. The BookRoll allows students to read digital contents. It has a feature like red or yellow markers to highlight some parts of the text. Students' can add memos to remember important points or bookmark pages to access them easily while they are reviewing the content. We analyzed more than 65,000 click-stream data that are collected from 72 students registered in an Elementary Informatics Course at a university. The course was offered to first-year undergraduate university students. Students used the BookRoll system to access course materials that were uploaded by the instructor once a week. Students were given 13 contents in different weeks of the course and their anonymized interactions (e.g. next, previous, jump, highlight, adding memo, etc.) were recorded by the system. Students' academic performance was evaluated based on eight quizzes across the semester and one final exam.

2.2 Preprocessing of Data

At the beginning of data preprocessing, features from the click-stream data were extracted. Extracted features were used to measure students' engagement. In feature selection, students' behavioral (e.g. total number of events, number of times open the system, etc.) and cognitive (e.g. red/yellow marker usage counts, memo counts, etc.) engagement with the system were taken into consideration. A brief description of all features is given in Table 2.

Table 2

Description of Features

Features	Description
Total event	Total number of events
Time	Total time spent on the BookRoll system in minutes
Unique Day	Number of different days that student use the system
Long event	Number of events longer than 3 seconds
Short event	Number of events less than or equal to 3 seconds
Next	Number of Next events
Previous	Number of Previous events
Open	Number of times that student open the system
Jump	Number of Jump events
Red marker	Number of red markers added by the student
Yellow marker	Number of yellow markers added by the student
Memo	Number of memos added by the student
Score	Final scores of the students

After extracting features, percentile rank transformation was applied to the raw data to create generalized engagement metrics. Percentile rank transformation was chosen as a method to create generalized engagement score because it can be used in various educational purposes such as comparing students' engagement in different courses, identifying at-risk students, monitoring their engagement, etc. Formula 1 was used to calculate percentile rank (PR) scores where f_b is the number of scores which are less than the score value of the percentile rank, f_w is the number of scores which have the same value as the score value of the percentile rank, and N is the number of scores. The percentile rank measures range from 0 to 1. Single engagement score was also calculated by taking the average of all engagement metrics and compared its performance against the other metrics.

$$PR = \frac{f_b + \frac{1}{2} f_w}{N} \quad (1)$$

2.3 Data Analysis

To analyze the relationship between engagement scores and students' academic performance, correlation analysis was performed. Here, the Pearson correlation coefficient between engagement metrics and students' final scores was calculated. In addition, to analyze the relationship between students' engagement score and their academic performance, Markov Chain analysis was performed. We also visually analyzed long term engagement patterns of students with different level of academic performance.

3. Results and Discussion

3.1 Correlation Analysis

Results of the correlation analysis between raw scores, transformed scores, and students' final scores is shown in Fig. 1. In raw data, total event, total time, unique day, long event, short event, next, previous, and open metrics' correlations with the final score are significant ($p < 0.01$). With the transformed data metrics' expect jump and red marker correlation with the final score are significant ($p < 0.01$). As shown the Fig. 1, almost all metrics' correlation increased after the transformation.

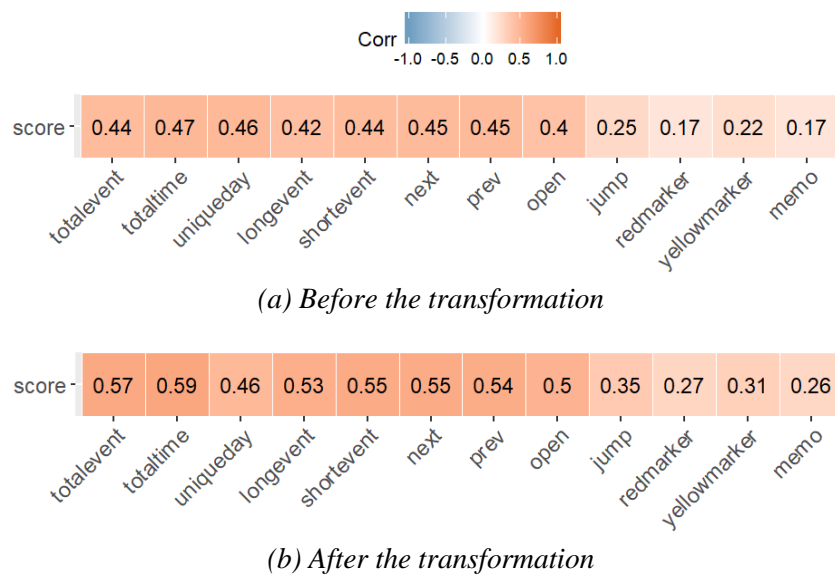


Figure 1. Engagement features correlation with the students' final scores before and after transformation.

In terms of single engagement score that was derived from the other metrics, a moderate ($r = 0.59$, $p = 0.01$) positive correlation between the engagement score and final score variable was noted. In Fig.2, with a scatterplot, we summarized the results.

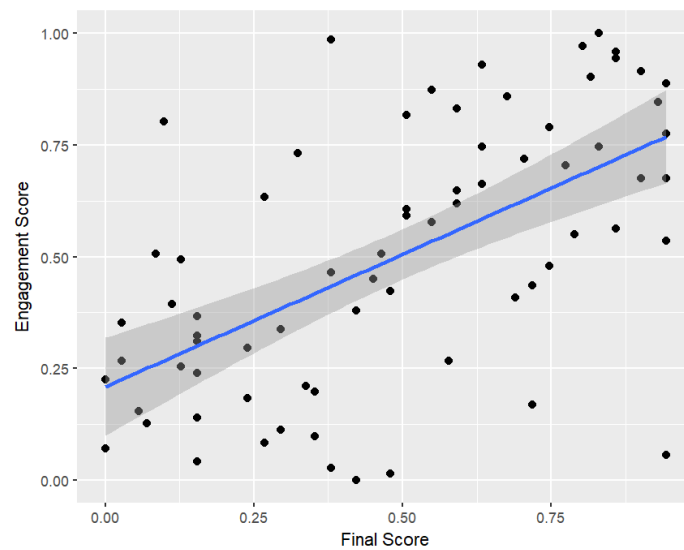


Figure 2. Correlation between Engagement Score and Final Score.

3.2 Markov Chain Analysis

We divided students into two groups based on their academic performance and single engagement score. Then, analyzed transitions between the level of engagement and level of academic performance (Fig. 3). We grouped them as Low Performers (final score $\leq .50$, $n = 36$), Low Engagement (engagement score $\leq .50$, $n = 36$), High Performers (final score $> .50$, $n = 36$), and High Engagement (engagement score $> .50$, $n = 36$). According to Fig. 3, low-engaged student’s probability to get a low score is 0.83 (83%) while probability to get a high score is 0.17 (17%). A similar pattern is observed for high-engaged students. The probability of getting a high score is 0.83 (83%) while probability to get a low score is 0.17 (17%). In other words, students are most likely to get scores related to their level of engagement.

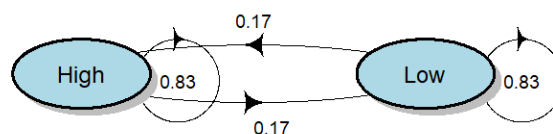


Figure 3. Transitions between engagement level and academic performance.

3.3 Engagement Pattern Analysis

To understand students’ engagement patterns across the semester, we chose four students with different academic performance levels (two high performers, two low performers) and visually compared their engagement patterns. Visualization of the weekly engagement patterns are shown in Fig. 4. Top two graphs show the weekly engagement of the high performer students while the last two show the low performer students’ engagement. In each graph, the grey line shows the class average. High performer students’ engagement is higher than the class average across the semester, and their average engagement is higher than 80% of the class. Low performer students, however, have much less activity than class average. For example, students in (c) were only active at the beginning and at the end of the semester. Students in (d) were only active for the first three weeks of the class and there is no activity after that. These graphs can be used to identify at-risk students at the beginning of the semester;

however, further investigation is required to see how common these patterns are among other low and high performers.

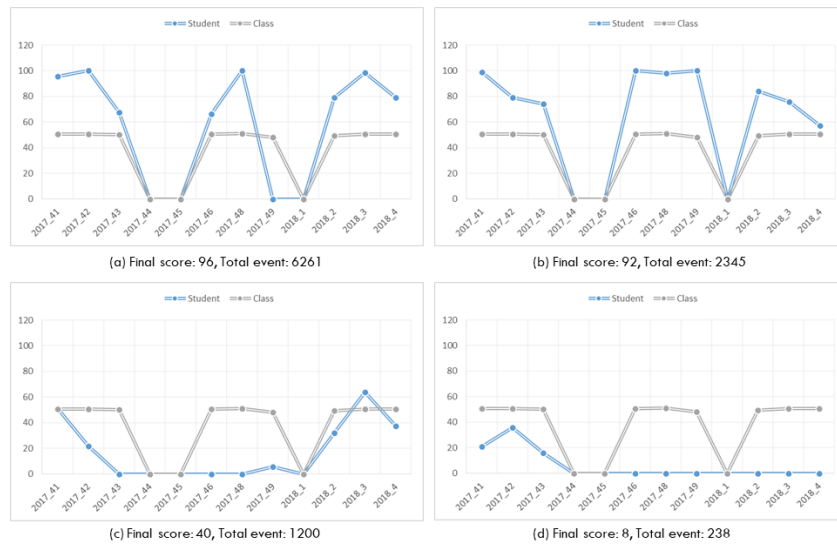


Figure 4. Students' weekly engagement patterns across the semester.

4. Learning Analytics Dashboard

By using the single engagement score, we developed three interventions for students and implemented them in our learning analytics dashboard (Fig. 5). These interventions are in the form of e-mail feedback, engagement score graph, and weekly engagement graph.

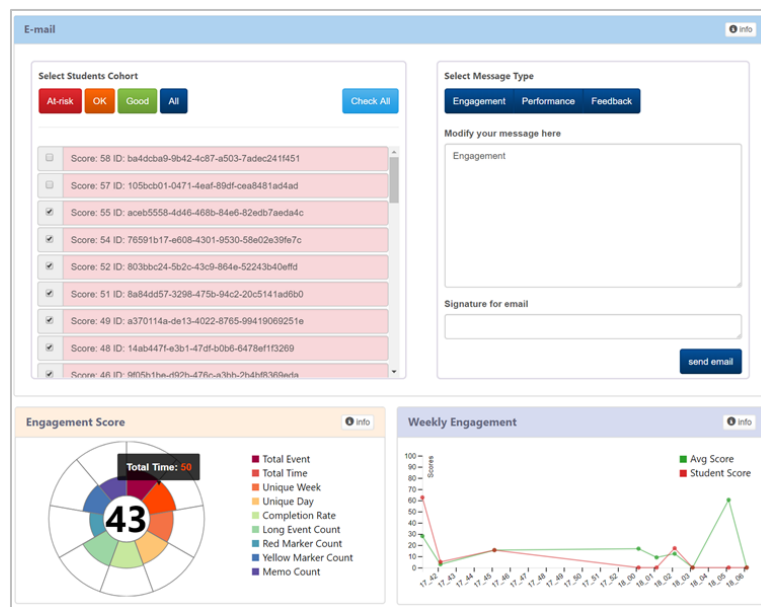


Figure 5. Learning analytics dashboard.

In the e-mail feedback module, instructors can filter and select students based on their engagement scores (e.g. at-risk, ok, good, all) and can send them an e-mail either using predefined templates or adding their own message. Engagement score graph shows students' engagement scores in each metrics (e.g. total time, total event, etc.) and his/her average engagement with the selected content. Weekly engagement graph shows students' average weekly engagement scores and class average. The

e-mail feedback module is developed only for instructors, students can only see engagement score graph and weekly engagement graph.

5. Conclusions

In this study, we aimed at analyzing relationships between students' academic performance and their engagement in the digital textbook reader system. We extracted features from students' reading logs to measure their behavioral and cognitive engagement with digital textbooks. Our results revealed that transformed scores' correlation with the final scores is higher than raw engagement scores. Moreover, a single engagement score derived from engagement metrics performed similarly in terms of relationships with the students' performance. A newly generated metric is also easy to understand by non-expert users.

Literature suggests that timely interventions planned based on engagement effective to increase learning outcomes (Arnold & Pistilli, 2012; Tanes, Arnold, King, & Remnet, 2011). Therefore, instructors can use our proposed single engagement score to monitor their students' engagement during the semester and can give them feedback accordingly. Markov model showed that students are most likely to exhibit academic performance close to their engagement level. This model can be used to predict students' academic performance based on their level of engagement. In the future, we will test the effects of these models along with developed dashboard interventions on students' engagement and academic performance. Our future studies will investigate more on various dimensions of engagement (such as emotional engagement) that can be considered while modeling learners' engagement.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S) 16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

- Abaci, S., Morrone, A., & Dennis, A. (February 22, 2015). Instructor engagement with e-texts. *Educause Review*. Retrieved from <https://er.educause.edu/articles/2015/2/instructor-engagement-with-etexts>
- Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an Early-Warning System for Spotting At-Risk Students by using eBook Interaction Logs. *Smart Learning Environments*, 6(4), 1-15. doi:doi.org/10.1186/s40561-019-0083-4
- Akçapınar, G., Majumdar, R., Flanagan, B., & Ogata, H. (2018). *Investigating Students' e-Book Reading Patterns with Markov Chains*. Paper presented at the 26th International Conference on Computers in Education (ICCE2018), Manila Philippines.
- Arnold, K. E., & Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success*. Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, British Columbia, Canada.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59-109. doi:10.3102/00346543074001059
- Hasnine, M. N., Akçapınar, G., Flanagan, B., Majumdar, R., Ogata, H., & Mouri, K. (2018). *Towards Final Scores Prediction over Clickstream Using Machine Learning Methods*. Paper presented at the 26th International Conference on Computers in Education (ICCE2018), Manila Philippines.
- Hu, M., & Li, H. (2017, 27-29 June 2017). *Student Engagement in Online Learning: A Review*. Paper presented at the 2017 International Symposium on Educational Technology (ISET).
- Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education*, 27, 54-63. doi:<https://doi.org/10.1016/j.iheduc.2015.06.001>
- Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2017). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments*, 25(2), 220-234. doi:10.1080/10494820.2016.1278391
- Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. *Computers & Education*, 57(4), 2414-2422. doi:<https://doi.org/10.1016/j.compedu.2011.05.016>

Using Clickstream to Understand Learning Paths and the Network Structure of Learning Resources: Using MOOC as an Example

Ming GAO^a, Jingjing ZHANG^{b*}, Di SUN^c & Jiang ZHANG^d

^a*Research Centre of Distance Education, Beijing Normal University, China*

^b*Big Data Centre for Technology-mediated Education, Beijing Normal University, China*

^c*IDDE, Syracuse University, USA*

^d*School of Systems Science, Beijing Normal University, China*

* jingjing.zhang@bnu.edu.cn

Abstract: Massive open online courses (MOOCs) have attracted great attention from the public and learners. However, their high dropout rates have been criticized over the past few years. There has been a body of work dedicated to using learning analytics to provide feedback for instructors and designers to improve learners' engagement and retention rates. However, the open and flexible nature of MOOCs has often been overlooked in these analytics studies. In this work, we used clickstream data to construct a flow network model in order to identify MOOC learners' learning paths and the network structure of available learning resources from an open system perspective. We found that learners tend to adopt linear learning paths within chapters and continue to watch video lectures in next chapters instead of taking quizzes at the end of the chapters, and they rarely review previous chapters during studies. We also found that learning paths of all learners have formed a centralised network structure which implies that certain learning resources in such a network structure dominate the way in which learners learn in MOOCs.

Keywords: MOOCs, learning behaviour, learning paths, network structure, learning analytics

1. Introduction

Massive open online courses (MOOCs) play an important role in higher education (Li & Stephen, 2013). MOOCs gather global education resources and provide opportunities for learners to take online courses from prestigious universities. In the early stages of MOOC development, MOOCs on Udacity, Coursera, and edX platforms attracted more than 100,000 registrants per course (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014). However, there have also been high dropout rates, with only a very small numbers of registrants completing most MOOCs successfully (Koller, Ng, Do, & Chen, 2013). It is critical to find out more about how to keep learners engaged and how to improve the dismal retention rates. A great deal of recent work has been done using machine learning methods to predict the learners who are most likely to drop out (e.g., Halawa, Greene, & Mitchell, 2014; Moreno-Marcos, Alario-Hoyos, Munoz-Merino, & Delgado Kloos, 2018; Xing & Du, 2018). While the accuracy rate of dropout prediction is high, the explanatory power of such studies is low. The selected variables, which have a strong impact on prediction, may not provide enough information for course designers and instructors to adapt MOOC design or content responsibly. Some research has addressed this by highlighting the importance of analysing learners' learning paths to provide more practical feedback for instructors and designers (Davis, Chen, Hauff, & Houben, 2016). For example, Wen and Rose (2014) investigated learners' habits through their sequential use of learning activities. Nevertheless, in analysing the learning paths of individuals, only a few studies have taken the open and flexible nature of MOOCs into consideration (e.g., Zhang, Lou, Zhang, & Zhang, 2019). For instance, a learner's use of time is not always continuous once they start online learning: they can stop watching a video or pause it, then restart or close it whenever they like.

In this work, we attempt to understand MOOC learners' learning paths and to explore the network structure of learning resources by modelling a flow network from learners' clickstream data. Specifically, we focus on analysing learners' learning behaviour in relation to learning resources

consisting of lecture videos and assignments, attempting to discover learners' learning behaviour or habits and to measure the network structure of learning resources from an open perspective.

2. Methodology

2.1 MOOCs Dataset

The dataset used in this work was gathered from a course run in XuetangX, titled “Introduction to Psychology” (2015 autumn semester). It contained 12 chapters (68 learning resources, including videos and assignments) and a final exam. The design of this course implied a traditional linear trajectory through the learning material (see Figure 1). Each chapter commenced with a video lecture, followed by an assignment to evaluate what the learners had attained from this chapter. Nearly 20,000 learners had registered for the course but only 10,359 actually accessed it.

In the dataset, learner ID, timestamp (the time at which the page was opened/closed), URLs, page title, stay time, and page type were saved in each log. Learners and learning resources could be identified uniquely through the learner ID and URLs. To take into account the flexibility of the learners' learning time, the timestamp was used to extract their learning sessions through sorting and segmenting their clickstream data. As a rule of thumb, online behaviours that occurred over 25.5 minutes apart were determined as separate sessions (Catledge & Pitkow, 1995). The 30-minute threshold was used in this study to delineate a sequence as a new session (see Figure 2). In other words, in cases where there was 30 minutes between the use of two different resources, we assumed that the learners had stopped learning from the former resource. Stay time was used to filter invalid click access, and page titles and types provided the basic descriptions of a web page.

Some of these course learners used mobile devices to access the course. Their data were not included in the analysis due to incomplete formats. Some logs associated with zero seconds (0s) of stay time on websites and other resources (except video lectures and assignments) were also excluded from the data analysis. After finishing some pre-processing work, 5,506 users with at least 1 second of log data who used web browsers to visit the course (63,060 logs) were included.



Figure 1. Screenshot of “Introduction to Psychology”

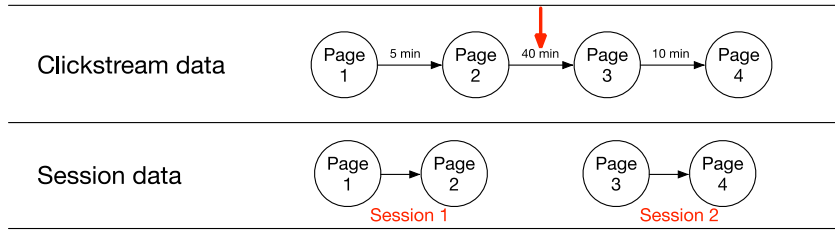


Figure 2. Transformation from clickstream data to session data

2.2 Flow Network Model

This work used the flow network model proposed by Zhang et al. (Zhang & Wu, 2013; Wu, Zhang, & Zhao, 2014) to model the openness of MOOCs. This model has been found useful in many different contexts, including assessing the impact of websites (Wu & Zhang, 2011), the “stickiness” of a web forum (Wu et al., 2014), and the geometric representation of Internet ecology (Shi et al., 2015). The flow network model is an open, directed, and weighted network in which nodes represent the websites and weighted edges represent traffics (the amount of users’ transition) between two websites. It can be constructed from the users’ clickstream data and its structure can be seen in Figure 3. The openness of the flow network means that it can connect the online and offline world by adding two special nodes, “source” and “sink”. The direction from the “source” to other nodes represents the point at which users go online; when users go offline, this is marked by the nodes with “sink”. In a balanced network, moreover, the amount of inflow of a node equals its outflow, except for the “source” and “sink”. As mentioned above, 68 nodes (the number of resources, including all video lectures and assignments) can be used to construct the flow network.

Based on the flow matrix hidden in the flow network, three indexes – A_i , C_i , and η – can be used to describe the attributes of this network. A_i measures the total flux of node i , and C_i measures the impact of node i on the whole network. After getting A_i and C_i for every node i , we can explore the existence of the allometric scaling law ($C_i \sim A_i^\eta$) (Jiang Zhang & Guo, 2010), a universal relationship found in river systems (Rodriguez-Iturbe & Rinaldo, 2001), living organisms (West, Brown, & Enquist., 1999), and international trade systems (Shi, Luo, Wang, & Zhang, 2013). The exponent η reflects the flow structure of the network. An example provided by Shi et al. (2013) is two networks with the same distribution of A_i having a different value of η . $A_i = [1, 2, 3, 4, 5]$, $C_i(1)=A_i^{0.5}=\{1, 1.4, 1.7, 2, 2.2\}$, $C_i(2)=A_i^2=\{1, 4, 9, 16, 25\}$. For the network with $\eta=2$, the largest two nodes – the fifth and fourth nodes ($25/(1+4+9+16+25)=45\%$, $16/(1+4+9+16+25)=29\%$ of impacts, respectively) – almost control the whole network. For the network with $\eta=0.5$, the largest node only dominates $2.2/(1+1.4+1.7+2+2.2) = 27\%$ of impacts. Thus, if $\eta > 1$, the network structure is centralised. In other words, some nodes control the flows of the network. If $\eta \leq 1$, that means the network structure is more decentralised. Each node plays the more equal role in the network. In the learning resources network, we can interpret A_i as the total amount of viewing frequency through resource i ; C_i reflects the influence of resource i on the entire network (both direct and indirect) and η reflects the flow structure of the resource network. The larger the η , the more centralised the flow structure (Wu & Zhang, 2013). More detail on these three variables is given in Zhang et al. (2010).

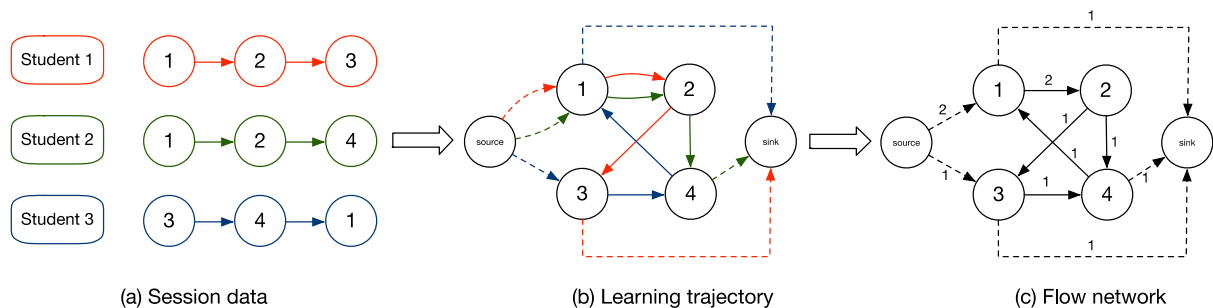


Figure 3. Process of flow network construction

3. Findings

The learning resource network was constructed based on students' clickstream data and visualised using visualisation software called Gephi (<https://gephi.org/>) (Figure 4(a)). The size of the nodes represents their importance (C_i). The colour of the nodes represents their state of flux. A red node indicates that the flux of outflow (excluding the flow to sink) was larger than the flux of inflow (excluding the flow from source), while a blue node indicates that its flux of inflow (excluding the flow from source) was larger than the flux of outflow (excluded the flux to sink). That is, red nodes show that there are more learners continuing to learn after they have completed this resource, and the blue nodes show that some of the learners leave these resources. The arrows between two nodes represent a learner's learning direction (learning path) between resources. The thickness of the edges indicates the viewing frequency.

3.1 Learning Paths

We almost found that red nodes (the flow of the outflow is larger than the inflow) are the first resource of each chapter (see Figure 4). This means that some learners might keep on learning after the first video lecture, while some may leave during any video lecture or assignment within the same chapter.

Given the non-formal nature of MOOCs, learners can determine their way of learning and do not have to follow the predesigned learning sequence, for example, employing nonlinear navigation (Guo & Reinecke, 2014). The complex links between resources show the nonlinear trajectories to some extent (see Figure 4(a)). However, almost all the thicker directed edges appear within the chapter. This indicates that most of learners still follow the predesigned learning path – a linear learning path – especially within each chapter.

To highlight the relationship between resources, we filtered some edges and left the backbone edges based on learners' frequent transitions between resources (shown in Figure 4(b)). No isolated point was left when the lower limiting value of edges was reduced to 133.344. Two separate learning resource communities (Chapters 11 and 12) were found with no links to other chapters and with thin edges. This depicts the learning paths of learners who finish the course successfully; these are the last two chapters of the course, and the topics were broken up into weeks and released one chapter at a time per week. On the other hand, this may reflect that some learners were just interested in the topic of these two chapters. We also found links between the last nodes of some chapters, for example between Chapter 4 and Chapter 10. The last node (resource) of a chapter was an assignment aiming to help learners evaluate their learning performance in this chapter. It indicates that some learners intend to continue watching the next chapter's video lectures instead of testing their learning performance at the end of a chapter. This was also verified by the links between the last-but-one resource of former chapters and the first resource of later chapters. It is interesting to see that no links return to former nodes. This indicates that fewer learners reviewed former content. Some activities and assignments that embed or mix previous content can be designed to help learners review and enhance what they have learned.

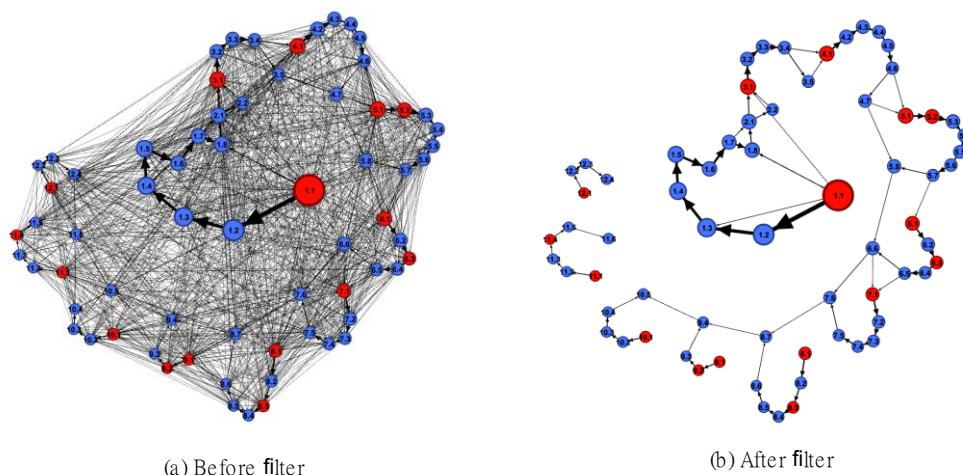


Figure 4. The flow structure of resources. The size of a node stands for the impact of the resource (C_i). The width of the line means the amount of flux. The colour represents the difference between inflow

and outflow: red represents that the flux of its outflow is bigger than the inflow; blue represents that the flux of the outflow is smaller than the inflow

3.2 The Network Structure of Learning Resources

As shown in Figure 4(b), some of the largest nodes (larger value of C_i), which reflect their importance, were almost all located in the first chapter, which gave an introduction to psychology. This indicates that the introduction may play an important role in the course learning resources network.

Further, we explored the impact of nodes on the flux of whole flow network. In other words, we examined whether some resources play an important role in connecting other resources to affect a learner's learning or understanding by exploring the relationship between the size of a node and its impact. As shown in Figure 5, it shows the A_i and C_i values of all resources on the log-log plot, and they have been fitted with a line. The minimum square error (R^2) was used to evaluate the performance of fitness. The value of R^2 is 0.93, which shows a high degree of fit. The exponent (η) is 1.08. This shows that the flow network structure is centralised that some resources, represented above the red line in Figure 5, have a larger impact (C_i) on the whole network than estimated. Namely, some resources affect a learner's learning or understanding through connecting related or necessary resources.

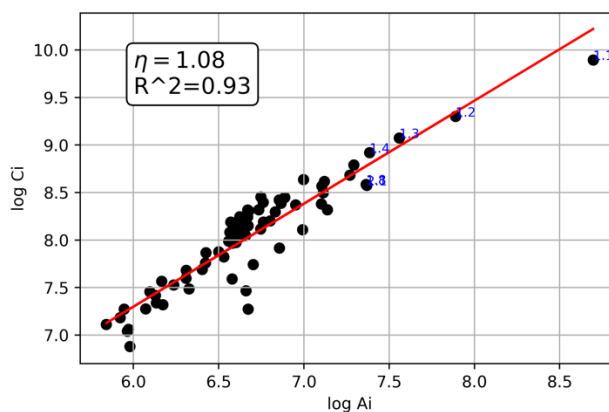


Figure 5. The power-law relationship between A_i and C_i for courseware

4. Conclusion and Future Work

In this study, we attempted to use a flow network model to improve our understanding of MOOC learners' learning behaviour and the relationship or structure of learning resources from an open perspective. We conducted an in-depth analysis on learners' learning paths and the relationship between the course resources. Some interesting results emerged from the learning path analysis, for example that learners tend to adopt a linear learning path within chapter learning; few learners review or enhance what they have learned. We also found that some resources play a role in connecting other resources, with an effect on learners' learning and understanding. In future, we will further explore learners' learning paths through different learning achievement groups and explore the learning resources structure in different subject categories to provide more practical feedback for course instructors and designers.

Acknowledgements

The authors would like to thank XuetangX for providing us with the clickstream data for this research. This project is funded by Student Research Foundation of Faculty of Education, Beijing Normal University [Project No. 1812201].

References

- Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27(6), 1065–1073. [https://doi.org/10.1016/0169-7552\(95\)00043-7](https://doi.org/10.1016/0169-7552(95)00043-7)
- Davis, D., Chen, G., Hauff, C., & Houben, G.-J. (2016). Gauging MOOC learners' adherence to the designed learning path. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 54–61). Retrieved from <http://www.derecho.unam.mx/cultura-juridica/eventos/programa-2-Curso-DyL.pdf>
- Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14* (pp. 21–30). <https://doi.org/10.1145/2556325.2566247>
- Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. In *Proceedings of the European MOOC Stakeholder Summit 2014* (pp. 58–65). Retrieved from <http://www.emoocs2014.eu/sites/default/files/Proceedings-Moocs-Summit-2014.pdf>
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in Massive Open Online Courses: in depth. *Educause Review*, 48(3), 62–63.
- Li, Y., & Stephen, P. (2013). *MOOCs and Open Education: Implications for Higher Education*. *Cetis*. <https://doi.org/10.4324/9781315751108-1>
- Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., & Delgado Kloos, C. (2018). Prediction in MOOCs: a review and future research directions. *IEEE Transactions on Learning Technologies*, <https://doi.org/10.1109/TLT.2018.2856808>
- Rodriguez-Iturbe, I., & Rinaldo, A. (2001). *Fractal River Basins: Chance and Self-Organization*. Cambridge: Cambridge University Press.
- Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a Massive Open Online Course? *Communications of the ACM*, 57(4), 58–65. <https://doi.org/10.1145/2500876>
- Shi, P., Huang, X., Wang, J., Zhang, J., Deng, S., & Wu, Y. (2015). A geometric representation of collective attention flows. *PLoS ONE*, 10(9), 1–21. <https://doi.org/10.1371/journal.pone.0136243>
- Shi, P., Luo, J., Wang, P., & Zhang, J. (2013). Centralized flow structure of international trade networks for different products. In *2013 International Conference on Management Science and Engineering 20th Annual Conference Proceedings* (pp. 91–99). IEEE. <https://doi.org/10.1109/ICMSE.2013.6586267>
- Wen, M., & Rose, C. P. (2014). Identifying latent study habits by mining learner behavior patterns in Massive Open Online Courses. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp. 1983–1986). New York: ACM. <https://doi.org/10.1145/2661829.2662033>
- West, G. B., Brown, J. H., & Enquist, B. J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science*, 284(5420), 1677–1679.
- Wu, L., & Zhang, J. (2011). The decentralized structure of collective attention on the Web. *ArXiv Preprint ArXiv:1110.6097*, 1–12.
- Wu, L., & Zhang, J. (2013). The decentralized flow structure of clickstreams on the web. *European Physical Journal B*, 86(6), 1–6. <https://doi.org/10.1140/epjb/e2013-40132-2>
- Wu, L., Zhang, J., & Zhao, M. (2014). The metabolism and growth of web forums. *PLoS ONE*, 9(8), 1–11. <https://doi.org/10.1371/journal.pone.0102646>
- Xing, W., & Du, D. (2018). Dropout Prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*. 57(3), 547-570. <https://doi.org/10.1177/0735633118757015>
- Zhang, J., Lou, X., Zhang, H., & Zhang, J. (2019). Modeling collective attention in online and flexible learning environments. *Distance Education*, 40(2), 278-301. <https://doi.org/10.1080/01587919.2019.1600368>
- Zhang, J., & Guo, L. (2010). Scaling behaviors of weighted food webs as energy transportation networks. *Journal of Theoretical Biology*, 264(3), 760–770. <https://doi.org/10.1016/j.jtbi.2010.03.024>
- Zhang, J., & Wu, L. (2013). Allometry and dissipation of ecological flow networks. *PLoS ONE*, 8(9), 1–8. <https://doi.org/10.1371/journal.pone.0072525>

Identifying and Comparing Elementary Students' Problem-Solving Behavior Patterns Using Lag Sequential Analysis

Bo YANG^a, Baoping LI^{ab*}

^a*Advanced Innovation Center For Future Education, Beijing Normal University, China*

^b*Faculty of Education, Beijing Normal University, China*

*libp@bnu.edu.cn

Abstract: This paper investigates the elementary students' problem-solving behavior patterns using the lag sequential analysis. 90 students in grade 5 were required to develop their own solution to the task in an online assessment system and all their interactive behaviors were recorded and then coded for sequential analysis. Comparing the analysis results between the higher-score and lower-score students shows that regardless of the prior knowledge and school exam performance, their differences in the problem-solving behavior pattern led to the discrepancy in task performance.

Keywords: problem solving, assessment task, behavioral pattern, lag sequential analysis

1. Introduction

Problem solving refers to the process of discovering a proper method of reaching a goal from an initial state. While educated adults may show equally good performance in their skills to solve problems, teenager students normally have different competence levels in using strategies to solve problems (Findings, 2014).

In recent years, some scholars have concentrated on assessing problem solving. Distlehorst et al. (2005) assessed students' problem solving by checking their performance in information acquisition, self-regulation and collaborative study in problem-based learning (PBL). Johnson et al. (2007) argued how students value the information could be evaluated through their information-accessing behaviors. Schweizer et al. (2013) developed MicroDYN to evaluate students' determining variable dependencies through manipulating the variables and observing the effects.

Scholars also investigated different factors influencing problem-solving performance. Kalyuga et al. (2010) and Greiff et al. (2015) argued that cognitive elements, such as exploring the problem, representing knowledge, and planning and evaluating the solution affect problem-solving skills. Sabourin et al. (2012) discovered that information gathering could improve problem-solving efficiency. OECD (2012) found that non-cognitive factors such as belief and motivation have a direct impact on the problem-solving process in the 2012 PISA results. Gyöngyvér et al. (2018) concluded that deductive and inductive ability and fluid intelligence level are significantly associated with problem solving.

In recent years, both hardware and software are competent to collect real-time behavioral data and scholars employed lag sequential analysis (Sackett & Richard, 1979; Bakeman & Gottman, 1986) to analyze the data. Lan et al. (2012) used lag sequential analysis to identify the behavior pattern of students' knowledge construction. Hou (2013) analyzed the differences in behavior between students in an online role-playing game. Yang et al. (2015) investigated the behavioral pattern and group interactive network in online English-to-Chinese translation. Malmberg et al. (2017) examined temporal sequences of regulated learning activities by lag sequential analysis of video data. Shukor et al. (2014) explored students' knowledge construction behaviors using lag sequential analysis.

In this paper, we aim to find and interpret the difference in students' behavioral pattern during the problem-solving process by using lag sequential analysis.

2. Assessment System

A general framework for the assessment task is provided in the assessment system and all the interactions between the respondent and system will be automatically recorded into log files for further analysis.

The system offers some general functionalities, as Figure 1 illustrates. For instance, a tip button is to the upper left of the task window and a small popup window showing the guideline and tips about this task will be displayed after clicking it. In addition, the restart and give up buttons are to the lower left and right of the task window, respectively. The restart button can be used to restart the task from the beginning and the give up button can be used to abort the task with zero score obtained.



Figure 1. General Functionalities of the Assessment System

We carried out our experiment and data analysis by using one of our assessment tasks where tent allocation in outdoor camping is used as the background setting. Respondents need to allocate tents to people by dragging women, girls, men or boys into the big, medium and small tents. Different tents can accommodate different numbers of people and two essential requirements for completing this task can only be found in the tip popup. The solution to the task can be submitted only when the following two requirements are satisfied: only people of the same gender can use the same tent and each tent must have at least one adult in it. Besides, the dragging action in this task can never be undone and the only way of clearing a false move is to click the restart button and do the task again.

3. Experiment Design and Analysis Method

In this experiment, respondents would independently finish the task within 10 minutes. In total, 90 elementary students in grade five participated.

We used lag sequential analysis to analyze the students' behavior sequences in two scale ranges: first all the behaviors in the whole problem-solving process were used and then only their dragging choices of people and tents in each new attempt after a restart were manipulated.

For the whole problem-solving process, all important behaviors including clicking the tip, dragging people into the tent, clicking the restart button, reading reference information, using the give up button and finally submitting the solution were used. Of these behaviors, clicking the tip is a little bit special since just clicking the button could never guarantee students' reading and understanding the information. As the assessment system would also automatically record the duration of the behavior, only those lasted for over 6 seconds when reading the tip of 30 Chinese characters would be recognized as effective reading and then used.

For the allocating choice, every first dragging behavior both when the task began and each time after they restarted the task by clicking the restart button was used and dragging different people (man, woman, boy and girl) into different tents (big, medium and small ones) was treated as different behaviors. Since the solution to the task will be submitted and graded only when the two essential requirements are satisfied and no false move can be undone in the task, the restart button may be

frequently clicked by students and whether and how students would change their dragging strategies in the next new turn based on their previous errors can show their problem-solving strategy.

The above-mentioned behaviors were obtained and directly coded for subsequent analysis, as Table 1 and 2 shows, respectively.

Table 1

The Coding Scheme for All the Behaviors in the Whole Problem-solving Process

Code	Behavior	Description
TP1	Read the tip	Read the tip at the beginning of the task
TP2	Click the tip	Read the tip during the problem-solving process
DR	Drag people	Drag people into the tents in allocation of tents
RE	Click restart	Click the restart button to clear and reload the current task
DA	Use reference	Click the “Information Center” to read reference
GU1	Click give up	Click the “Give up” button and the popup menu appears
GU2	Confirm give up	Confirm aborting the task with zero score obtained
CT	Click continue	Click the “Continue” button and return back to the task
SU	Click submit	Click the submit button to submit the solution

Table 2

The Coding Scheme for the First Dragging Behavior in Each New Attempt

Code	Behavior	Description
WB	Woman-big tent	Drag a woman into the big tent
WM	Woman-medium tent	Drag a woman into the medium tent
WS	Woman-small tent	Drag a woman into the small tent
GB	Girl-big tent	Drag a girl into the big tent
GM	Girl-medium tent	Drag a girl into the medium tent
GS	Girl-small tent	Drag a girl into the small tent
MB	Man-big tent	Drag a man into the big tent
MM	Man-medium tent	Drag a man into the medium tent
MS	Man-small tent	Drag a man into the small tent
BB	Boy-big tent	Drag a boy into the big tent
BM	Boy-medium tent	Drag a boy into the medium tent
BS	Boy-small tent	Drag a boy into the small tent

4. Results

4.1 Correlation Results

During our experiment, we also collected the Chinese and Math scores of the students in the last final exam. Then Pearson correlation coefficients were calculated and the results showed that different exam scores did not lead to significant differences in task performance.

Table 3

Correlation Between the Exam Scores and Task Performance

	Chinese score	Math score	
Pearson correlation	0.13	0.05	Task performance
P-value	0.2827	0.6465	Task performance

4.2 Results of the Lag Sequential Analysis

Since our assessment task requires no prior knowledge and students' performance was also independent of their school scores, we assumed that the possible determining factor might be their problem-solving behaviors and strategies during the task process.

The full mark of the task is 100 and the average score is 56.3. In all the 85 students from whom valid behavioral data was obtained, 49 obtained higher scores than the average while 36 got lower scores. As a result, lag sequential analysis was conducted for higher-score and lower-score groups, respectively and the results are as follows.

4.2.1 First Analysis Using All the Task Behaviors

The behavioral transition of the significant sequences (with z-scores greater than 1.96) is illustrated for the two groups in Figure 2 and 3, respectively.

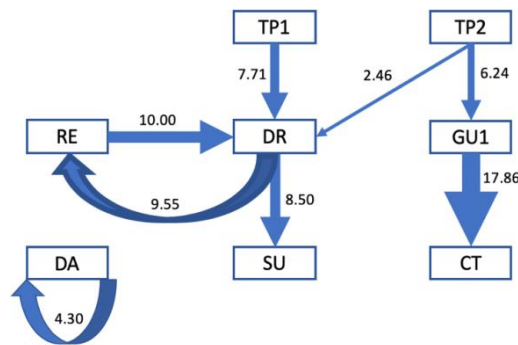


Figure 2. Behavioral Transition Diagram of All the Problem-Solving Behaviors for the Higher-Score Group

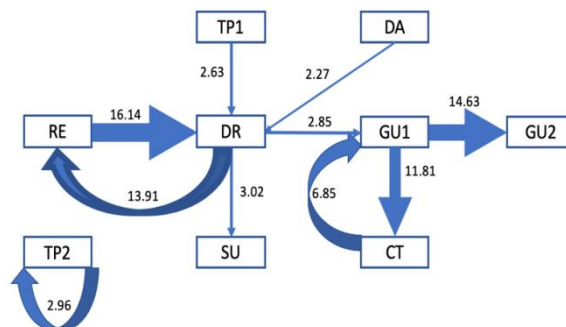


Figure 3. Behavioral Transition Diagram of All the Problem-Solving Behaviors for the Lower-Score Group

According to the above figures, the higher-score and lower-score groups share some significant behavior sequences such as $GU1 \rightarrow CT$, $DR \rightarrow RE$, $DR \rightarrow SU$, $RE \rightarrow DR$, and $TP1 \rightarrow DR$. However, $TP2 \rightarrow GU1$, $DA \rightarrow DA$ and $TP2 \rightarrow DR$ are only significant for the higher-score group while $GU1 \rightarrow GU2$, $CT \rightarrow GU1$, $TP2 \rightarrow TP2$, $DR \rightarrow GU1$ and $DA \rightarrow DR$ are only significant for the lower-score group.

4.2.2 Second Analysis Using First Dragging Behaviors

The behavioral transition of the significant sequences is illustrated in Figure 4 and 5.

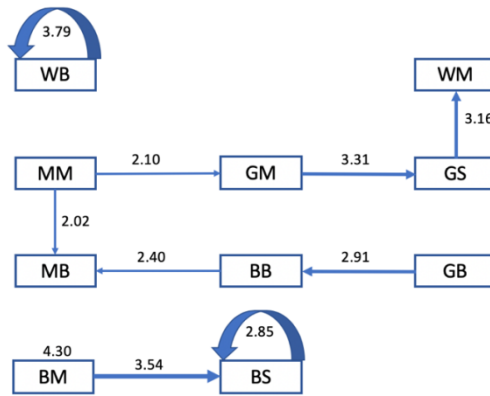


Figure 4. Behavioral Transition Diagram of the First Dragging Behaviors for the Higher-Score Group

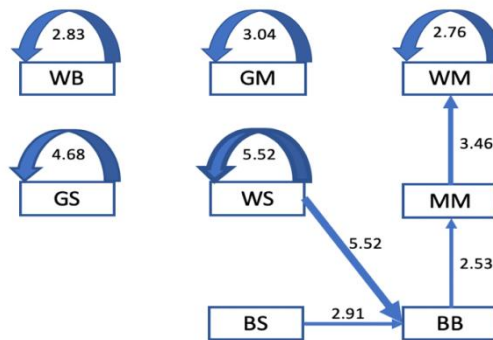


Figure 5. Behavioral Transition Diagram of the First Dragging Behaviors for the Lower-Score Group

According to the results, the two groups share a significant behavior sequence $WB \rightarrow WB$. However, other significant sequences are quite different between the two groups.

5. Discussion and Conclusion

We could draw several implications from the results:

First, in the second sequential analysis, it is obvious that for the higher-score students, various combinations of people and tents were used at the new attempts with only two repeated sequences while their lower-score peers also did comparatively well but tended to repeat their choices in different new attempts. This difference might help result in their different task performance.

Second, in the first sequential analysis, the $GU1 \rightarrow GU2$ sequence ranked as the second most significant sequence for the lower-score group while for the higher-score group, the $GU1 \rightarrow CT$ sequence appeared as the most significant one and no sequence including $GU2$ showed significance. This indicated that among all the task behaviors, the higher-score students may click the give up button but they tended to return to the task whereas their lower-score peers could really give up. In addition, the $CT \rightarrow GU1$ sequence shown by the lower-score group implied that they could repeat clicking the give up button after they cancelled it whereas no similar sequence initiated by CT for the higher-score students indicated that they tended not to try giving up again after a previous cancelation.

Moreover, in the first sequential analysis, both groups had the $TP1 \rightarrow DR$ sequence that indicated they all read the tip at the task beginning and then began dragging. However, the $TP2 \rightarrow DR$ for the higher-score group suggested that they also turned to the tip during the task and then started dragging again. Since the information of the tip is essential to obtaining the proper solution, naturally we can assume that the helpful tip might be used to instruct their dragging behaviors. On the contrary, the $TP2 \rightarrow TP2$ sequence appeared significant in the lower-score group and it indicated these students would just repeat reading the tip but without the $TP2 \rightarrow DR$ sequence, the tip might not make a difference to reaching their task goals. It seemed that the higher-score students were more target-oriented and their lower-score peers distracted by the tip information to some extent.

All the above differences are related the behavioral strategies representing students' positive character and attitude. This is supported by studies conducted by other scholars. O'Connell (2000), Jonassen (2000), Au et al. (2003), and Erdemir (2009) argued about the importance of positive character and attitude in problem-solving ability from different aspects in their individual research.

Acknowledgements

This work was supported by Open Funding Project of the Key Laboratory of Modern Teaching Technology, MOE of PRC (Grant No. SYSK201802).

References

- Au, S., & Parameswaran, N.. (2003). Problem Solving with Attitude and the Theory of Reasoned Action. *Problem solving with attitude and the theory of reasoned action*.
- Bakeman, R., & Gottman, J. (1986). *Observing Interaction*. New York: Cambridge University Press.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis (2nd ed.)*. Cambridge: Cambridge University Press.
- Distlehorst, L. H., Dawson, E., Robbs, R. S., & Barrows, H. S. (2005). Problem-based learning outcomes: the glass halffull. *Academic Medicine*, 80(3), 294-299.
- Erdemir, N.. (2009). Determining students' attitude towards physics through problem-solving strategy. *Asia-Pacific Forum on Science Learning and Teaching*, 10(2), 21.
- Findings, K. (2014). *PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-life Problems (Volume V)*. OECD.
- Greiff, S., Wüstenberg, S., & Avvisati, F. Computer-generated log-file analyses as a window into students' minds? a showcase study based on the PISA 2012 assessment of problem solving[J]. *Computers & Education*, 2015,91: 92-105.
- Gyöngyvér, Molnár, Greiff S, Fischer A, et al. Empirical study of computer-based assessment of domain-general complex problem-solving skills[M]. *The Nature of Problem Solving*. 2018.
- Hou, H. T., & Wu, S. Y. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: a case study. *Computers and Education*, 57(2), 1459-1468.
- Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. *Journal of Experimental Psychology Learning Memory & Cognition*, 33(3), 461-474.
- Jonassen, D. H.. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63-85.
- Kalyuga,S., Renkl, A., Paas, F., Facilitating flexible problem solving: a cognitive load perspective[J]. *Educational Psychology Review*, 2010(22):175-186.
- Lan, Y. F., Tsai, P. W., Yang, S. H., & Hung, C. L. (2012). Comparing the social knowledge construction behavioral patterns of problem-based online asynchronous discussion in E/M-learning environments. *Computers & Education*, 59(4), 1122-1135.
- Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, 49, 160-174.
- O'Connell, S. (2000). Introduction to problem solving. *Strategies for the elementary math classroom*. Portsmouth, N.H: Heinemann.
- OECD. (2012). *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-life Problems*. Paris: OECD Publishing. 2014
- Sabourin J, Rowe J, Mott. B. W., et al. Exploring inquiry-based problem-solving strategies in game-based learning environments[C]. *International Conference on Intelligent Tutoring Systems*. Springer-Verlag, 2012:470-475.
- Sackett, G. P., & Richard Holm (1979). A fortran program for lag sequential analysis of contingency and cyclicity in behavioral interaction data. *Behavior Research Methods and Instrumentation*, 11(3), 366-378.
- Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: complex problem solving predicts school grades beyond working memory capacity. *Learning & Individual Differences*, 24(2), 42-52.
- Shukor, N. B. A., Tasir, Z., Meijden, H. A. T. V. D.,& Harun, J. (2014). Exploring students' knowledge construction strategies in computer-supported collaborative learning discussions using sequential analysis. *Journal of Educational Technology & Society*, 17(4), 216-228.
- Yang, X., Li, J., Guo, X., & Li, X. (2015). Group interactive network and behavioral patterns in online English-to-Chinese cooperative translation activity. *Internet and Higher Education*, 25, 28-36.

A Tool for Learning of Cognitive Process by Analysis from Exemplar Documents

Wasan Na Chai, Taneth Ruangrajitpakorn, and Thepchai Supnithi

Language and Semantic Techonolgy Laboratory,

National Electronics and Computer Technology Center, Pathumthanee, Thailand

{wasan.na_chai,taneth.rua,thechai.sup}@nectec.or.th

Abstract: Metacognition is difficult to train, and most of students lack the knowhow to practice it. This work proposes a thought analysis tool from reading to help training in understand logical connections between sentences. The tool allows learners to explicate types of logical expression of the writing article as the example of strategy to convince readers. By analysis and annotating the text in a controlled environment, it is expected that learners can learn from thinking about author's cognitive process, and apply them to think about their own thought and make a strategic planning when they are in a role to write. The tool is designed for learners to assign several kinds of annotations including logical statement type, keyword and logical linking between sentence to the proper writing text by self-analysis under the supervision of coaches. The experiment results show that the tool helps to apparently improve learners' cognitive performance in composing proper essay via the training on analysis reading. In comparison to other sample groups with lecturing and coaching without the tool, the improvement of the tool users was significantly greater in term of logical relatedness, logical completeness and convincible power.

Keywords: Metacognition, Concept annotation tool, Explicating thought, Logic learning

1. Introduction

Metacognition is a regulatory system involving in understanding and controlling ones' own cognitive performance. Metacognition literally refers to a process of 'thinking about thinking' or 'an ability to know what one knows and what one does not know' (Mahdavi, 2014). Metacognition nurtures ones' ability to think and learn via awareness of how they learn, an evaluation of their learning needs, generating strategies to meet these needs and then implementing the strategies. Developing metacognition is undoubtedly important for learning how to learn and to make use of knowledge/skill. In development of metacognition, metacognition mostly develops with practice (Papaleontiou-Louca, 2003). Harris (2003) pointed out that key elements in developing metacognition including to practice having strategies for planning, monitoring, and evaluating of use.

In practicing metacognition, self-regulation approaches are suggested by letting learners to specify their strategies for planning, monitoring and evaluating their learning. The key is to ask learners to make their thoughts explicit since their vague conceptual thought will be more collected. Besides, making thoughts explicit can cause an intervention of running thought and making learners aware/reflect of their cognition process. There are several implemented tools and teaching strategies elaborating the idea including concept-mapping tool (Chevon, 2014), annotation tools (PJ Rich, M Hannafin, 2009) and wrapper modules to control learning environment. The studies on these works show the sign of development on metacognitive skills.

Metacognition can explicitly be noticed in ones' language usage. The writing essays and articles show author's thought of strategies to convince readers. Furthermore, the use of logical statements is a good manifestation of metacognitive skills since it shows how well learners grapple with and apply their knowledge and skill. This work aims to study the use of control environment by the metacognition training tool in practicing of analysis on logical statement from reading articles. The tool allows users to concept-mapping and annotate logical relations in context as a good example for practicing on using their metacognition. The study in this work is conducted on the effect of learning by

analyzing from good examples and their improvement in using the learned skill. The rest of the paper is organized as follows. Section 2 explains a design of the thought analysis tool via reading. Section 3 provides experimental setting, result and discussion. Last, Section 4 gives a summary of the paper and a future plan of the research.

2. Tool Design and Usage

This section describes on the design of the tool for controlling environment for practicing metacognition. The thought analysis tool via reading is extended of the tool (W. Na Chai, 2017) for analysis of written articles in terms of logical statement, relation among context and the key concept of the writing. The main focus is for users to analyze the hidden logical relation of the sentence and the method how the author uses to convince readers. Namely, it tasks user to think about what author thinks. By doing so, when the reader turns to be a writer, they are expected to think and to be aware about their cognition in writing process.

The tool provides functions including annotating a keyword (key concept) of a statement, a type of a statements, and indicating how arguments relating to other. With the annotation, hidden relation of logical links of content is explicitly revealed. An overview of the thought analysis tool via reading is show in Figure 1.

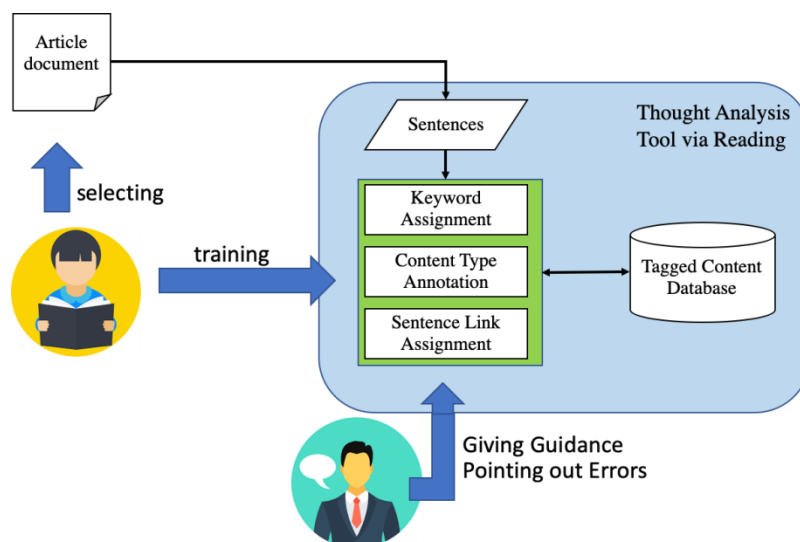


Figure 1. An Overview of the Thought Analysis Tool via Reading

The input of the tool is a text (currently allows Thai and English only) from a writing article. The tool asks users to split their text into a sentence level. The tool then asks users to annotate the sentence regarding logical concepts hidden in the context. The annotation includes keyword (key idea) of the sentence, statement logical type and its link to other sentences. This should reveal the logical strategies of the article author to convince readers towards their goal. Then, the annotation will be reviewed by the coach for guiding towards correction of the annotation. The coach in this tool plays a role of validators of annotation, namely pointing out the incorrect annotation but not giving the correct annotation. With supervision of coach, the learner users will be prevented from misunderstanding and false-belief from shallow thinking and unknowingness.

2.1 Keyword Assignment

For each sentence, learner users are asked to provide keyword(s) which they think capable to represent the core concept of the sentence. In assigning keyword(s), users are allowed to fill the box with the term appeared in the sentence context. The number of keywords available for assigning is limited to 1 to 3 terms per sentence. The chosen term can be either a single word or compound words.

This function is designed for users to analyze the terms in the sentence for the most significant representation to the concepts of the sentence. This aims to let users realize importance of term selection regarding contextual appropriateness.

2.2 Logical Type of Statement Selection

In this function, a list of logical type is designed to be annotated to the sentence. With exhaustive review on guidelines for writing, many types are given and used in the previous version of the tool (W. Na Chai, 2017). However, the users submit a complain that there are too many types to select on, and some overlaps to other in which causes a difficulty and confusion in selection. In this work, the list is reviewed and revised for the main concepts of logical statement in a two-level hierarchical structure as exemplified in Table 1.

Table 1. A List of Statement Types for Annotation and Their Definition

Type		Definition
Top	Bottom	
General	<ul style="list-style-type: none"> - Declaration of fact - Declaration of opinion 	<ul style="list-style-type: none"> - Statements that relay fact information - Statements that relay opinion information
Addition	<ul style="list-style-type: none"> - Giving details - Emphasizing 	<ul style="list-style-type: none"> - Statements that provide specific details of core term of other statements - Statements that are restated again to give a sense of significance
Opposition	<ul style="list-style-type: none"> - Contradict part 	<ul style="list-style-type: none"> - Statements that relay a contradiction of other statements
Illustration	<ul style="list-style-type: none"> - Giving example - Giving demonstration 	<ul style="list-style-type: none"> - Statements of the actual case/instance of the circumstance in other statements - Statements to demonstrate the circumstance in other statements
Cause-effect	<ul style="list-style-type: none"> - Cause part - Effect part 	<ul style="list-style-type: none"> - Statements about the cause of the event/incident - Statements about the effect/result of the caused event/incident
Sequencing	<ul style="list-style-type: none"> - Prior part - Following part 	<ul style="list-style-type: none"> - Statements that occur in prior of a continuous event/incident - Statements that occur in following of a continuous event/incident
Condition	<ul style="list-style-type: none"> - Condition part - Result part 	<ul style="list-style-type: none"> - Statements about the condition that triggers the event/incident - Statements about the result that is triggered by completing the condition

Users are asked to assign one of these types to all sentences. Only the types of the bottom are allowed to be annotated. With the types, users are expected to visualize author's implicit method and strategy in convincing readers from the connection of sentence. Users will also learn a pattern of a writing style and look deeper into logic instead of the surface of the content.

2.3 Sentence Linking

Relation of sentences indicates the link of thoughts from author's ideas in their work. From observing the proper published articles from genuine sources, we find that most of the sentences are related to form a logical network to convince reader. Thus, we expect reader to learn the relational expression from the good writing.

According to statement types, there is a link between types such as cause and effect. Learners are asked to align the relation to connect the sentence based on their analysis. The assignment of the logical relation is to fill the sentence ID of other sentence(s) regardless of assigned statement types. In case of a new concept unrelated to any previous concepts, users are allowed to assign 'none' to the sentence.

2.4 Providing a Summarized Goal of the Article

After all the annotations are made, this function is to ask for a summary concept of the article. Users are asked to give a short summary representing a goal of the reading article. The fill-in box allows up to 250 syllables. This part is important to get the understanding of readers' thought whether they can detect the core of the content after all the annotations are made. This information can help to realize users thought and view about the aim of author which can significantly help learners to concretely establish their own aim when they need to write.

2.5 Feedback from Coach

Once learner users submitted their annotations, they are stored into the database. Coaches in this tool have a role to monitor the submitted annotation and to validate them carefully. The coach is asked to provide feedback on the annotations. The options for the feedback 'agree' and 'disagree' while the 'disagree' has the option to provide additional comment on how the coach disagrees on the annotation. Particularly, the coaches are not suggested to give the correct answer directly, but the guidance on how to think to reach that conclusion. This will help to instill the correct cognitive process and the chance for learners to think about their way of thinking.

3. Experiments

3.1 Experiment Design and Setting

For an overview of this experiment, we aimed to study the effect of the tool in controlling environment for learning of logical analysis. The experiments were to find improvement of learners' metacognition towards the understanding and applicability of convincing writing regarding logical expression usage. The convincing writing was regarded as a realization of goal, strategy to achieve the goal and the performance of the chosen strategy. The learning was conducted by studying from a good example in which in this case was reading the good articles. The testing was measured in the writing from the written essay. A flow of experiment is drawn in Figure 2.

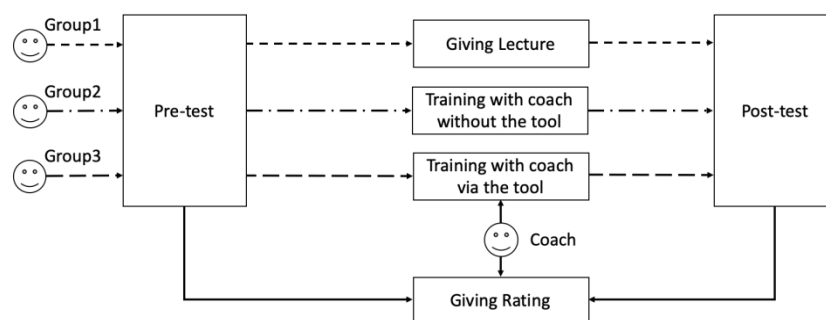


Figure 2. Experiment processes

There were two roles of participants in this experiment. The first is learners, and the second is coaches. Learners were a university Thai student who was a tested subject while coaches took a role of validator who pointed out mistakes from learners and provided guidance in the process. Learner participants were a university undergraduate student. The learner participants were randomly separated into three groups of 15 participants each from a total of 45 learner participants. The three groups were assigned with a different learning process as follows.

- Group1: providing a lecture on the logical analysis from coaches
- Group2: training on analysis with coaches
- Group3: training on analysis via the tool with coaches

For all groups, pre-test and post-test were conducted by asking participants to write an essay of their choice of languages between Thai and English for 400-600 words. The topic of the essay could be chosen from the provided topic-list or by themselves. The topics from the list were a common but meaningful topic such as 'importance of learning English' or 'Definition of superhero in your opinion'. On the other hands, the participant's own topic required an approval from a coach. After the pre-test,

Group2 and Group3 had been trained by the coach by choosing and learning from an academic article as an example for 7 days. Meanwhile, participants from Group1 were given a lecture regarding logical statement in convincing writing. The post-test then was conducted together at the day after the training of Group2 and Group3 were done. As restriction, the topic of the pre-test and post-test for individual could not be the same topic. The essays in blind from both pre-test and post-test were rated in scale of 1 (lowest) to 5 (highest) by three coaches for the aspects given in Table 2. The difference of the rating was a measurement of the improvement.

Table 2. *Focused Aspects for Rating and Definition*

<i>Aspects</i>	<i>Definition and Rating Criteria</i>
<i>relatedness</i>	<i>Checking connection between sentences; the more quantity and quality sentences are correctly logically linked in proper context, the higher rating is given.</i>
<i>keywording</i>	<i>Checking for the using of key terms in sentences; the more proper choose and use, the higher rating is given</i>
<i>completeness</i>	<i>Checking for a completion of statement, i.e. the use of both 'cause and effect' or both 'condition and result'; rating deduction from missing the pair of relation.</i>
<i>overall convincible</i>	<i>Checking the overall convincible of the written essay regarding giving sufficient evidence and reasoning for the statement.</i>

3.2 *Experimental Results and Discussion*

From the experiment, the essays from participants were all written in Thai. The average rating results from three coaches of pre-test and post-test were calculated to find the difference. The higher difference of the post-test than the pre-test indicated the improvement of the learner and vice-versa. The results are given in Table 3.

Table 3. *Improvement Result Separated by Groups and Focused Aspects*

Group	relatedness	keywording	completeness	overall convincible	Summation
1	0.09	0.09	-0.11	0.11	0.18
2	0.11	0.78	0.18	0.27	1.33
3	1.27	0.80	1.00	0.96	4.02

The results from Table 3 indicates that the group using the tool (Group3) improves the best for all focused aspects. Group2, who were trained by coaches without the tool, also has an improvement, especially on the keywording aspect. For Group1, there is a little improvement while has a minus score in a completeness aspect. In overall results, Group3 obtains the highest improvement while Group2 comes in second.

These results signify the potential of using the tool to increase users understanding and applicability of their knowledge and skill from practicing via the tool. It is also notable that the students can improve one's own thinking process by training from analyzing another people thought. They become more realizing of their thinking process and are able to make use of them. By observing the results of the Group3, their post-test essay was full of the complete logical expression and appropriate terms. They were more convincible when comparing with their pre-test. For those students from Group2 who were also trained closely by coaches, their improvement is greatly shown in the selection of key terms in the context, but a few improvements in other aspects. Without the tool to visualize the implicit logical relatedness among sentences, the training alone cannot make learners to conceptualize the logical relation in their cognition. Despite being told, lacking practicing on these aspects results in little to none improvement.

When looking into results based on individuals' base ability, we obtain the rating improvement results given in Table 4. The ratings for the learners in Group2 and Group3 with low pre-test rating were notably increased; however, the increase was not much for those with pre-test rating higher than 10. This indicates that the training, especially with the tool, works well to those who lacked analysis skill or did not know how to analyze logical relations among sentence. On the other hands, the learners who performed well in the pre-test gain a little since they have the skill. In details, the learners with good

pre-test rating results mostly increase the outcome for keywording and overall-convincible, respectively.

Table 4. *Improvement Result based on Pre-test Rating*

Group	Rating score by pre-test				
	≤6	>6, ≤8	>8, ≤10	>10, ≤12	>12
1	1.00	1.50	-0.17	0.33	-0.89
2	2.17	1.60	1.22	0.67	0.33
3	5.67	5.22	4.73	1.67	0.78

After the experiments, learner participants were asked for opinions regarding the learning. Participants from Group2 and Group3 gave remarks that ‘they never know that reading articles could be used to learn for how to write’. They also mentioned that they can visualize the logical relations of sentences, and this results in understanding author’s aim of the article as well as context comprehensibility. With this understanding, they also are motivated to read more articles since it is easier to obtain knowledge than prior. For those participants from Group1, most of them revealed that they did not fully understand the thinking process of the author as well as their own cognitive process while reading and writing. They were clueless on what to be focused when reading and writing, and let their mind astray from objectives.

4. Conclusion and Future Work

This paper proposes a thought analysis tool via reading. The tool, despite involving in reading, does not aim to support language learning or improve reading skill, but it is designed to help learners to draw out the thought of the article author by mapping the hidden logical relation of sentences in the writing pieces. Concept mapping by annotating type of statement to the sentences assists to reveal main concepts and a connection between sentences. By making the logical relation explicit, learners are trained and learned to think about what author do to convince readers as a strategy and the pattern used to express the statements. This is expected that by learning to think about other’s thinking learners are affected to improve their own cognitive process as to think about their thinking when they hold a role of writer.

From the experiment, the results signified that the thought analysis tool via reading improved users in terms of becoming more strategic planner and making a convincible writing piece. In comparison to other methods including lecturing and training without the tool, the improvement of the tool user group was clearly higher. Moreover, the results of improvement based on pre-test rating indicated that the tool works best for the groups of learners with the low cognitive skill proficiency while it showed less effectiveness to those with high pre-test rating.

To improve our work, we plan to apply the tool to train a thinking process on other subjects such as art designing, movie script writing, translating and researching. We also plan to visualize the annotated data of connected logical expression into a graph-based representation and study the found patterns. Furthermore, more experiments to different age-groups and knowledge background will be conducted for studying the effect of the thinking training and their improvement in metacognition.

References

- M. Mahdavi. (2014). An Overview: Metacognition in Education. *International Journal of Multidisciplinary and Current Research*, Vol.2, Issue May/June, 2014, Pages 529-535.
- E. Papeontiou-louca. (2009). The concept and instruction of metacognition. *An international journal of teachers' professional development* ,Volume 7, 2003 - Issue 1
- K. R. Harris, S. Graham, M. Brindle, K. Sandmel. (2009). Metacognition and children's writing. *Hand Book of Metacognition in Education*.
- MP Chevron. (2014). A metacognitive tool: Theoretical and operational analysis of skills exercised in structured concept maps. *Perspectives in Science* Volume 2, Issues 1–4, 2014, Pages 46-54.
- PJ Rich, M. Hannafin. (2009). Video Annotation Tools: Technologies to Scaffold, Structure, and Transform Teacher Reflection. *Journal of Teacher Education*, Volume: 60 issue: 1, page(s): 52-67.
- W. Na Chai, T. Ruangrajitpakorn, T. Supnithi. (2017). A Tool for Data Acquisition of Thinking Process through Writing. *International Conference on Computers in Education*.

E-book Learner Behaviors Difference under two Meaningful Learning Support Environments

Jingyun WANG^{a*}, Atsushi SHIMADA^b & Fumiya OKUBO^c

^a*Research Institute for Information Technology, Kyushu University A, Japan*

^b*Faculty of Information Science and Electrical Engineering, Kyushu University, Japan*

^c*Department of business administration, Takachiho University, Japan*

*warmplam@gmail.com

Abstract: In this paper, we present an ontology-based visualization support system for e-book learners, which provides not only a meaningful receptive learning environment but also a meaningful discovery learning environment. Those two environments are developed to help e-book learners to effectively construct their knowledge frameworks. A series of experiments were conducted on four undergraduate classes instructed by two professors (A and B): two classes (one guided by A and the other guided by B) were assigned as control groups and studied with one e-book chapter in receptive learning environment while another two classes (one guided by A and the other guided by B) were assigned as experimental groups and studied with the same e-book chapter in discovery learning environment. For analyzing the learner behavior, K-means clustering algorithm is performed not only by considering the number of total command actions and the cumulative duration of stay on target pages as learner features, but also by considering the duration of stay on each target page (in total 15 pages) as learner features. Learners' behavior differences in e-book system are examined and discussed.

Keywords: Clustering, map, meaningful learning, discovery learning, reception learning

1. Introduction

Nowadays, many countries, especially Japan and Korea, has started to use digital textbook instead of traditional textbook due to its various advantages such as cost-saving and higher portability (Shepperd, Grace, & Koch, 2008; Shin, 2012; Yin, et al, 2018). Since the learning behavior of e-book users are recorded in logs and can be readily accessed for further analysis, when evaluating the learning effect of a learning support system, the examination of e-book logs is essential.

In the previous work (Wang, Ogata, and Yamada, 2017), an ontology-based visualization support system, is designed and developed to help e-book learners to effectively construct their knowledge frameworks. Two learning modes are provided in this system: (a) reception comparison mode, in which learners are provided directly with complete versions of relation maps; and (b) cache-cache comparison mode, where all information concerning relations is hidden at the first stage of learning, and in the second stage learners are encouraged to actively create them. In this paper, we will exploit traces of learner activities and discuss the behavior differences of e-book learners while they studied with these two modes.

2. A Visualization Support System for E-book Learners

Facilitation of the visualization support of meaningful learning (structure (Ausubel, 1963; Ausubel et al., 1978).) requires descriptions of the information about all the knowledge points and their relations. In this study, a Knowledge Point (KP) is defined as "a minimum learning item which can independently describe the information constituting one given piece of knowledge in a specific course." The learner can understand a KP via its own expression or can acquire it through practice. It is suggested that the domain knowledge needed by the learning support system should be automatically extracted from an ontology designed and developed on the basis of the content of the e-books.

To construct a demonstration unit, first we adjust the ontology design method described by Wang et al. (2014) and apply it to the development of a course-centered ontology for an existing

computer science course (called COCS). The ontology consists of about 100 KPs and 20 kinds of relations, extracted and defined based on an analysis of the content of all the e-books of this computer science course. Moreover, an ontology-based visualization support system (called VSSE), which supports not only meaningful receptive learning but also meaningful discovery learning, is implemented to help e-book learners to efficiently develop their conceptual framework (Wang, Ogata, and Yamada, 2017).

2.1 Meaningful Reception Learning Environment

One of the main functions of VSSE is that the KPs appearing in any page range of any e-book, along with their upper concepts, can be displayed in a relation map. In the reception comparison mode, learners are provided directly with complete versions of relation maps (Wang, Ogata, and Yamada, 2017). As can be seen in Fig. 1, users of the e-book system can select a specific e-book and input any page range in the reception comparison interface. VSSE will display all the KPs appearing in the searched pages along with their related KPs. For example, Fig.1 displays: red nodes, which represent the KPs that appear in pages 1 to 20 of e-book A03; blue nodes, which represent related KPs that do not appear in those pages but have essential relations with the KPs represented by the red nodes; and pink nodes, which represent the upper concepts of the KPs represented by red or blue nodes. When the user places the mouse on any node in this relation map, the essential properties (such as definition and explanation, represented by the data properties of one individual in COCS) of that KP will be listed, while for every arc in the relation map, a statement of the relation will be displayed (for example, the displayed relation axiom “prescribe” from “ASCII” to “ASCII Character Set” in Fig. 1). Therefore, users can conveniently find the essential properties of every KP and all its related KPs from this visualization map. All that information is extracted automatically from COCS.

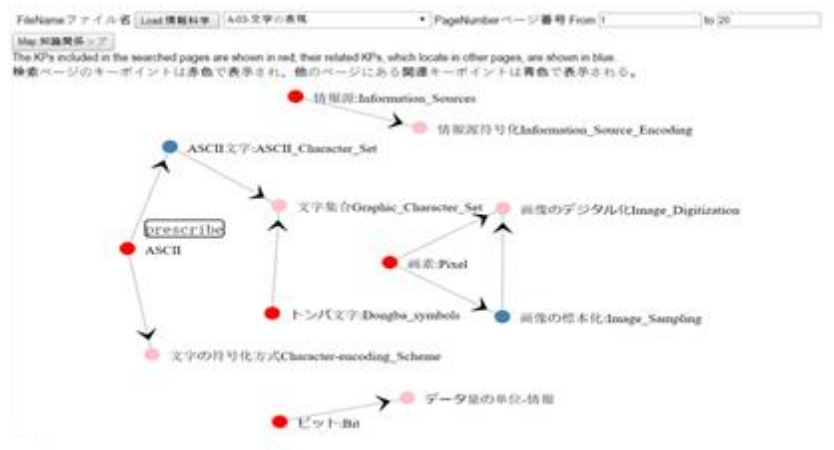


Figure 1. The relation map of the search pages

2.2 Meaningful Discovery Learning Environment

To encourage active engagement in meaningful learning, another environment (called “cache-cache comparison” mode) integration with discovery learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Bruner, 1961) is also presented (Wang, Ogata, and Yamada, 2017). Considering KPs and relations as the building blocks of course relation maps, “cache-cache comparison” mode in VSSE hides all the relations in expert relation maps and guides learners to seek to discover those hidden relations. The learners engage in an active learning process when they struggle to complete relation maps. Fig. 2 shows an instance of “cache-cache comparison” mode: the range of interest to the learner is pages 1 to 20 of e-book A-03. First, as shown in Fig. 2, “cache-cache comparison” mode displays all the KPs that appear in the page range of interest in red; the related KPs that do not appear in the pages of interest in ranges in blue; and their upper concepts in pink. Then firstly the learner is required to classify the KPs by connecting them to their pink upper concepts; next, the learner is encouraged to find out the relations between KPs by connecting red nodes or connecting red nodes to blue nodes. The descriptions of the

relation arcs made by the learner can be modified and saved anytime. After the learner completes the relation map, she/he can click the “Compare with experts” button. Finally, all the relations extracted from the ontology will be displayed as red lines. The learner can easily compare the red lines with the black lines that she/he has made.

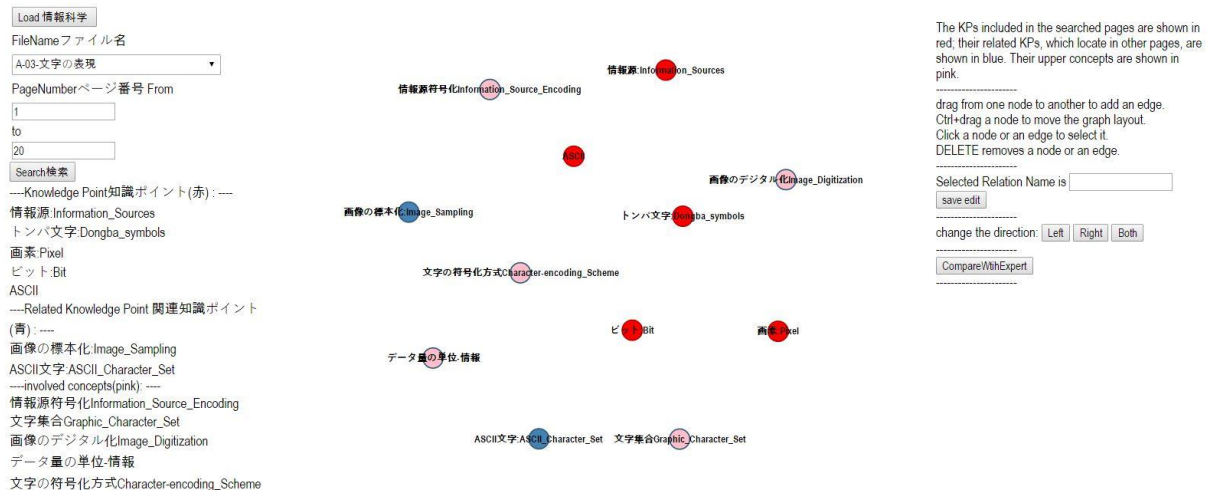


Figure 2. An instance of “cache-cache comparison” mode

3. Participants and Experimental Procedures

To explore the learning performance differences (including e-book behaviors, learning achievement and perception of learning) between participants who studied with the two environments presented in previous section, three hundred and sixty-seven first-year undergraduates from four classes belong to the same computer science curriculum at a Japanese University participated in the experiments of this study. Two of the classes instructed by Professor A, were randomly assigned as experimental group (namely E_A, 72 students) and control group (namely C_A, 103 students), respectively, while the other two classes instructed by Professor B were also randomly assigned as experimental group (namely E_B, 146 students) and control group (namely C_B, 46 students), respectively. Before the experiment, all the participants had studied this computer science course for 14 weeks within learning support systems (Moodle, e-portfolio system Mahara and an E-book system called BookLooper (Yin, et al, 2015)). And weeks before the experiment, all participants have received the instruction related to the e-book c03, entitled, “Digital Image Processing.” During the experiment, the participants took pre-test at the beginning of the class and the control groups reviewed the e-book C03 with reception comparison mode support while the experimental groups reviewed with cache-cache comparison mode support. The first 15 pages of e-book C03 were chosen as target learning content for 30-minuts review activity. In those 15 pages 11 KPs appear. After the study, all of them took another post-test and a questionnaire related to this learning activity.

4. E-book Behavior Analysis using K-Means Clustering Algorithm

Whenever a user interacts with BookLooper, one data log containing user ID, the date, time, learning content ID, page number, user action (such as “OPEN”, “NEXT”, “PREV”, “SEARCH JUMP”, “PAGE JUMP” and so on), and other data, will be recorded in her/his reading history (Yin, et al, 2015). In this paper, we intend to discuss the learner e-book behavior differences under the support of the two different meaningful learning environments mentioned in Section 2. The e-book logs of all the participants who studied the target pages with the support of VSSE are analyzed by means of K-means clustering algorithm. K-means clustering is one of the data mining approaches applied to identify learners with similar features and assign them into clusters. In this paper, basic K-Means clustering by considering K=2 and K=3 was performed, respectively.

4.1 Analysis on the number of total command actions and the cumulative duration of stay

Firstly, we consider the number of total command actions and the cumulative duration of stay (the total amount of time spent on the target 15 pages) as 2 features of learners. These two features are the typical variables used to explain the individual diversity in e-learning behaviors ((Yin, et al, 2018). It is needed to mention that, in the clustering analysis of this paper, when a learner stays more than 10 minutes on one page, that page is considered left unattended after 10 minutes. During a 30-minutes learning activity, spending more than 10 minutes in a page is unusual considering the moderate difficulty of the learning content; therefore, 10 minutes will be considered as the duration of stay. The 2-means clustering result is shown in Table 1. The average cumulative duration of stay of learners in Cluster 0 was significant longer than those in Cluster 1 ($F(1, 193) = 460$, Cluster 0: Mean=787s, SD=277s, Cluster 1: Mean =200s, SD=190s); the average number of total command actions of learners in Cluster0 was significant higher than those in Cluster1 ($F(1, 153) = 272$; Cluster 0: Mean =105, SD=47, Cluster 1: Mean =32, SD=20). Although, lower than 20% of students in both classes guided by Professor A are found in Cluster 0, the percentage of students in Cluster 0 in E_A Class (17%) is still 3 times more than C_A Class (5%). In the other hand, 69% of students in E_B class are identified in cluster 0 while only 22% of students in C_B class are identified in cluster 0.

Table 1

The 2-means clustering result by considering the number of total command actions and the cumulative duration of stay as learner features

Cluster	Label	Average cumulative duration of stay (sec)	Average number of total command actions	N	Percentage inside label	Percentage inside cluster
0	C_A	561	74	5	0.05	0.04
	E_A	813	90	12	0.17	0.09
	C_B	718	80	10	0.22	0.08
	E_B	802	110	101	0.69	0.79
1	C_A	138	30	98	0.95	0.41
	E_A	176	31	60	0.83	0.25
	C_B	302	32	36	0.78	0.15
	E_B	286	40	45	0.31	0.19

The 3-means clustering result is consistent with 2-means clustering result. The average cumulative duration of stay of learners ($F(2, 107) = 395$, Cluster 0: Mean=912s, SD=318s, Cluster 1: Mean =640s, SD=208; Cluster 2: Mean =135s, SD=128s) and the average number of total command actions of learners ($F(2, 112) = 261$; Cluster 0: Mean =149, SD=41, Cluster 1: Mean =65, SD=25; Cluster 2: Mean =30, SD=20) in these 3 Clusters differ significantly (Cluster 0>Cluster 1>Cluster 2;). There are 31% of E_B class student, 4% of E_A students, 4% of C_B students are found in cluster 0 and they had the longest average cumulative duration of stay and highest average number of total command actions; 50% of E_B class student, 39% of E_A student, 26% of E_A students and 9% of C_A students are found in cluster 1 and they have the medium average cumulative duration of stay and medium average number of total command actions; the rest of the students in cluster 2 had the shortest average cumulative duration of stay and lowest average number of total command actions.

The differences between classes guided by difference professors maybe caused by the learning style difference of the professors. But both 2-means and 3-means clustering results suggest that the learners in experimental groups who study with the support of meaningful discovery learning environment had longer cumulative duration of stay and higher number of total command actions than those in control groups who studied with meaningful receptive learning environment.

4.2 Analysis on the duration of stay on each target page

Secondly, to explore the reading behavior pattern of learners, we consider the duration of stay on each target page (the time spent on each target page, in total 15 variables) as 15 features of learners. The 2-means clustering results are shown in Table 2 and Figure 3. For the average cumulative duration of stay on the target 15 pages, learners in Cluster 0 outdistanced than those in Cluster 1 in almost every page except the first page which only displays the title of the chapter (here we consider $p < 0.001$ as significant standard). As shown by Figure 3, compared to learners in Cluster 1, learners in Cluster 0 tends to stay much longer in page 2 ($F(1, 145) = 85.39$, Cluster 0: Mean = 149s, SD=149s, Cluster 1: Mean = 24s, SD=52s), page 3 ($F(1, 153) = 109.58$, Cluster 0: Mean = 215s, SD=169s, Cluster 1: Mean = 51s, SD=72s), page 4 ($F(1, 143) = 21.14$, Cluster 0: Mean = 46s, SD=89s, Cluster 1: Mean = 9s, SD=30s), page 8 ($F(1, 141) = 32.28$, Mean of Cluster 0=70s, SD=114s, Cluster 1: Mean = 11s, SD=34s), page 9 ($F(1, 131) = 15.77$, Cluster 0: Mean = 43s, SD=109s, Cluster 1: Mean = 5s, SD=16s), page 10 ($F(1, 144) = 25.88$, Cluster 1: Mean = 57s, SD=94s, Cluster 1: Mean = 14s, SD=32s), and page 11 ($F(1, 132) = 32.90$, Cluster 0: Mean = 58s, SD=100s, Cluster 1: Mean = 7s, SD=17). After further examining the contents in those 7 pages, it is found that: 2 KPs and 2 relations appear on page 2; 5 KPs and 5 relations (which is one important part of the target map) appear on page 3; 2 KPs appears on page 4; 4 KPs and 3 relations (which is another important part of the target map) appears in the range between page 8 to 11. This means that these 7 pages which explain important KPs and relations are the key part of target learning content. As shown in Table 2, although lower than 20% of students in both classes guided by Professor A are found in Cluster 0, the percentage of students in E_A Class (19%) is still 3 times more than C_A Class (4%). In the other hand, 68% of students in E_B class is found in cluster 0 while only 24% of students in C_B class is found in cluster 0.

The 3-means clustering results shows learners in three clusters (114 learners in cluster 0, 18 in cluster 1, and 235 in cluster 2) differ significantly in page 2 ($F(2, 43) = 48.48$, Cluster 0: Mean = 165s, SD=153s, Cluster 1: Mean = 58s, SD=59s, Cluster 2: Mean = 22s, SD=48s), page 3 ($F(2, 44) = 57.97$, Cluster 0: Mean = 231s, SD=171s, Cluster 1: Mean = 86s, SD=77s, Cluster 1: Mean = 51s, SD=71s), page 4 ($F(2, 41) = 18.22$, Cluster 0: Mean = 29s, SD=38s, Cluster 1: Mean = 151s, SD=199s, Cluster 2: Mean = 8s, SD=26s), page 8 ($F(2, 41) = 15.29$, Mean of Cluster 0=72s, SD=117s, Cluster 1: Mean = 45s, SD=83s, Cluster 2: Mean = 12s, SD=35s), page 9 ($F(2, 42) = 8.69$, Cluster 0: Mean = 49s, SD=115s, Cluster 1: Mean = 8s, SD=20s, Cluster 2: Mean = 4s, SD=15s), page 10 ($F(2, 41) = 34.80$, Cluster 1: Mean = 33s, SD=30s, Cluster 1: Mean = 228s, SD=175s, Cluster 2: Mean = 12s, SD=21s), and page 11 ($F(2, 41) = 19.05$, Cluster 0: Mean = 63s, SD=105s, Cluster 1: Mean = 22s, SD=25, Cluster 2: Mean = 7s, SD=15s). Although students in cluster 1 in average stayed longer time on Page 4, 5 and 10 than those in the other two clusters, there is only a small number of students in cluster 1 (18 in total). Except those 3 pages, the average duration of stay of learners in Cluster 0 outdistanced those in other two cluster 2 in almost every page except the first page which only displays the title of the chapter.

For summary, the results of 3-means clustering and 2-means clustering are almost consistent. Both results suggest that the participants in experimental groups had longer average duration of stay in the pages which explain important KPs and relations than those in Control groups.

Table 2

The 2-means clustering result by considering the duration of stay on each target page as features

Cluster	Label	N	Percentage inside label	Percentage inside cluster
0	C_A	4	0.04	0.03
	E_A	14	0.19	0.11
	C_B	11	0.24	0.09
	E_B	100	0.68	0.78
1	C_A	99	0.96	0.42
	E_A	58	0.81	0.24
	C_B	35	0.76	0.15
	E_B	46	0.32	0.19

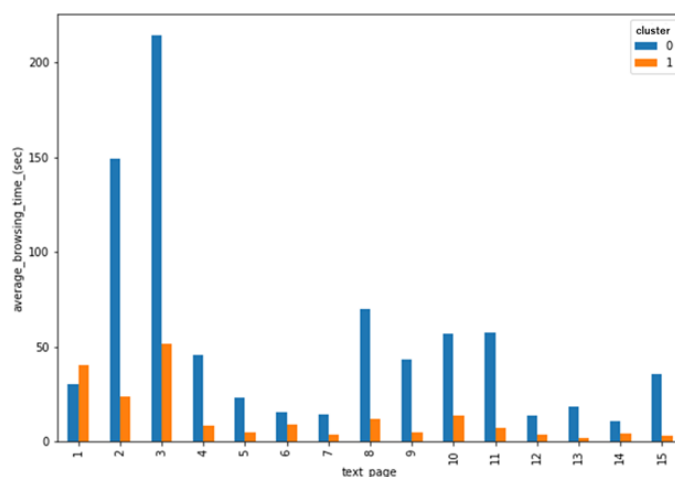


Figure 3. The duration of stay on each target page(2-means)

5. Conclusion and future work

In this paper we present an ontology-based visualization learning support system which provides not only meaningful reception learning environment but also meaningful discovery learning environment for e-book learners. A series of experiments were conducted on an undergraduate computer science course of a Japanese university. To examine the e-book learners' behavior differences under the support of these two different learning environments, the log data of all the participants was analyzed by using K-means clustering approach. The results suggest that compared to the participants who studied with meaningful reception learning environment, those who studied with meaningful discovery learning environments had longer average cumulative duration of stay and higher average number of total command actions; moreover, they also had longer average duration of stay in the pages that explain important KPs and relations. In the future work, learning achievement and learning perception differences will be analyzed and discussed for a better evaluation of these two meaningful learning environments.

Acknowledgements

The research is supported by KAKENHI Grant Number 17K17936.

References

- Alfieri, L, Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. R. (2011). Does Discovery-Based Instruction Enhance Learning? *Journal of Educational Psychology*, 103 (1), 1-18.
- Ausubel, D. (1963). *The Psychology of Meaningful Verbal Learning*. New York: Grune & Stratton.
- Ausubel, D., Novak, J., & Hanesian, H. (1978). *Educational Psychology: A Cognitive View (2nd Ed.)*. New York: Holt, Rinehart & Winston.
- Bruner, J. S. (1961). *The act of discovery*. Harvard educational review.
- Wang, J., Mendori, T., & Xiong, J (2014). A Language Learning Support System using Course-centered Ontology and Its Evaluation," *Computer & Education*, 78, pp 278-293.
- Wang, J., Ogata H., & Shimada A. (2017). A Meaningful Discovery Learning Environment for E-book Learners. *Proceedings of IEEE Global Engineering Education Conference 2017*, 1158-1165.
- Yin C-J., M. Yamada, Oi M., Shimada A., Okubo F., Kojima K., and Ogata H. (2018). Exploring the Relationships between Reading Behavior Patterns and Learning Outcomes Based on Log Data from E-Books: A Human Factor Approach, *International Journal of Human-Computer Interaction*, DOI: 10.1080/10447318.2018.1543077
- Yin C-J., Okubo F., Shimada A., Oi M., Hirokawa S., and Ogata H. (2015). Identifying and Analyzing the Learning Behaviors of Students using e-Books", *Proc. International Conference of Computers on Education 2015*.

Application of Programming Learning Support System to Object-Oriented Language

Satoru KOGURE^{a*}, Kento OGASAWARA^a, Koichi YAMASHITA^b,
Yasuhiro NOGUCHI^a, Tatsuhiro KONISHI^a & Yukihiro ITOH^c

^a*Faculty of Informatics, Shizuoka University, Japan*

^b*Faculty of Business Administration, Tokoha University, Japan*

^c*Shizuoka University, Japan*

*kogure@inf.shizuoka.ac.jp

Abstract: In the field of programming education, learners often cannot fully understand the behavior of algorithms and programs, because it is difficult for them to capture a realistic image of their behavior. Many researchers have proposed methods to visualize “the behavior of algorithm programs” as a solution to address this problem. Many programming learning support systems that incorporate these methods. We have developed a programming learning support system, TEDViT, that allows teachers to control the method of visualization. However, TEDViT is a programming language compatible with the C programming language, hence, the system cannot handle visualization in object-oriented languages. We developed and evaluated a prototype system that can visualize the proposed model.

Keywords: Programming learning support system, visualization, object-oriented programming language

1. Introduction

It is essential that beginning programmers are able to grasp and understand the program behavior correctly. However, it can be difficult for beginners to grasp concrete behavior. As a way to solve this problem, many program visualization (PV) systems have been proposed to visualize the program behavior (Pears, 2007, Moreno et al., 2014, Guo, 2013). These PV systems can only visualize the behavior of the program created by the learner based on specific methods decided by the tool developer. Therefore, it is not possible for teachers to freely depict what they want to explain in their teaching materials. To solve this problem, we have developed TEDViT (Teacher Explaining Design Visualization Tool), a C language behavior visualization system that can reflect the teacher's the explanation intention (Yamashita et al., 2018). We conducted many practical programming classes using TEDViT. The results obtained indicate that TEDViT can contribute to the understanding of the program behavior.

In this research, we propose Object-Oriented Conceptual Model (OOC Model) as a framework that can reflect the teacher's intention in the visualization of an object-oriented language. The OOC Model can express the structural relationship between classes and instances, which is one of the essential concepts to learn object-oriented languages. We constructed a prototype learning support system for Java that realized the OOC Model visualization. Herein, we report the experimental findings concerning the usability and usefulness of the prototype system we created.

2. Visualization for an Object-Oriented Language

To investigate problems of visualization for an object-oriented language, we need to select a target language. Although there are many object-oriented languages, there are some differences in the language specifications of each language. In this research, we focus on Java as an example of an object-oriented language. This research considers only the object-oriented language specific issues of the Java language.

In Java, basic concepts such as variables, arrays, and repeating structures are common to other programming languages. There is no need to newly consider the visualization of common concepts in this study, because their visualization is already possible using, for example, Jeliot 3 (Moreno et al., 2014), Python Tutor (Guo, 2013), TEDViT, among others. We investigated the concepts that require Java-specific visualization in Java programming classes offered in the first year of the computer science department. As a result of excluding the above-mentioned common parts, we only focus the relationship class and instance. We do not consider the exception handling and IO streaming topics in this study.

When learning the class and instance topics in object-oriented languages, it is necessary to understand the three concepts, (C1) the class definition and instance generation concepts (the relationship between a defined class and a generated instance), (C2) the inheritance concept (the relationship between a superclass and a subclass), and (C3) the polymorphism concept (invoking method of each class by the same-name method). The C3 cannot be visualized on the existing system. The requirement is to be able to specify the relationship between “class definition” and “instance” and to be able to specify which class’s method actually started. To visualize these relationships, we need a mechanism that can synchronously display the “program”, “class diagram” and “instance diagram”.

We propose the Object-Oriented Conceptual Model (OOC Model) as a visualization model that can realize the three concepts. To understand the program behavior of object-oriented languages, it is necessary to visualize diagrams showing the relationship between classes, methods, and instances. To show relationships between classes and instances, we can use the concepts of “class diagrams” and “object diagrams” used in UML diagrams. However, it is necessary to develop a framework that emphasizes class and instance (including fields and methods) in synchronization with class-instance relationships and program step execution.

3. System Implementation

Figure 1 shows an overview of the developed Prototype system. The left side shows the program, and the right side shows the OOC Model. The learner can proceed with single-step program execution by pressing the [>] button. In the Figure 1, the public void printMemberInfo() on the program side is surrounded by red lines. This indicates that the program is currently running up to this line. This example shows the situation after executing ml[1].printMemberInfo() has been executed. The ml[1] is an instance of SpecialMember, and the printMemberInfo() method defined in SpecialMember class should be activated. By highlighting both the program display on the left (the printMemberInfo () method in class SpecialMember has been invoked) and the OOC Model on the right (ml[1].printMemberInfo() and the printMemberInfo() method of the SpecialMember class) at the same time, the learner can gain a deep understanding of polymorphism.

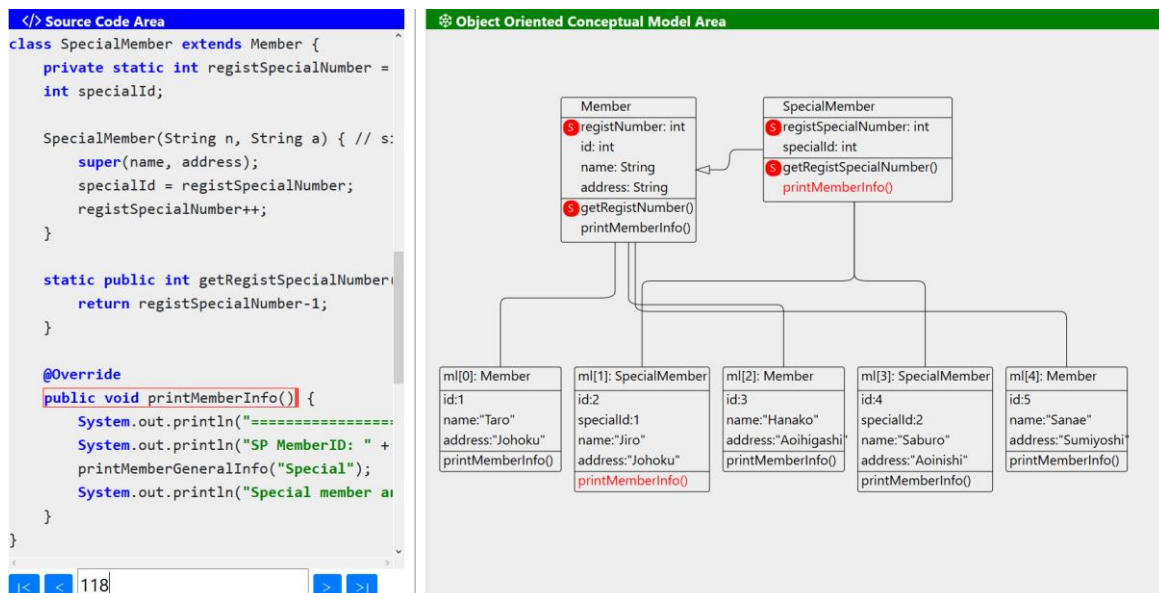


Figure 1. A Screen Shot of Prototype PV System.

4. Experimental Evaluation

The evaluation experiments investigate whether the following three hypotheses hold. **H1** is “the use of Java TEDViT raises the level of understanding of Java-specific concepts,” **H2** is “the OOC Model is an aid to learners of programming,” and **H3** is “by using Java TEDViT, it is possible to understand the flow of a program’s operation, predict of its output, and understand the reasons for its behavior.” The subjects for the experiments were 17 university students in the information system area. We gave the subjects the following procedure: (1) View a system tutorial, (2) Answer a questionnaire before the experiment, (3) Use the system, and (4) Answer a questionnaire after the experiment.

First, the rate of increase in self-evaluation for class and instance (**C1**) was high at 82.3%. Next, the rate of increase in self-evaluation for inheritance (**C2**) was 76.5%. We noted the opinion that “step flow of the constructor was understood well by executing step-by-step in the system”. Finally, the rate of increase in self-evaluation for polymorphism (**C3**) was 64.7%. It is considered that this did not improve much because the support system did not actively explain polymorphism. This result weakly supports that **H1** holds.

Moreover, we obtained subjective questionnaire on the usefulness of the system after the experiment. The results show that although the system does not adversely affect learning, it does not contribute to a clear improvement in understanding. Therefore, it is judged that this questionnaire result rejects **H2** and **H3**. On the other hand, there were many positive opinions on the mechanism of proposed prototype PV system itself, and we obtained the result that the expectation for practical use can become higher by further improvement.

5. Conclusion

In this research, we examined the visualization of program behavior of an object-oriented language and proposed an Object-Oriented Conceptual Model (OOC Model). We also constructed a prototype of a PV system that allows teachers to control the visualization of program behavior based on the OOC Model. Evaluation of the constructed system was carried out by 17 university students. The results suggested that the use of a prototype system improves the understanding of object-oriented learning concepts. Although the evaluation of the usefulness of the system was not good, we obtained the finding that it would be useful if the system were improved.

In the future, therefore, we plan a number of enhancements, to include improving the user interface of the system, simplifying the rule setting method for the teacher, and using the system in an actual classroom setting.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Numbers JP16K01084, JP18K11567, 19K12259 and, JP19K12265.

References

- Guo, P. J. (2013). Online Python Tutor: Embeddable Web-Based Program Visualization for CS Education. *ACM Technical Symposium on Computer Science Education (SIGCSE)*.
- Moreno, A., Sutinen, E., & Joy, M. (2014). Defining and evaluating conflictive animations for programming education: the case of Jeliot ConAn. *Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14)* (pp. 629-634). New York, NY, USA.
- Pears, A., Seidman, S., Malmi, L., Mannila, L., Adams, E., Bennedsen, J., Devlin, M., & Paterson, J. (2007). A survey of literature on the teaching of introductory programming. *ACM SIGCSE Bulletin*, 39(4), 204-223.
- Yamashita, K., Tezuka, D., Kogure, S., Noguchi, Y., Konishi, T., & Itoh, Y. (2018). A learning support system for visualizing behaviors of students' programs based on teachers' intents of instruction. *Proceedings of the 26th International Conference on Computers in Education* (pp.761-766).

Learning Evidence Analytics Framework (LEAF) in Practice: A2I2 based Teacher Adoption Approach

Rwitajit MAJUMDAR^{a*}, Jayakrishnan M. Warriem^b, Hiroyuki KUROMIYA^a, Gökhan AKÇ APINAR^{a,c}, Brendan FLANAGAN^a & Hiroaki OGATA^a

^aAcademic Center for Computing and Media Studies, Kyoto University, Japan

^bNPTEL, Indian Institute of Technology Madras, Chennai, India

^c Computer Education & Instructional Tech, Hacettepe University, Turkey

*dr.rwito@gmail.com

Abstract: Learning Analytics (LA) platforms can gather data from the teaching-learning interactions during a course. While there have been previous discussions regarding the individual tools, limited scholarship describes the utility of a LA framework for supporting evidence-based teaching-learning practices. We have proposed LEAF, a framework to bridge that gap. We implement the framework in a platform by integrating LMS, learning behaviour sensors such as an ebook reader, learning analytics dashboard and an evidence portal through Learning Tools Interoperability (LTI). The platform was then made available to teachers from different colleges in India to orchestrate their course offering for one semester. This paper describes the design of the teacher training module for the adoption of the platform based on the A2I2 model as its theoretical basis. The A2I2 model explicitly focuses on encouraging scholarship of learning and teaching among participating teachers and thus is an ideal candidate for utilizing an evidence-based framework.

Keywords: LEAF, A2I2, Evidence-based Education and Learning, TEEL

1. Introduction and Background

There are 10,426 of engineering and technology colleges in India affiliated to the All India Council for Technical Education (AICTE). However, not all of them set up their own learning management system (LMS) and use e-learning tools to potentially enrich the teaching and learning experiences. In class, teachers follow traditional didactic teaching lecturing with the aid of presentation tools such as powerpoint. In the context of India, earlier attempts were in the form of faculty development programs (FDPs) to nudge higher education teachers from didactic classroom-based teaching to adopt more learner-centric active learning strategies using educational technology. These training programs were offered in face-to-face (Warriem, Murthy & Iyer, 2013), asynchronous online (Warriem, Murthy & Iyer, 2015) and even large-class MOOCs (Warriem, Murthy & Iyer, 2016). This cross-country collaborative work investigates the utility of providing a learning platform integrated with learning analytics framework to the college teachers to set up and conduct their semester-long courses. Our Learning Evidence Analytics Framework (LEAF) includes existing e-learning and learning analytics tools connected through Learning Tools Interoperability (LTI) to gather and analyse learning logs captured in standard xAPI format (Flanagan & Ogata, 2017; Ogata et.al. 2018, Majumdar et.al. 2019). Over a period of one semester, interested teachers and their students from colleges across India accessed Moodle as LMS, BookRoll, an ebook based learning system and LAViEW, a learning analytics tool which had the dashboard and initial design of the evidence portal. This article describes our approach to design a training module for teachers who are adopting any learning analytics framework for the first time in their teaching practice. Based on it we study how the teachers participated across the phases of adopting the framework and how the platform was used by their students while they conduct courses in their college over a semester.

2. Implementing LEAF for instructors in India

The developed framework was implemented as service through the LET lab Moodle and we provided access to the interested Indian instructors. Further, we designed the TEEL workshop, followed by a TEEL coordinating course on the Moodle. The various activities involved to introduce the framework followed A2I2 model over four phases. Table 1 provides the details of the phases and the mapping with the A2I2.

Table 1 *Design of Adoption phases based on A2I2 model*

Adoption phase	Description	A2I2 phase	Outcomes
Faculty Development Program	A MOOC on “Mentoring Educators in Educational Technology” was offered to college teachers as FDP.	Attain	The teachers were introduced to the basics of offering a Small Private Online Course and designing active learning strategies while using online learning management systems like MOODLE
Workshop for exposure to the LA platform	A workshop with hands on activities to get introduced to the TEEL framework	Align	The teachers link the knowledge gained in the MOOC to the practice system introduced during the workshop
Preparations for course offerings	The teachers create the course material and prepare the Moodle to offer the course	Integrate	The teacher integrates the LMS and the tools offered in their own semester course
Orchestration and Reflection	The course is run with students and teachers conduct a reflection on the activities offered.	Investigate	The teacher offers the course and links reflects on the course with the help of the dashboards and the surveys given to the students.

As seen in Table 1, the starting point of the workshop was a MOOC specifically aimed at helping teachers understand the several aspects involved in the creation, offering and analysis of online courses. As part of the MOOC, the participating teachers collaboratively (in groups of 4) developed a module of a Small Private Online Course (SPOC) using MOODLE. They also participated as learners in the SPOCs created by their peers and thereby generated the learner data for each of the course. To complete the MOOC, the participants had to submit a summarized report regarding perceived and actual engagement and learning in the course and their direction of refinement based on that.

It is followed by a workshop on the TEEL infrastructure with the LA platform. The exposure in the workshop regarding accessing analytics intended to highlight the relevant features available in the LA platform. Once the teachers are familiar with the analytics infrastructure they will start uploading their course-related resources and start the teaching & learning activities. During the orchestration of course activities, the LAView system will help the teacher in reflecting on the students’ actions and refine their own instructional activities.

3. Participant engagement across phases of adoption

We wanted to investigate *What is the extent of teachers’ engagement across different phases of the framework adoption?* The LA framework was introduced to the teachers by following the A2I2 model and sent an invitation to adopt LEAF platform to 533 different college teachers in India. These teachers registered in a MOOC course titled Mentoring Educators Educational Technology. Participation in all the four phases was voluntary. Those teachers who participated in each phase were the sample of this study. The MOOC had the largest amount of participation and there was consistent attrition across the multiple phases. This was expected as the effort was based on a voluntary participation model. 144 of the teachers involved in the TEEL workshop and 75 of them prepared for their course. However, if we take a look at the geographic spread of the

participants (Fig 2a), we see that even with this attrition there was a significant amount of diversity in the regions that people are participating from. This diversity is also seen in the variety of courses intended to offer (Fig 2b) from varying areas of STEM as well as Management, Pharmacy and Humanities found representation. Later 28 of them created their course and 22 of them orchestrated on the platform. Total 628 of their students were active on the platform out of the 1,463 registered in the TEEL Moodle.

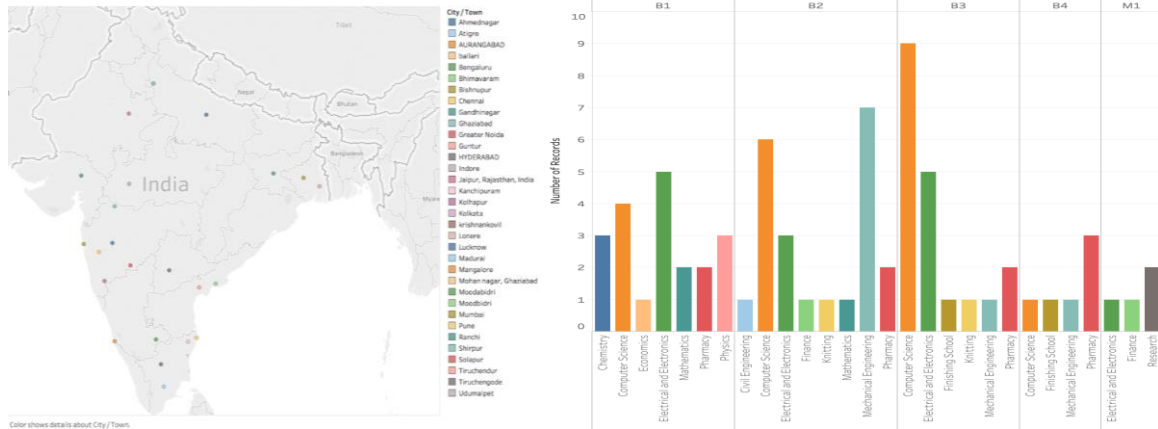


Figure 2. Location of colleges and courses planned by the participants in the preparation phase

We have presented the initial descriptive analysis of the courses offered. We further plan to conduct a detailed analysis of the teaching-learning experience with the data logs collected in the LEAF platform. We envision the evidence portal to emerge as a central platform for dialogues between teachers and researchers. Teachers would provide the actual context data and issues from the field (their courses) where the researchers can then collaboratively ideate about possible solutions and further systematically evaluate the result of its implementation and save them as a case for evidence.

Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304, NEDO Special Innovation Program on AI and Big Data 18102059-0 and JSPS KAKENHI Research Activity Start-up Grant Number 18H05746 (F.Y.2018), 19K20942 (F.Y.2019).

References

- Flanagan, B., & Ogata, H. (2017) Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In Proc. of the 25th ICCE 2017, Christchurch, New Zealand.
- Majumdar R., I Boticki., Ogata H., International Workshop on Technology-Enhanced and Evidence-Based Education and Learning , In Companion Proc. of the 9th LAK, Tempe, USA.
- Murthy, S., Iyer, S. and Warriem, J.M.. (2015). ET4ET: A Large-Scale Faculty Professional Development Program on Effective Integration of Educational Technology. J. of Edu Tech. & Society, 18(3), 16-28.
- Ogata H., Majumdar R., Akçapınar G., Hasnine M.N., Flanagan B., Beyond Learning Analytics: Framework for Technology-Enhanced Evidence-Based Education and Learning, In Proc. of the 26th ICCE 2018, Manila
- Warriem, J.M., Murthy, S., and Iyer, S.(2013). A2I: A Model for Teacher Training in Constructive Alignment for Use of ICT in Engineering Education, In Proc. of 22nd ICCE 2013, Nara, Japan.
- Warriem, J.M., Murthy, S., and Iyer, S.. (2015). Sustainability at Scale: Evidence from a Large Scale Teacher Professional Development Program. In Proc. of 23rd ICCE 2015, Hangzhou, China.
- Warriem, J.M., Murthy, S., and Iyer, S.(2016). Shifting the focus from Learner Completion to Learner Perseverance: Evidences from a Teacher Professional Development MOOC. In Proc. of 24th ICCE 2016, Mumbai, India.

Extending Program Visualization System Based on Teacher's Intent of Instruction to Support Learning Dynamic Data Structures

Koichi YAMASHITA^{a*}, Yuta HIRAMATSU^b, Satoru KOGURE^b, Yasuhiro NOGUCHI^b,
Tatsuhiko KONISHI^b & Yukihiro ITOH^c

^a*Faculty of Business Administration, Tokoha University, Japan*

^b*Faculty of Informatics, Shizuoka University, Japan*

^c*Shizuoka University, Japan*

*yamasita@hm.tokoha-u.ac.jp

Abstract: In this paper, we describe an extension of TEDViT, the program visualization system that can reflect teachers' intent of instruction to support learning dynamic data structures. We extend TEDViT to enable the definition of drawing rules using values of target program variables and pointers corresponding to the execution status when the rule is fired. The evaluation results reveal that our extension could significantly reduce the amount of drawing rules and the time required for defining program visualizations. It suggests that our approach would support program visualizations with dynamic data structures to a certain degree.

Keywords: Programming education, program visualization system, program visualization design, dynamic data structures

1. Introduction

To date, several program visualization (PV) systems have been developed to support novices who are learning programming (Pears et al., 2007). We have developed a PV system called Teacher's Explaining Design Visualization Tool (TEDViT) that allows teachers to provide not only a target program, but also its visualization policy (Kogure et al., 2014). The teacher can define, by providing the drawing rules to TEDViT, which variable is visualized with which drawing objects and when they are visualized in the process of program execution. However, TEDViT becomes complex when it comes to visualizing program behaviors with dynamic data structures (DDS) because defining using drawing rules is costly.

In this paper, we describe an approach to extend TEDViT so that it can visualize the program behaviors with DDS. The reasons for the high cost of defining drawing rules on DDS in TEDViT are difficulties in expressing a certain location of main memory for drawing rules and assigning a certain drawing position of objects. To address these issues, we extend the drawing rule representations by enabling the possibility of expressing certain data on memory in a more abstract way. We also increase the abstraction level for position expression of drawing objects, thereby enabling the possibility of assigning a relative position to other drawing objects.

This paper proves that our extension could enable the visualization of program behaviors processing linked list, which is one of the typical DDS, at relatively low cost. We evaluate how much cost is reduced when defining a drawing ruleset in our extension. The evaluation results show that our extension could significantly reduce costs when defining a drawing ruleset in extended TEDViT.

2. Visualizing Behaviors of Programs with Dynamic Data Structures

TEDViT allows teachers to define the policy for drawing a status of the target domain world based on their intents of instruction. Teachers can create or edit a configuration file independently from the target program file, which defines a visualization policy comprises a set of drawing rules, each of which is a

CSV entry as in Figure 1. The drawing rule shown in Figure 1 means that when the statement with ID “10” in the target program is executed, TEDViT draws a circle object and assigns the object ID “OBJ1” to it. The corresponding variable is indicated as “i,” and hence the value of the variable *i* is drawn inside OBJ1. OBJ1 is drawn at position (x1, y1) with a line, background, and inner character colors of black, white, and black, respectively, according to the indicated values in the rule.

```
state==10, create, OBJ1, circle, i, x1, y1, black, white, black
```

Figure 1. Example of the drawing rule in TEDViT.

When defining PV with DDS, teachers often demand the dynamic location of drawing objects depending on the program behaviors. For example, in a linked list node insertion, the target node to be inserted is naturally drawn near the linked list, at the position where the target is inserted into. Teachers may also demand to highlight some drawing objects according to specific timing, to provide natural-language explanations by accompanying some objects with balloon messages, etc. However, the original TEDViT required that the drawing objects and their positions be statically indicated by object ID and absolute coordinates in pixel or grid measure. Hence, we extend TEDViT to be capable of indicating drawing objects by using values of target program variables corresponding the execution status when the rule is fired, and of indicating coordinates relative to any other drawing objects in drawing rules. In drawing rule description, teachers can describe the following functions to indicate target objects and their positions that are origins of relative coordinates;

- `xname(ObjectID)`, `yname(ObjectID)`: finds x-/y-coordinate of the drawing object with the ID *ObjectID* given by the argument.
- `xvari(VariableName)`, `yvari(VariableName)`: finds x-/y-coordinate of the drawing object corresponding to the variable with the name *VariableName* given by the argument.
- `xaddr(PointerExpression)`, `yaddr(PointerExpression)`: finds x-/y-coordinate of the drawing object corresponding to the memory location referred by the pointer *PointerExpression* given by the argument. The value of *PointerExpression* depends on the execution status when the rule is fired.
- `searchObjectbyValue(PointerExpression)`: finds drawing object corresponding to the memory location referred by the pointer *PointerExpression* given by argument.
- `searchObjectbyname(VariableName)`: finds drawing object corresponding to the memory location that stores the value of variable with the name *VariableName* given by argument.

In rule definition, teachers sometimes need to describe variable names for indicating memory locations corresponding to drawing objects. We also extend TEDViT to be capable of describing the following function:

- `searchVariable(PointerExpression)`: finds variable name corresponding to the memory location referred by the pointer *PointerExpression* given by the argument.

3. Experiment

Since PVs generated by the extended TEDViT basically follow the ones generated by original TEDViT, it is expected that the extended TEDViT would provide the same level of learning effects as the original TEDViT. Therefore, we conducted an experiment to evaluate the cost to define drawing rules for PV with DDS. The number of subjects were 8: two teachers who have experience of teaching programming, two students who have experience as programming teaching assistants (TAs), and four students who have the same level of programming experience as programming TAs.

At the beginning of the experiment, we explained an outline of the experiment, the defining of PVs, and the use of original and extended TEDViT to the subjects, for 30 min. Then, the subjects defined PVs of a program consisting of about 40 statements. They defined PVs in two-time intervals, first with the original TEDViT in 60 min, and then with the extended TEDViT in 60 min. To reduce the order effects, we divided the subjects into groups A and B, each of which included one teacher, one TA, and two students, and swapped the order in which they used the original and the extended TEDViT. The subjects in group A used the original TEDViT first, while those in group B used the extended TEDViT

first. For each of the two rule definitions, we measured the minutes taken to define entire drawing rules and the number of defined rules for every subject.

Table 1 shows the results of the experiment. The value in parentheses that is appended below the rule definition time with original TEDViT indicates the time estimated to complete the entire rule definition. Given that the subjects had to define the drawing rules within a time limit of 60 min, no subject except #8 could complete their rule definition within that time. We estimated the ruleset completion times using the ratio of the number of the total rules we expected to be answered to the number of drawing rules defined in 60 min.

Table 1

The Defining Time and the Number of Drawing Rules

Subject		#1	#2	#3	#4	#5	#6	#7	#8
Group		A	A	A	A	B	B	B	B
N of rules	w/ Original	46	40	58	39	42	18	18	70
	w/ Extended	8	8	10	8	8	8	8	8
Time (min.)	w/ Original	60*	60*	60*	60*	60*	60*	60*	55
	(Estimated)	(88)	(102)	(70)	(104)	(97)	(226)	(226)	
	w/ Extended	17	29	26	21	30	51	59	23

(* represents the time for subjects who did not finish the rule definitions.)

We can see that the numbers of drawing rules to define PV and the rule definition intervals are significantly reduced by using the extended TEDViT. Note also that the average number of rules in the extended TEDViT is 24.8% of those in the original version. The same tendency can be seen in group A, 18.8%, and group B, 29.8%. Based on the estimated times, the subjects could have completed their rule definition with the extended TEDViT in 28.3% of the interval required with the original, on average. The same tendency also can be seen in group A, 26.2%, and group B, 30.0%. These results suggest that the cost to define PVs with DDS could be significantly reduced by our extension.

4. Conclusion

In this paper, we describe an extension of a PV system called TEDViT to enable the visualization of program behaviors with dynamic data structures. We extend TEDViT to enable the possibility of specifying the target and its position through values of target program variables corresponding the execution status when the rule is fired. It means to enhance the abstraction level of rule descriptions, such as specifying a drawing object referred by a certain pointer, specifying a drawing position like “in the vicinity of a certain object,” etc. The evaluation results revealed that the extended TEDViT could significantly reduce the amount of drawing rules and the time to define PVs likewise for both groups. This suggests that the extended TEDViT could enable to visualize the program behaviors processing linked list, which is a typical DDS, at relatively low cost. Therefore, it can be concluded that our extension would help to define TEDViT PV with DDS to a certain degree.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Numbers JP16K01084, JP18K11567, and JP19K12259.

References

- Kogure, S., Fujioka, R., Noguchi, Y., Yamashita, K., Konishi, T., & Itoh, Y. (2014). Code reading environment according to visualizing both variable's memory image and target world's status. Proceeding of the 22nd International Conference on Computers in Education, 343-348.
- Pears, A., Seidman, S., Malmi, L., Mannila, L., Adams, E., Bennedsen, J., Devlin, M., & Paterson, J. (2007). A Survey of Literature on the Teaching of Introductory Programming. *ACM SIGCSE Bulletin*, 39(4), 204-223.

An Index System of Education Information Resources Selection Based on Analytical Hierarchy Process

Jiahui RUAN^{a*}, Yiling HU^a, Xiaoling PENG^a

^a*Department of Education Information Technology, East China Normal University, China*

*rjhskye@gmail.com

Abstract: In the face of massive learning resources, learners are confronted with resource selection dilemmas and cognitive load problems. In order to help learners or educators to choose personalized resources better, we combed the literature reviews on classic learner models, and from the perspective of learners' personalized characteristics, we analyzed and constructed an index system of education information resource selection based on college student preferences by using analytical hierarchy process (AHP), then the weight of each factor in the index system was determined through Delphi method, a specific course was taken as an example afterwards, which verified that the index system constructed is effective and scientific.

Keywords: Analytical hierarchy process, education information resources, personalized characteristics, learner preferences

1. Introduction

The contradiction between the surge of education information resources and the differences of learners' resource demands has led to information overload. Constructing a scientific index model of resource selection and mapping strategy has become hot off the press. In this case, what learners need indeed are to be urgently explored. Despite learner-centered instruction has long been the core idea of open education, the analysis on students' explicit and implicit needs seem to be continually neglected more or less, which is the very beginning point of realizing precise online education service. In order to improve the design effects of online learning systems and the quality of individualized learning resources recommendation, it is necessary to build a comprehensive, accurate, scientific and reasonable learner profile. The learner profile is a structured representation and storage of learners' preferences, cognition, abilities and situations, and then chooses appropriate schemas to describe learners' explicit and implicit characteristics. The purpose of the current study is first, to establish a learner characteristics profile served as a hierarchical structure model to construct an index model of education information resource selection and second, to verify the effectiveness of the index model to guide our further refinement on the models. The research question is proposed as follows:

RQ: How to construct a hierarchical index evaluating model to solve multi-criteria decision making with uncertain factors and help different learners select an optimal learning resource?

2. Construction of Index Model for Education Information Resource Selection

2.1 Brief Introduction of Analytical Hierarchy Process

Numerous highly uncertain factors are affecting learners' preferences for resource selection in specific time and places, some of which may even conflict with each other. Analytical hierarchy process (AHP), a qualitative and quantitative method, can be applied to decision analysis in various uncertain situations (Emrouznejad & Marra, 2017). Not rare are research on the application of AHP in the field of education, which can be roughly divided into the following five categories: (a) the quality evaluation of learning resources, (b) teaching quality evaluation, combined with Data Envelopment Analysis (DEA) or

Multi-choice Goal Programming (MCGP) method to reduce the non-computational error of AHP (Thanassoulis et al., 2017; Gurung & Phipon, 2016), (c) evaluation of students' learning quality, highly requiring a diversification assessing indicators which can reflect the internal and logical relationship, as well as conciseness and expansibility, (d) allocation of education resources, which is decided by balancing the weights and relationships among different indicators, (e) analysis of phenomenal factors, including the construction of problem models and the analysis of factors restricting or proposing suggestions through empirical methods.

2.2 Establishment of Hierarchical Structure Model

Through the comparative analysis of typical learner characteristics profiles and interviews with several experts as well as college students in related fields, this study proposes a learner characteristics profile based on college students' online learning preferences and personalized features. The complex problems of influencing factors of college students' education information resources selection can be divided into four levels: objective level O , first-class factor (indicator) level C , second-class factor (criterion) level P and alternative level A . The alternative level includes teaching plan (A_1), teaching courseware (A_2), micro-lesson video (A_3) and case-based video recording (A_4). The final hierarchical model is then established as is shown in Figure 2.

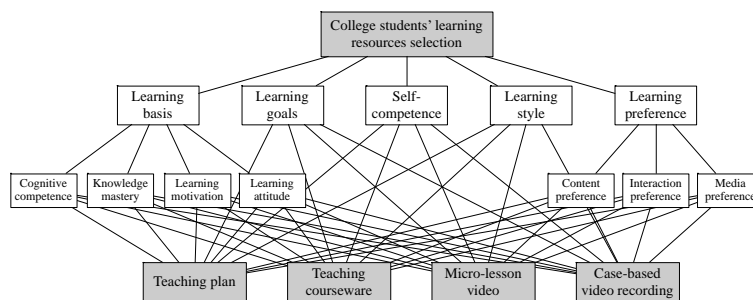


Figure 2. Hierarchical structure model.

2.2.1 Construction of Judgment Matrix

In order to obtain more reasonable and authoritative data, and therefore determine the scale of the impact of learner characteristics on resource selection, four experts were invited to compare the factors in the hierarchical structure model, assigning relative importance of each item as shown in Figure 3.

A	Importance comparison															B		
Learning basis	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning preference
Learning basis	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning style
Learning basis	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Self-competence
Learning basis	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning goal
Learning preference	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning style
Learning preference	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Self-competence
Learning preference	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning goal
Learning style	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Self-competence
Learning style	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning goal
Self-competence	◀9	◀8	◀7	◀6	◀5	◀4	◀3	◀2	1	▶2	▶3	▶4	▶5	▶6	▶7	▶8	▶9	Learning goal

Figure 3. Evaluate the relative importance of learner characteristics for choosing resources.

2.2.2 Hierarchical Single Ranking and Consistency Check

Through iterative feedback and modification, a judgment matrix $E_k (k = 1, 2, 3, 4)$ is therefore obtained for each expert at the factor level. Hierarchical single ranking refers to the ranking order of comparing all elements in this layer with those in the upper layer. In this study, MATLAB is used to calculate the maximum eigen root λ_{max} of matrix E_k , corresponding eigenvector W and weight vector w by sum method. The weight vector results are shown in Table 4.

Table 4

Combination weight vector result

Indicator level	Weight	Rank	Criterion level	Weight	Rank
Learning goal C_1	0.5050	1	/	/	/
Learning basis C_2	0.1334	3	Cognitive competence p_1	0.0554	1
			Knowledge mastery p_2	0.0474	3
			Learning motivation p_3	0.0118	7
			Learning attitude p_4	0.0188	4
Learning preference C_3	0.0739	4	/	/	/
Learning style C_4	0.0559	5	Content preference p_5	0.0486	2
			Interaction preference p_6	0.0119	6
			Media preference p_7	0.0133	5
Self-competence C_5	0.2318	2	/	/	/

3. Results

Course “Animation design and production” was taken to verify the validity of the above model results, resource preference selection experiments on 60 sophomores in two classes of a university in China were carried out. Each student has a computer to receive the distribution of four kinds of learning resources. Each type of resource has two files and strictly conforms to the curriculum standards and schedule which has been applied in practice before.

Students were required to select two of their favorite resources after browsing 8 learning resources and fill in the questionnaire. After eliminating and counting the invalid data, a total of 114 resource choices of 57 students were collected. The results of comparison between the weights of index model and experiments are shown in Figure 4.

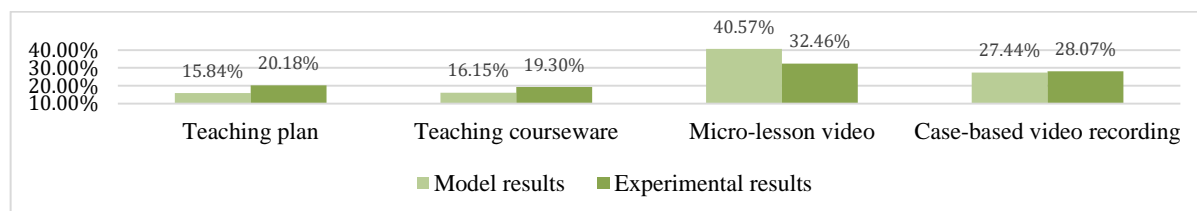


Figure 4. Comparison of the results of model and experimental data weights

4. Conclusion

This study constructed a learner characteristic model based on the personalized features of college students through comparative analysis of several classical learner characteristic models. Expert surveys and analytical hierarchy process were carried out to assign index weight of each learner characteristic factor. A course was taken and five learner characteristics with their sub-criteria factors in the index model were successfully verified through the experiment. The study provides a theoretical support for the application of learner models, as well as the design and development of algorithms for college students’ personalized resource recommendation. More systematic selection and evaluation of learner characteristics combined with interpretive structural model methods are to be improved, to solidify the authority of learner models applied to specific research.

References

- Emrouznejad, A., & Marra, M. (2017). The state-of-the-art development of AHP (1979-2017): a literature review with a social network analysis. *International Journal of Production Research*, 1-23.
- Gurung, S., & Phipon, R. (2016). Multi-criteria decision making for supplier selection using AHP and TOPSIS method. *International Journal of Engineering Inventions*, 6(2), 13-17.
- Thanassoulis, E., Dey, P. K., Petridis, K., Goniadis, I., & Georgiou, A. C. (2017). Evaluating higher education teaching performance using combined analytic hierarchy process and data envelopment analysis. *Journal of the Operational Research Society*, 68(4), 431-445.

Developing E-Book Page Ranking Model for Pre-Class Reading Recommendation

Christopher C.Y. YANG^{a*}, Gökhan AKÇ APINAR^{b,c}, Brendan FLANAGAN^b & Hiroaki OGATA^b

^aGraduate School of Informatics, Kyoto University, Japan

^bAcademic Center for Computing and Media Studies, Kyoto University, Japan

^cDepartment of Computer Education and Instructional Technology, Hacettepe University, Turkey

*yang.yuan.57e@st.kyoto-u.ac.jp

Abstract: In this paper, we propose an E-Book Page Ranking (EBPR) method to rank e-book pages from the original learning material automatically. The proposed method ranks all the e-book pages by the class probabilities retrieved from machine learning models. The top-ranked e-book pages are then selected to form the pre-class reading (preview) recommendation. The proposed method extracts image features and text features from e-book page contents as well as the e-book usage features from students' previous reading logs. In this paper, we test the performance of the proposed model with two different cases, with and without past e-book usage data. The experimental results showed the improvability of the model after taking into account learners' past e-book usages.

Keywords: E-book page ranking, educational recommender systems, e-book usage logs, preview reading

1. Introduction

Preparing for lectures is crucial for learners and has numerous advantages over their learning. A study conducted by Dreher and Sammons (1994) indicated that students who previewed the learning material before the class were able to answer the questions in exams better than students who did not preview. However, learners are not always willing to prepare for the class, especially when learning materials are too long for them. Another study conducted by Shimada et al. (2017) revealed that giving learners a subset of learning material (important pages) instead of the whole content, increased their preview behaviors and overall learning performances. In the context of e-learning, the characteristics of important document page/e-book page and the associated content features have been mentioned and applied in several articles (Neto et al., 2002; Shimada et al., 2017). Additionally, in this paper, we assumed that the learners' past e-book usage that related to the target e-book material should be taken into account when considering e-book preview recommendation. The proposed method can be potentially anticipated to well reduce the overhead for course preview material creation on e-book system. In this study, we attempt to answer the following two research questions:

1. What is the best-performed machine learning algorithm for the proposed E-Book Page Ranking (EBPR) method?
2. Can we improve the performance of EBPR by adding features to the model related to learners' past e-book usage?

2. Machine Learning-Based E-Book Page Ranking Method

2.1 Data Collection and Feature Extraction

The entire performing process is described as follows: we use BookRoll system (Ogata et al., 2015) which is a digital textbook reading system that contains plenty of functions such as page turning, internal learning content searching, page jumping, annotation creation, annotation transfer across

e-book revisions (Yang et al., 2018), etc. Learners' reading behaviors while using BookRoll store in the database of BookRoll. In this study, two types of data were collected from BookRoll as the input of the proposed method. The first type of data is text contents and image contents from the original e-book material that contains 91 e-book pages in BookRoll named Semantic Web Services with 38 learners enrolled under the period of 3 weeks in the previous semester. The second type of data is the recorded 10147 e-book reading events that related to the same e-book material. The reading events tracked by BookRoll has been described in the previous article (Ogata et al., 2015). In the feature extraction, we applied text processing techniques, image processing algorithms including background subtraction method and inter-frame difference method, as well as the educational data mining methods to extract text, image, e-book usage features from the collected e-book page contents and students' e-book reading events as described in Table 1.

Table 1

Description of the Extracted Features

Category_Index	Feature Name	Feature Description
Text_01	<i>TotalChar</i>	Total characters in a page
Text_02	<i>AvgTFIDF</i>	Average of TFIDF (Term Frequency-Inverse Document Frequency) value
Text_03	<i>Similarity to title</i>	Cosine similarity to the title of the content
Text_04	<i>Similarity to keywords</i>	Cosine similarity to the keywords of the content
Text_05	<i>Page-Page cohesion</i>	Sum of cosine similarities to the rest pages
Text_06	<i>Punctuation</i>	Total occurrence of punctuations in a page
Image_01	<i>Background subtraction</i>	Foreground pixels in a page
Image_02	<i>Background subtraction + Inter-frame difference</i>	Absolute foreground pixel differences with previous page and next page (choose the higher value)
Usage_01	<i>Marker</i>	Total number of marker added in a page
Usage_02	<i>Memo</i>	Total number of memo added in a page
Usage_03	<i>Bookmark</i>	Total number of bookmark added in a page
Usage_04	<i>UniqueVisit</i>	Total number of learners visited a page
Usage_05	<i>TotalTime</i>	Total reading time on a page
Usage_06	<i>AvgTime</i>	Average reading time on a page per learner
Usage_07	<i>TotalEvent</i>	Total clicking events in a page
Usage_08	<i>AvgEvent</i>	Average clicking events in a page per learner

2.2 Modeling, E-Book Page Ranking, and Generation of E-Book Preview Recommendation

To train the machine learning models we asked one lecturer to label important e-book pages in his learning content (gold-standard). The lecturer labeled 45 out of 91 pages as important pages (recommended for preview). Rest of the pages considered as less important for preview. We formed our problem as a binary classification problem, however, we obtained class probabilities as the e-book page ranking output instead of class labels, which gave us the flexibility of ranking pages based on their importance. The top-ranked e-book pages are then selected to form the pre-class reading recommendation.

3. Test Results

To evaluate the ranking performance of the proposed EBPR method, we first selected the top-ranked 45 e-book pages from each model (the same number of pages are labeled as important page by the lecturer). Model performances were evaluated by using metrics such as precision, recall, and Area Under the Curve (AUC) along with the 3-folds cross validation. To evaluate the improvability of models after taking into account learners' past e-book usages, we conducted two different experiments which are past e-book usage feature exclusion and inclusion, respectively. As shown in Table 2, in the first experiment, e-book usage features were excluded from the training process, resulting of the

best-performed model MultilayerPerception with precision 0.667, recall 0.667, and AUC 0.67. In the second experiment, e-book usage features were included from the training process of models, the observed result in this experiment indicates that the best-performed model was still MultilayerPerception with precision 0.756, recall 0.756, and AUC 0.758.

Table 2

Performance of each model (excluding / including students' past e-book usage logs)

Model	Precision	Recall	AUC
<i>DTNB</i>	0.422 / 0.556	0.422 / 0.556	0.429 / 0.560
<i>JRip</i>	0.489 / 0.556	0.489 / 0.556	0.494 / 0.560
<i>RandomForest</i>	0.556 / 0.600	0.556 / 0.600	0.560 / 0.604
<i>J48</i>	0.533 / 0.511	0.533 / 0.511	0.538 / 0.516
<i>BayesNet</i>	0.422 / 0.422	0.422 / 0.422	0.429 / 0.429
<i>GaussianNaïveBayes</i>	0.467 / 0.467	0.467 / 0.467	0.472 / 0.472
<i>LogisticRegression</i>	0.622 / 0.733	0.622 / 0.733	0.626 / 0.736
<i>MultilayerPerception</i>	0.667 / 0.756	0.667 / 0.756	0.670 / 0.758

4. Conclusion and Future Work

In this paper, we proposed a machine learning-based E-Book Page Ranking (EBPR) method for the recommendation of e-book preview that can be integrated into any e-book systems. We ranked the input e-book pages by the predicted class probabilities accordingly. After the ranking process, the top-ranked e-book pages are selected to form the recommendation of preview material for learners before a class. We compared several classification models and reported the best-performed model *MultilayerPerception*, which answered our first research question. We evaluated the performances of e-book page ranking in different conditions through two experiments. The statistical results shown in Table 2 reported that when including learners' past e-book usage features, the overall performance of e-book page ranking will be improved, which answered our second research question. In the future, we will look for more e-book page samples from lecturers in different domains as the training samples to investigate and evaluate the ranking performance of the model and the most important features.

Acknowledgments

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

- Dreher, M. J., & Sammons, R. B. (1994). Fifth graders' search for information in a textbook. *Journal of Reading Behavior*, 26(3), 301-314.
- Neto, J. L., Freitas, A. A., & Kaestner, C. A. (2002, November). Automatic text summarization using a machine learning approach. In *Brazilian Symposium on Artificial Intelligence* (pp. 205-215). Springer, Berlin, Heidelberg.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).
- Shimada, A., Okubo, F., Yin, C., & Ogata, H. (2017). Automatic Summarization of Lecture Slides for Enhanced Student Preview—Technical Report and User Study—. *IEEE Transactions on Learning Technologies*, 11(2), 165-178.
- Yang, C.C.Y., Flanagan, B., Akcapinar, G., & Ogata, H. (2018). Maintaining reading experience continuity across e-book revisions. *Research and practice in technology enhanced learning*, 13(1), 24.

Discovering the Impact of Student Communities in Educational Recommendations

Yong Zheng

Illinois Institute of Technology

Chicago, Illinois, USA

yzheng66@iit.edu

Abstract: Recommender systems have been applied as one of the technology-enhanced learning techniques for educations. Recently, multi-stakeholder recommender systems were proposed to balance the needs among different stakeholders, especially when there are conflicts of interests. In this paper, we create student communities by using the clustering technique, and seek the impact of these communities in the educational recommendations. Our experimental results identify the best community which can improve the multi-stakeholder recommendations.

Keywords: recommender system, stakeholder, education, clustering, student community

1. Introduction and Related Work

Recommender systems, as one of the technology-enhanced learning methods, has been introduced to the area of educational learning. For example, it is able to suggest books for K-12 users (Pera, 2016), recommend after-school programs (Burke, 2011), or suggest appropriate citations (He, 2010) in paper writings. We successfully have utilized the personality traits to build effective personality-aware recommendation algorithms (Zheng, Subramaniyan, 2019), and develop multi-stakeholder recommendations (Zheng, Ghane, Sabouri, 2019) for the course project recommendations.

In the traditional recommender systems, the receiver of the recommendations is the only stakeholder in the environment. Recently, researchers argue that the perspective of other stakeholders may be also important (Burke, 2016; Zheng, 2017) to be considered in the recommendation process. We made the first attempt to develop multi-stakeholder recommendation methods for educational learning (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a) in order to balance the needs of both instructors and students. More specifically, in the context of course project recommendations, some students may try to select easier projects while the instructors suggest students to work on more challenging projects. Therefore, a multi-stakeholder course project recommendation model is necessary to be built in order to balance the needs of students and instructors.

We have successfully applied the utility-based multi-stakeholder recommendation approaches (Zheng, 2018) to the context of course project recommendations as. In addition, we provide solutions to alleviate the over-/under-expectations (Zheng, Ghane, Sabouri, 2019) and help better capture the different perceptions of the students and instructors (Zheng, 2019a) in our existing work respectively. In this paper, we try to discover the impact of student communities in the multi-stakeholder educational recommendations by creating student communities based on clustering student information.

2. Educational Data and Methodologies

We use the same educational data as the one in our previous work (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a). Each student should select at least three liked and disliked topics of the course, and provide an overall rating to them. In addition, students were asked to rate each selected project on three criteria: how interesting the application area is (App), how convenient the data processing will be (Data), how easy the whole project is (Ease). Both the overall rating and the multi-criteria ratings are in scale 1 to 5. There is a total of 3,306 rating entries given by 269 students on 70 Kaggle data sets. Each rating entry is associated with both overall and multi-criteria ratings. In addition, we have collected the

demographic information (i.e., age, gender, nationality) and personality traits represented in five personality dimensions (McCrae, 1992) of the students.

The idea behind the frame work of the utility-based multi-stakeholder recommendations is that the multi-objective learning techniques can be applied to balance the needs of different stakeholders, while we need to build the utility function for each stakeholder in the system. There are different ways to build the utility function. We first claimed that the multi-criteria ratings can be utilized to build these functions (Zheng, 2017). The utility-based multi-stakeholder educational recommendation models have been proposed accordingly (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a).

The utility function in these systems is represented by the similarity between a rating vector on an item, and an expectation vector. These two vectors are mapped to the multiple criteria in the system. In our data set, we have three criteria – App, Data and Ease. An expectation vector by the student (i.e., student expectations) indicates which items they may prefer in terms of the three criteria. By contrast, the expectation vector of the instructor (i.e., instructor expectations) is defined as the instructor’s minimal requirements in the system. Both students and instructors will give multi-criteria ratings to the items. Note that the ratings by students can tell student preferences, while the ratings given by instructors indicate how appropriate the item can be selected as projects from the perspective of the instructors. Therefore, the utility of an item t from the perspective of a student s ($U_{s,t}$) can be represented by the similarity of student expectations and the rating vector on the item by the student. By contrast, the utility of an item t from the perspective of an instructor p ($U_{p,t}$), can be represented by the dissimilarity between the expectation and evaluation rating vectors based on the “Data” and “Ease” dimensions. We can calculate a *utility score* which is a combination of the item utility from the perspective of students and instructors, i.e., $utility\ score = \alpha \times U_{s,t} + (1 - \alpha) \times U_{p,t}$, while α ($0 < \alpha < 1$) is the weight factor. The optimal value of α may not be 0.5, due to the different distribution of $U_{s,t}$ and $U_{p,t}$. We use the multi-objective learning technique to learn the optimal value of α . We collected the instructor expectations in advance, since there is only one instructor. We can learn the student expectations in advance by the utility-based multi-criteria recommendation models (UBRec) (Zheng, 2019b), or we can learn them in the process of multi-objective learning. In this paper, we decide to learn them in advance. An evaluation metric “score” which is the difference between utility gain and utility loss. For more details, please refer to our previous work (Zheng, 2019a).

In this paper, we would like to discover the impact of student communities in the multi-stakeholder educational recommendations. More specifically, we can update the *utility score* to be $\alpha \times (\beta \times U_{s,t} + (1 - \beta) \times U_{c,t}) + (1 - \alpha) \times U_{p,t}$, while β is another weight factor in $[0, 1]$. In comparison with the previous utility score, we use a linear combination of $U_{s,t}$ and $U_{c,t}$, where we use c to represent the student community and $U_{c,t}$ therefore as the utility of the item from the perspective of community c . We decide to use X-Means (Pelleg, 2000) to cluster students into different communities, while we can utilize the following feature sets for the purpose of clustering -- demographic information (F_d), personality traits (F_p) and ratings on the items (F_r).

3. Experiments and Results

We use the same setting in our previous work (Zheng, 2019a), e.g., we use 5-fold cross validation, and the ϵ -MOEA as the multi-objective optimizer in the MOEA library¹. We use three baseline approaches – UBRec (Zheng, 2019b) which only considers the student preferences, SolBasic which is the basic multi-stakeholder solution (Zheng, 2018) and SolCorr which is the improved multi-stakeholder solution which considers preference corrections (Zheng, 2019a).

The experimental results can be observed from Figure 1. The basic multi-stakeholder solution SolBasic failed to beat UBRec which is a traditional recommendation model. Once we applied the preference corrections (Zheng, 2019a), SolCorr is able to perform the best among these baselines. Other solutions are the ones that we incorporate the student communities. We can observe that all of these solutions are able to beat the SolBasic. By using F_d , F_p and F_r respectively, we can observe that F_r is the most effective feature set, while the performance based on clusters using F_d performs the worst.

When we combine the feature sets together to produce clusters, the experimental results vary. The performance was even decreased, when we combined F_d with other feature sets. For example, the

¹ MOEA, <http://moeaframework.org>

score by using F_d+F_p is lower than the one using F_p only. It infers that F_d is a set of noisy features. A closer look in F_d can reveal that most of the students are from a same country, and most of the students are in the same age group. That's probably the reason why F_d seems to be not useful in our experiments. The combination of F_p and F_r turns out to be the optimal feature set to create the clusters if we combine two feature sets only. However, the score is still lower than the SolCorr. The best solution is the F_p+F_r+Corr in which we utilize F_p and F_r to create student communities, and also apply the preference corrections (Zheng, 2019a). In comparison with SolCorr, the score was improved by around 2%.

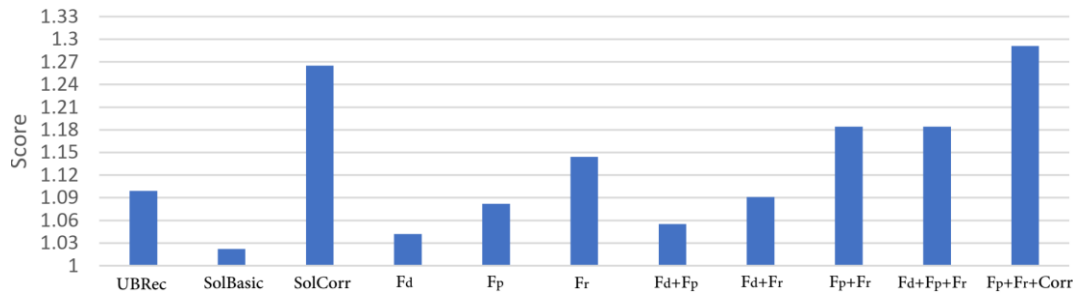


Figure 1. Experimental Results

Recall that β is used to fuse the utility of the item from the perspective of students and the communities. In most of the proposed solutions, the value of β varies from 0.6 to 0.8, while the β is 0.72 in the best solution F_p+F_r+Corr . The contribution by the student communities is 28%, which confirms the effective impact by the student communities.

4. Conclusions and Future Work

In this paper, we found that the student communities could be able to improve the multi-stakeholder educational recommendations, if we use the appropriate feature sets to create the student communities. In our future work, we plan to apply A/B test to further examine the effectiveness of these models.

References

- Burke, R. D., Abdollahpouri, H., Mobasher, B., & Gupta, T. (2016). Towards Multi-Stakeholder Utility Evaluation of Recommender Systems. In *ACM UMAP (Extended Proceedings)*.
- Zheng, Y. (2017). Multi-stakeholder recommendation: Applications and challenges. arXiv preprint arXiv:1707.08913.
- Zheng, Y., Ghane, N., & Sabouri, M. (2019). Personalized Educational Learning with Multi-Stakeholder Optimizations. In *Adjunct Proceedings of the ACM UMAP*. ACM.
- Zheng, Y. (2019). Multi-Stakeholder Personalized Learning with Preference Corrections. In *Proceedings of the 18th IEEE International Conference on Advanced Learning Technologies (ICALT)*. IEEE.
- Zheng, Y., & Pu, A. (2018). Utility-Based Multi-Stakeholder Recommendations By Multi-Objective Optimization. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 128-135). IEEE.
- Pera, M. S., & Ng, Y. K. (2013). What to read next?: making personalized book recommendations for K-12 users. In *Proceedings of the 7th ACM Conference on Recommender Systems* (pp. 113-120). ACM.
- Burke, R., Zheng, Y., & Riley, S. (2011). Experience Discovery: hybrid recommendation of student activities using social network data. In *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems* (pp. 49-52). ACM.
- He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th international conference on World wide web* (pp. 421-430). ACM.
- Zheng, Y., & Subramaniyan, A. (2019). Personality-Aware Collaborative Learning: Models and Explanations. In *International Conference on Advanced Information Networking and Applications* (pp. 631-642). Springer.
- Zheng, Y. (2019). Utility-based multi-criteria recommender systems. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 2529-2531). ACM.
- Pelleg, D., & Moore, A. W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML* (Vol. 1, pp. 727-734).
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2), 175-215.

An Investigation of a Medical Terminology Learning Environment with a Robot and a Tablet

Yoko SAKAMOTO^{a*} & Nobuhiro SAKATA^b

^a*Division of Languages and Humanities (English), Dokkyo Medical University, Japan*

^b*Center for Information & Communication Technology, Dokkyo Medical University, Japan*

*y-saka@dokkyomed.ac.jp

Abstract: In this study, we have examined a medical terminology learning environment with a humanoid communicative robot and a tablet. In the environment with the robot participants can see the whole body of the robot and the tablet. In the tablet only environment the body of the robot was covered with white imitation paper. Total fifteen pairs of English medical terms were selected with the same number of syllables and the same stress pattern with similar sounds. Participants were divided into two groups. One group studied the medical terms with the robot while the other group did with the tablet. In the second learning session conducted after two weeks the participants studied with the another environment. During the session quizzes of the medical terms were given before and after the study. The average scores of the quizzes improved to a similar extent in the both robot and tablet groups. However, in the memory retention test, which was taken two weeks later of the study, showed that the average score was 4.3 in the environment with a robot although 2.4 in the tablet only environment. After the investigation more participants suggested that learning with a robot was fun and could be helpful in memory retention. In fact, we observed that many of participants looked at the face of robot during the study with a robot. We think that the presence of the robot might affect psychological environment. In conclusion, the observation suggested that retention of medical terms might be affected by the learning environment with a robot.

Keywords: Medical terminology, robot, tablet, language learning

1. Introduction

Humanoid communicative robots have been reported to provide effective support in caring for autistic children (Kojima et al. 2008). Ishiguro et al (2016) also showed a new method of using robots for patient education. These previous studies suggest the simplicity of communication with a robot can be helpful to support who have anxiety or need care. When we think of the second language education, language anxiety and social anxiety can be a factor to reduce the learning effect. Thus, we have developed the robot learning environment for medical students. In this study, we aimed to investigate the differences between learning English medical terms with a robot or a tablet.

2. Materials and Methods

2.1 Materials

Fifteen medical terms were presented via a humanoid communicative robot or a tablet.

2.1.1 Medical Terminology

Fifteen pairs of English medical terms were used as stimulus words. Each pair was chosen with the same number of syllables, the same stress pattern, and where possible, similar sounds. For example,

adenoma and trachoma, peritoneal and peritonitis, hypoxia and hypoplasia etc. The left of each pair was presented via a robot and the right was presented via a tablet.

2.1.2 Communicative Robot Environment

A communicative robot (“Pepper”) produced by Softbank Robotics (Tokyo, Japan) was used in this investigation (Figure 1). This robot can recognize and speak English and can show pictures and letters on a 10.1-inch touch-panel display (tablet). The height of this robot is about 120 cm. The robot makes gestures such as nodding and moving its arms. A software development kit (Choregraphe) was used to develop the medical term learning materials. The robot asks the Japanese meaning of 15 medical terms orally and students can answer by touching the tablet, which shows four alternative responses. After checking the meaning, the robot asks the students to pronounce the word with a robot.

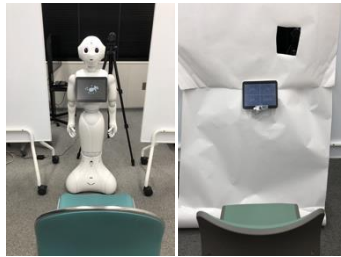


Figure 1. The learning environments with a communicative robot (left) and a tablet only (right)

2.1.3 Tablet only Environment

In order to make similar condition with a robot environment, the body of the robot was covered with white imitation paper (Figure 1). The content of the tablet included the same learning procedure with the robot. The tablet asks the Japanese meaning of 15 medical terms with sounds and students can answer by touching the display, which shows four alternative responses. After checking the meaning of the medical term in Japanese, the tablet asks the students to pronounce the word with a tablet.

2.1.4 Participants

Total 22 first grade medical students participated in this investigation. They were divided into two groups. Half of the participants firstly learned with a robot and then learned with a tablet two weeks later, while the other half of them learned with a tablet first and then with a robot two weeks later.

2.2 Methods

Each student participated in the trial in a separate room. The trial consisted of three steps, including a pre-trial questionnaire and a pre-trial word quiz, robot learning or tablet learning, a post-trial questionnaire and a post-trial word quiz. Two weeks later, the participants were counterbalanced and firstly they took the same word quiz as two weeks before to be checked the memory retention of the medical terms and followed the five steps.

Step 1: In the pre-trial questionnaire, students were asked whether they had learned with a robot or a tablet. The participants did a pre-trial word quiz, which asked the students to write in Japanese the meaning of the medical terms to be learned with a robot or a tablet.

Step 2: The participants learned with the robot or the tablet to check 15 English medical terms presented as a game. First, the robot or the tablet asked a question orally and the students answered by touching a display that showed four alternative answers. When the student answered correctly, the robot or the tablet asked the participant to pronounce the medical word. After the robot or the tablet recognized the word the student pronounced, the next question was presented.

Step 3: In post-trial questionnaire, students were asked whether learning with a robot or a tablet was helpful to remember English medical terms and whether the learning was fun. The students also did a post-trial word quiz, which asked the students to write in Japanese the meaning of the medical terms which was learned with a robot or a tablet.

3. Results

We checked the word quiz and put one point per one correct answer. Table 1 and 2 are the average score of the word quiz for two subject groups.

Table 1

The Average Score of the Word Quiz of the First Session of the Investigation

	Subject Group 1	Subject Group 2
	Robot trial	Tablet trial
Pre-trial word quiz	2.2	1.1
Post-trial word quiz	6.4	5.1

Table 2

The Average Score of the Word Quiz of the Second Session Conducted Two Weeks Later

	Subject Group 1	Subject Group 2
Memory retention quiz	4.3 (Robot)	2.4 (Tablet)
	Tablet trial	Robot trial
Pre-trial word quiz	1.6	1.8
Post-trial word quiz	5.8	8.2

The results of the word quiz showed that the average score of the robot learning was almost the same with the average score of the tablet learning on the first session, but two weeks later, the result of the memory retention quiz showed that the learning with the robot was slightly higher in the average score by two points than the average score of the tablet learning. Moreover, Subject Group 2 showed better average score after learning with the robot (8.2) compared to the tablet trial group (5.8).

From the post-trial questionnaire more participants mentioned that learning with a robot was fun and could be helpful in memory retention. We also observed that many of participants looked at the face of robot during the study with a robot.

4. Discussion and Conclusion

These results have suggested that a learning environment for medical terms with a robot or with a tablet might have some differences. We think that the presence of the robot might be important in the learning environment. In conclusion, from the point of view of the average scores after each session, we observed that there were similar learning effects when using a robot and tablet. However, retention of vocabulary might be affected by the learning environment with a robot.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16K02888 and Dokkyo Medical University Grant Number 2017-08. We would like to thank all participants of the trial at Dokkyo Medical University.

References

- Kojima, H, Nakagawa K, Yasuda Y. (2008). Teaching autistic children via communication with a robot. *Information Processing* **49**(1), 36-42. In Japanese.
- Ishiguro, K, Majima Y, Sakata, N. (2016). Deployment of ARCS Model and Utilization of Communication Robot in Patient Education. *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies* (BIOSTEC 2016) 5: HEALTHINF, 371-376.

Reconstruction of Concept Map to Promote Learners' Comprehension on New Knowledge

Didik Dwi PRASETYA^{ab*}, Triyanna WIDIYANINGTYAS^b, Tsukasa HIRASHIMA^a, & Yusuke HAYASHI^a

^a*Department of Information Engineering, Hiroshima University, Japan*

^b*Department of Electrical Engineering, Universitas Negeri Malang, Indonesia*

*didikdwi@um.ac.id, didik@lel.hiroshima-u.ac.jp

Abstract: This study investigated a new extension of reconstruction Kit-Build (KB) concept map, which combined with the open-ended style that calls KB-Mixed. To confirm the learners' comprehension of new knowledge, we compared KB-Mixed and open-ended concept map. This study involved 55 2nd-grade university students who were divided into two groups, experimental and control. Students in the experimental group requested to construct concept maps using the KB-Mixed method, while those in the control group using the open-ended concept map, which calls Extended Scratch-Build. The results indicated that KB-Mixed had a significant impact on learners' comprehend, which was represented by the post-test score. Moreover, the concept maps score assessed by expert judgment showed a significant correlation difference to the post-test score.

Keywords: Concept map, reconstruction, comprehension, new knowledge

1. Introduction

Concept maps are graphical tools for teaching, learning, organizing, and representing knowledge introduced by Novak and Gowin (1984) based on the meaningful theory offered by Ausubel (1963). Kit-Build (KB) is a closed-ended style concept map proposed by Hirashima (2015) aimed to support automatic assessment. In a KB, a learner is provided to nodes and links that are composed of teacher's concept map, and the learner is required to make a concept map by combining them. KB consist of three main stages: (1) A teacher creates a concept map that will become a goal map; (2) The KB system will decompose the goal map into concepts and links called "kit"; and (3) The students will be asked to reconstruct the concept map from the provided kit. Therefore, KB map requests a learner to reconstruct the original map by using provided components, and it can be called "reconstructional concept map" (Hirashima, 2018). Some previous studies have revealed many positive effects on the KB concept map to improve learning outcomes (Pailai et al., 2017; Hirashima, 2018), both individually and in groups. However, as a closed-ended concept map framework, KB considers less facilitate knowledge structure comparing to the open-ended concept map.

This study focused on a new extension of KB concept map, which is integrating the open-ended method that calls KB-Mixed. There are two phases of concept map construction as a unity, which is the main characteristic of KB-Mixed to increase meaningful learning and promote learners' knowledge building. To confirm the learners' comprehension of new knowledge, we investigated KB-Mixed and comparing with an open-ended style, which calls the Extended Scratch-Build concept map. In this comparison, we focus on a concept map that provides extensibility through two-phase construction. The present study hypothesizes that experimental groups that utilize KB-Mixed will promote learner's comprehension of new knowledge.

2. Methods

Participants in this study were 2nd-grade university students majoring of informatics engineering. Before determining the group, the lecturer gave a pre-test to ensure that the two groups are

homogeneous ($p = .736 > 0.05$). Fifty-five students were divided into two groups randomly, which consists of 27 students for the control group and 28 students for the experimental group. Students in the control group were requested to make concept maps using Extended Scratch-Build approach, while the experimental group using KB-Mixed.

This study was conducted in the “Basis Data 1” (Database 1) subject which was delivered in Indonesian. We involved an experienced lecturer who is taught this material to conduct theoretical learning in both control and experiment classes. The control and experimental group used the same learning environment, including classrooms, personal computers specifications, and Internet connection. Before the experiment was conducted, at the previous course meeting, participants had been given an introduction to concept maps. Furthermore, participants were also instructed to make concept maps on the introduction of a database topic. This experiment used map construction design twice (phase 1 and phase 2) in one lecture. The main difference between the control group and the experimental group lies in creating the concept map in phase 1, where the control group used the open-ended approach, while the experimental group utilized the KB system. Next, in phase 2, the two groups equally extended the previous concept map by adding new concepts and links.

This study involved two measurements to identify learners' performances: (1) pre-test and post-test; and (2) concept map score. The pre-test was designed to examine whether students in the control group and the experimental group had equivalent knowledge regarding related instructional design. Pre-test and post-test design used the same multiple-choice questions, where the post-test is randomly presented. Pre-test and post-test evaluations were carried out by a class teacher because this study was in class experimental. The teacher who conducted the assessment here was also a teacher who taught in class, a senior teacher with more than ten years of teaching experience in database subjects. The assessment of the concept map is based on propositions by adopting the quality rating of propositions proposed by Osmundson (1999). Propositions scores ranged from inappropriate or incorrect connections (score = 0 point) to most scientific understanding (score = 3 points). There were two midpoints represent practical understanding (score = 1 point) and the scientific understanding of but has limited explanatory power (score = 2 points).

3. Results and Discussion

Post-test was used to measuring learners' understanding after getting an intervention, whether there is an increase or not. The results of the post-test of both control and experimental groups were analyzed by the Mann-Whitney U test and are shown in Table 1. According to the table data, it can be seen that can see that in the mean rank and sum of ranks items experimental group had the highest score. The results revealed a statistically significant difference in learners' learning performance between the control group and the experimental group after the intervention ($U = 259.0, p = .030 < 0.05$).

Table 1

The Mann Whitney-U results of the post-test for both groups

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Control group	27	23.59	637.00	259.0	-2.164	.030
Experimental group	28	32.25	903.00			

The second measurement which conducted here was students' learning performance related to creating concept maps. The same teacher evaluated the concept map based on the rubric that was set at the beginning. This study also employed Mann-Whitney U statistical analysis to identify map score in both groups. Table 2 show that the experimental group outperformed control group with the significant difference ($U = 75.0, p = .000 < 0.05$).

Table 2

The Mann Whitney-U results of the map score for both groups

Groups	N	Mean Rank	Sum of Ranks	U	Z	p
Control group	27	16.78	453.00	75.0	-5.104	.000
Experimental group	28	38.82	1087.00			

In this study, we also examined whether there was a correlation between the post-test score and map score on the experimental group and the control group. The statistical correlation analysis was performed using Spearman bivariate correlation test. The test results state that there was a significant correlation difference between the post-test score and map score ($p = .018 < 0.05$) with a weak positive correlation coefficient ($r = .318$). It can mean that the use of approaches in creating concept maps can affect student learning outcomes.

4. Conclusion

The present study provided a new extension of the reconstruction Kit-Build concept map to promote learners' comprehension of new knowledge. The findings revealed that the experimental group outperformed the control group in terms of students' achievement indicated by the post-test score. In the results of creating a concept map, the experimental group that used KB-Mixed approach also score better than the control group which used extended scratch-build concept map. Correlation test between the concept map score and post-test score also shows a significant correlation, and emphasize that the use of concept maps methods has a positive effect on learning outcomes.

This study also states that the experimental group with the KB-Mixed concept map approach facilitates learners' comprehension of new knowledge based on their prior knowledge structure. Although the present study did not focus on knowledge building, the concept map results show that KB-Mixed had an excellent potential to facilitate learners' knowledge structure. The stage offered by KB-Mixed, which consists of two interrelated stages, can present meaningful learning well. However, studies of knowledge building and structural thinking still need to be explored more deeply. Therefore, for future work, maybe it would be more appropriate if we involved an analysis of knowledge building and structural thinking.

Acknowledgments

I would like to grateful for the financial support from the Islamic Development Bank (IsDB) for PhD scholarship in partner with Universitas Negeri Malang (UM), Malang, East Java, Indonesia. Also great thanks to my sensei in the Department of Information Engineering, Hiroshima University, Japan.

References

- Ausubel, D. P. (1963). *The Psychology of Meaningful Verbal Learning*. New York, Grune and Stratton.
- Hirashima, T., Yamasaki, K., Fukuda, H., & Funaoi, H. (2015). Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, 10(1), 17.
- Hirashima, T. (2018). Reconstructional concept map: automatic Assessment and reciprocal reconstruction. *International Conference on Learning Innovation (ICLI)*.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.
- Osmundson, E., Chung, G. K. W. K., Herl, H. E., & Klein, D. C. (1999). Knowledge mapping in the classroom: A tool for examining the development of students' conceptual understandings. Report: CSE-TR-507.
- Pailai, J., Wunnasri, W., Yoshida, K., Hayashi, Y., & Hirashima, T. (2017). The practical use of Kit-Build concept map on formative assessment. *Research and Practice in Technology Enhanced Learning*, 12(1), 20.

Beyond Just Following Data: How Does Visualization Strategy Facilitate Learning Analytics Design?

Jiixin CAO^{a*} & Yanjie SONG^a

^a *Department of Mathematics and Information Technology, The Education University of Hong Kong, *jxcao@s.edu.hk*

Abstract: In this poster, we reviewed 38 articles on learning analytics research, focusing on the data visualization interface designs. After examining the original ideas their interface design, a new *visualization strategy* was proposed to categorize and characterize them premised on their principles and approaches in four types, namely, (1) directly-presented, (2) outcome-oriented, (3) process-oriented, and (4) theory-oriented. Then, how these types of the visualization strategy could help facilitate learning analytics design and make data-interpretable by users was presented.

Keywords: Visualization strategy, Learning dashboard, Design principle, Learning analytics

1. Introduction

Learning analytics is defined by the Society for Learning Analytics Research (SoLAR) as “the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which learning occurs (Siemens, 2010, para. 6)”. In the past, the learning analytics community focused more on how to make use of educational datasets and identify potential information in learning contexts, but less on how the analysis results could make sense to users through different types of the visualization design. This is why we need to examine the original ideas of visualization design in the learning analytics tool. As a result, we proposed a new *visualization strategy* to categorize and characterize the principles and approaches for the visualization design. In this paper, we reviewed 38 articles containing interface designs of visualization in learning analytics research among 2010 and 2019 to identify the visualization design strategies of existed tools.

2. Related literature

Data visualization is a paradigm tool to support the judgment, inference, and decision-making in learning analytics (Alhadad, 2018). The design of visualization plays roles in (1) executing data interpretation into understandable and readable dimensions (Dzemyda, Kurasova, & Žilinskas, 2013), (2) incorporating relevant visualizations to support awareness, self-reflection, and sense-making of stakeholders (Ochoa, Suthers, Verbert, & Duval, 2014), and (3) organizing factors to support and transfer actionable decisions (Sacha et al., 2017). Thus, the design of visualization in learning analytics could influence the effect of learning analytics to gain insights into educational issues.

When a learning analytics tool is embedded in a learning system/platform for educational use, the data visualization is commonly employed as a bridge to demonstrate analytical results to general users and make educational data interpretable by them. However, many features included in the current visualization design have failed to meet their purposes as they claimed. Some studies have found that the majority of learning dashboards, and standard forms of the visualization in learning analytics, consist of a basic pie chart, line graph or scatterplots without actionable information for teaching and learning (Schwendimann et al., 2017).

Learning analytics research uses data to inform decisions for a diverse set of educational stakeholders (e.g., learners, instructors, and administrators), which is a significant difference comparing to traditional educational research (Dragan, Shane, & Abelardo, 2016). If having no actual impact, learning analytics development might be reduced to "simplistic rhetoric of quick technological fixes" (Dragan et al., 2016, p. 1). How to design visualization to facilitate wider stakeholders rather than the statistical experts is crucial to achieving the expected impact on implementing learning analytics in practice. However, currently, there is no clear picture of visualization designs adopted in current learning analytics research. This review study aims to solve the issue.

3. Methods

The searching range of the literature review was set from 2010 to 2019 using the keywords "visualization", and "learning dashboard" because "learning dashboard" is a popular visualizing tool in current literature. One hundred fifty-nine articles were found in the areas of education, machine learning or other domains. Then, we used the keywords "educational", or "education" to filter the articles. Finally, 38 articles were identified, which contain the description of visualization or learning analytics dashboards in detail.

Open coding was adopted to categorize and characterize the visualization strategies in the selected learning analytics designs. Three themes were identified: (1) their selection of indicators for demonstration, (2) their cognitive approaches intended by a viewport, and (3) the forms of visual content organization on the user interface (mainly dashboards).

4. Results

This study distilled four types of visualization strategies in reviewed cases and listed in Figure 1. They are (1) directly-presented, (2) outcome-oriented, 3) process-oriented, and 4) theory-oriented.

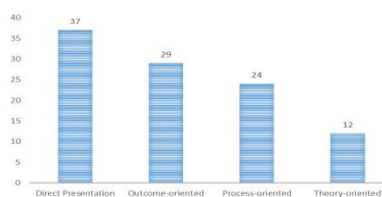


Figure 1. Types of visualization strategies in reviewed cases.

The directly-presented data is the fundamental realization of learning analytics visualization, which presents some scattered datasets from sensors or system logs that might contain educational concerns. Designers adopt statistical graphs or tables to outline data distribution in its original form. Users need to interpret data by clustering and regression from scattered table or plot to get desired information. By doing so, less information is lost.

The outcome-oriented visualization strategies focus on checkpoints, such as the score of quiz or the completion rate. Some critical performance could be identified by comparing results with standards, history, or peers through scale bars or scores. This type of strategies highlights information like achieving some milestones or grades but ignores the meaning of changes in progress.

The process-oriented visualization strategies attempt at representing learning trails. Such kind of visual designs makes indicators continuous in forms of the path and able to "paint out" the footprints in the learning process. By tracking the progress, users could find out and reflect on their personal learning experiences better.

However, stakeholders need a certain level of statistical skills to understand and interpret some of the artifacts generated from the three strategies mentioned above, such as clustering and regression. Thus, it remains a barrier for users to make sense and interpret the information in a certain background.

The theory-oriented visualization strategies align educational theories with organizing both selected and continuous indicators to address educational concerns holistically. The indicators adopted in the type of strategies are presented in a hierarchy, which is aligned with educational theories, such as

five phases in inquiry-based learning. An educational theory generally recognizes the nature of a specific setting and organizes key checkpoints and processes, so that interpretation of the indicators in context is more insightful and for educational decision-making.

The characteristics of four types of visualization strategies in learning analytics are also presented in Table 1. In order to make the audience understand more intuitively, the demonstration models of different types of visualization strategies in learning analytics is presented in Figure 2.

Table 1
Characteristics of Types of Visualization Strategies in Learning Analytics

Strategy	Directly-presented	Outcome-oriented	Process-oriented	Theory-oriented
Indicator	distributed	selected	continuous	hierarchical
Focus	phenomenon /behavior	checkpoint/result	trail	explanation of potential insight
Approach	cluster/regression	compare	track	interpret
Possible form	scatted table/plot	scale / score	trend / path	theory-based feedback

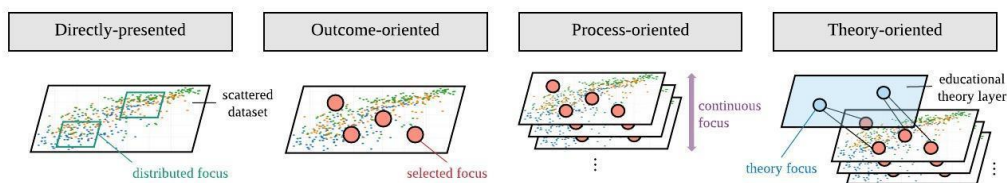


Figure 2. Demonstration models of different types of visualization strategies.

5. Limitation and future work

A primary limitation of this poster is that the categorized strategies were classified only based on the simple standards according to the user interface without in-depth analysis, to gain profound insights into their design principles. In a further study, we are going to identify a holistic framework of visualization design and presentation to deepen the understanding of orientations included in different types of strategies.

References

- Alhadad, S. S. (2018). Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science. *Journal of Learning Analytics*, 5(2), 60-85.
- Dragan, G., Shane, D., & Abelardo, P. (2016). How do we start? State and Directions of Learning Analytics Adoption: International Council for Open and Distance Education.
- Dzemyda, G., Kurasova, O., & Žilinskas, J. (2013). Multidimensional Data and the Concept of Visualization *Multidimensional Data Visualization: Methods and Applications* (pp. 1-4). New York, NY: Springer New York.
- Ochoa, X., Suthers, D., Verbert, K., & Duval, E. (2014). Analysis and reflections on the third Learning Analytics and Knowledge Conference (LAK 2013). *Journal of Learning Analytics*, 1(2), 5-22.
- Sacha, D., Sedlmair, M., Zhang, L., Lee, J. A., Peltonen, J., Weiskopf, D., . . . Keim, D. A. (2017). What you see is what you can change: Human-centered machine learning by interactive visualization. *Neurocomputing*, 268, 164-175.
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41.
- Siemens, G. (2010). *1st International conference on learning analytics and knowledge 2011*. <https://tekri.athabasca.ca/analytics/>

MAIN CONFERENCE C4
CLASSROOM, UBIQUITOUS, AND MOBILE TECHNOLOGIES ENHANCED
LEARNING (CUMTEL)

- FULL PAPER -

RESEARCH ON THE INFLUENCE OF ROBOT TEACHING ON THE CREATIVITY OF SECONDARY SCHOOL STUDENTS UNDER THE BACKGROUND OF STEM EDUCATION 377

JIANGXIANG ZHANG, FANGTING JIANG, CHU LIU

RESEARCH ON THE IMPACT OF E-SCHOOLBAG ON STUDENTS' DEVELOPMENT: FROM THE PERSPECTIVE OF PERSONAL LEARNING SKILLS..... 386

CHU LIU, JUEQI GUAN & YOUQUN REN

INFLUENCE OF FINANCIAL COURSE ON EIGHTH GRADE STUDENTS' FINANCIAL CONCEPTS, MATH MOTIVATION, MATH ANXIETY IN TAIWAN 394

YU-CHING HSU, MEI-WEN NIAN, CHANG-HSUAN YANG, YUAN-HSUAN LEE, JIUN-YU WU

GENETICUS INVESTIGATIO: A CLASSROOM-BASED TECHNOLOGY-ENHANCED LEARNING ENVIRONMENT FOR PROBLEM-SOLVING PROCESS SKILLS IN GENETICS 402

ANURAG DEEP, SAHANA MURTHY & JAYADEVA BHAT

AUGMENTED REALITY IN EDUCATION: THREE UNIQUE CHARACTERISTICS FROM A USER'S PERSPECTIVE 412

JULE M. KRÜGER, ALEXANDER BUCHHOLZ & DANIEL BODEMER

- SHORT PAPER -

ACTOR-NETWORK THEORY APPROACH USING M-LEARNING TECHNOLOGIES IN THE PUBLIC SENIOR HIGH SCHOOL AS PEDAGOGY 423

MAMERTO T. GONEDA, FRANCIS ARLANDO L. ATIENZA, JOHN PAUL P. MIRANDA, ROLANDO R. FAJARDO, DOMINIC C. CABAUATAN

THE PILOT IMPLEMENTATION USING AN ADAPTED TECHNOLOGY ACCEPTANCE MODEL TO EVALUATE AN INNOVATIVE USE OF SMARTPHONE FOR SCIENTIFIC INVESTIGATION PROGRAMME IN TERTIARY EDUCATION 429

SIEW WEI THO, TIEN TIEN LEE, SADIHA BAHAROM, FARIDAH LISA SUPIAN, NURUL SYAFIQAH YAP ABDULLAH & NURUL AINA SYAKIRAH ZAINAL ABIDIN

PROPOSAL FOR DEVICELESS LEARNING ENVIRONMENTS INSTEAD OF ENVIRONMENTS USING SMART DEVICES 435

KOZO MIZUTANI

A SEMANTIC TAG-BASED ENHANCED LEARNING RECOMMENDATION APPROACH FOR ENHANCING STUDENT LEARNING EXPERIENCES 441

TIEN-YU HSU, JUN-MING CHEN, HSIN-YI LIANG

DESIGNING PRE-TEST QUESTIONS AS PHONE NOTIFICATIONS: STUDYING THE EFFECTS OF A MOBILE LEARNING INTERVENTION. 447

INGRID YVONNE HERRAS, DON ROMIELITO N. ABANES, NICO B. DEL ROSARIO & JONATHAN DL. CASANO

MOOD: A MOBILE PHONE-ENABLED EDUCATIONAL DATA COLLECTION PLATFORM..... 453

PEI-YAN YUAN, A-MEI DU, YONG-BO LI, XIAO-YAN ZHAO, CHUAN WANG & XIAO-DONG WANG

ANALYSIS OF CURRENT SITUATION OF CLASSROOM INTERACTIVE TEACHING BASED ON MOBILE DEVICES: A CASE STUDY OF MIDDLE SCHOOL MATHEMATICS CLASSROOM TEACHING 460

JING-WEN PAN, DI-FEI WANG, MIAN LIU

MULTIMODAL RECORDING SYSTEM FOR COLLECTING FACIAL AND POSTURAL DATA IN A GROUP MEETING	466
YUSUKE SONEDA, YUKI MATSUDA, YUTAKA ARAKAWA & KEIICHI YASUMOTO	
STUDY OF AUGMENTED REALITY INTERACTION MEDIUMS (AIMS) TOWARDS COLLABORATIVELY SOLVING OPEN-ENDED PROBLEMS	472
PRATITI SARKAR, PRABODH SAKHARDANDE, UTSAV OZA & JAYESH PILLAI	
SUPPORTING JOB-HUNTING STUDENTS TO LEARN JOB-HUNTING RELATED TERMS WITH SCROLL EBOOK AND INCIRCLE	478
NORIKO UOSAKI, KOUSUKE MOURI , TAKAHIRO YONEKAWA, CHENGJIU YIN, & HIROAKI OGATA	
- POSTER PAPER -	
THE ADOPTION OF FACEBOOK MOBILE APPLICATION FOR MANAGING LEARNING	484
EDWIN THENG FUNG SIEW, & CHENG EAN (CATHERINE) LEE	
CHARACTERISTICS ANALYSIS FOR TECHNOLOGY ENHANCED LEARNING MATURITY: A QUALITATIVE APPROACH	487
AMALIA RAHMAH, HARRY B. SANTOSO & ZAINAL A. HASIBUAN	
INTEGRATION OF CLOUD-BASED MOBILE LEARNING TO IMPROVE STUDENTS' CREATIVE THINKING IN A VISUAL ARTS COURSE.....	490
JING-WEI LI, IWEN HUANG & HUI-YU JIANG	

Research on the Influence of Robot Teaching on the Creativity of Secondary School Students under the Background of STEM Education

Jiangxiang ZHANG^{a*}, Fangting JIANG^b, Chu LIU^c

^a *Department of Education Information Technology, East China Normal University, China*

^b *Chengdu Dongchen Foreign Language School, China*

^c *Department of Education Information Technology, East China Normal University, China*

* *zoe_jxiang@163.com*

Abstract: With the development of society, the requirements for the cultivation of learners are different. Nowadays, STEM education has become a boom in various countries, aiming to improve the comprehensive ability of learners in science, technology, engineering and mathematics. Robots are becoming an integral component of our society and have great potential in being utilized as an educational technology. Wonder B is an electronic module that can be freely spliced. This article uses Wonder B as the platform, combines with the actual learning situation of learners, tries to conduct teaching program design research, and aims to provides References for the teaching ideas and method of robot education. This study used the Williams Creativity Tendency Questionnaire to survey 40 Students and analyze s the data to explore the influence of robot teaching on their creative tendencies.

Keywords: STEM, robot teaching, primary and secondary students, creativity tendency

1. Introduction

The continuous implementation of curriculum reform have led to great changes to cultivate learning. In the 21st century, STEM education has got great attention in improving learns' comprehensive ability of science, technology, engineering and mathematics. The rise of Robot education provides an effective way for STEM education to cultivate computational thinking ability, interdisciplinary ability, problem solving ability and teamwork ability. The Robot education in primary and secondary schools in China has developed rapidly, and now it can be used as a carrier for basic education courses, and can also serve as a carrier for extracurricular activities in schools. Although the Robot education in China has developed for several years, there are many problems, such as lack of a standard curriculum, lack of product specifications to contest winners as the value orientation. Middle school is an important period for students to cultivate their ability meanwhile Scientific thinking ability has a strong guiding role in the process of student growth. Innovative ability is based on innovation thinking, which is the core strength of national competition and a strong support for economic development and social progress. Robots involve multiple subject areas in teaching practice of robot education. Recent study found that on the process of teaching robots has mat great challenges. A lot of robot teaching just stay in a superficial way which not only invalidated for connecting vertical communication with knowledge, but also increased the burden of students. Therefore, this study combined the learner's actual learning situation based on a platform called "Wonder B" and try to design the teaching project, aiming at providing reference for the teaching ideas and methods of the robot education curriculum under the background of STEM education.

2. Literature review

Robot education has always been a hot issue in foreign education research. The earliest educational robot came from Professor Papert of Massachusetts Institute of Technology who founded the

Laboratory of Artificial Intelligence in the 1960s. In 1994, the Massachusetts Institute of Technology established the “Design and Construction of LEGO Robot” course to improve the design and creativity of engineering design students. In recent years, Carnegie Robot Institute of the United States publicly released a few robot courses based on LEGO, VEX and other different educational robots. Such as Robot Science, ROBOTIC Intermediate Course, VEX Robot Course (Version 2.0), STEM-CAD Modeling Course, Electronics Course, etc. Foreign countries have also made many attempts in robot classroom teaching. In 2004, Mataric proposed that although robots seem to make a good teaching and learning tool, and have certain appeal to students of all ages, but in the application of educational robots, we should still pay attention to their teaching methods. (Mataric, 2004). In 2009, Alimisis and Kynigos suggested that the use of robot in a school environment should not be limited to focusing on technology, it should consider appropriate educational concepts (Alimisis & Kynigos, 2009). For example, Professor Eguchi listed three successful robot projects that show that the integration of STEM, programming and computational thinking can make students a future-adapted person. (Eguchi, 2014). Scaradozzi et al. believe that the introduction and use of robots as a course subject can teach children the technical basis and give them other types of life value, so they tracked the robot-themed and the entire student in the fifth grade of Italian primary school. During the fifth grade, it was found that children can demonstrate better learning skills, including technology and teamwork (Scaradozzi, Sorbi, Pedale, Valzano, & Vergine, 2015).

The robot educational in China has been gradually improved during the past 20 years. The earliest robot educational in China began in 1996, Dr. Wei Weimin took the lead in proposing the concept of educational robots internationally and created the first educational robot brand “Capability Storm which aims to cultivate learners’ technical literacy and scientific knowledge through educational robots. It was guided by robot projects and competitions which allows learners to “learning while doing” and “learning while having fun”. It can also construct a personal knowledge system to solve practical problems in life. Practice relevant domestic robot education of basic education can be traced back to 2000, Beijing Jingshan School has incorporated Robot Education into information technology curriculum in the form of scientific research projects, and has taken the lead in the teaching of robot in primary and secondary schools in China. (Zhang Lifang, 2015). In 2001, Shanghai Southwest Education School and Luwan High School began to explore and experiment with robot popularization education in the form of “school-based curriculum”. In September 2005, Harbin first officially introduced robots into classroom teaching in the city. 41 schools such as Teacher Attached and Provincial Experimental Middle School opened the course of “Artificial Intelligence and Robot” (Zhang Guomin & Zhang Jianping, 2008). In 2003, LEGO MINDSTORM EDUCATION was introduced to China to stimulate children’s technology interest and creativity. Lego robot education is no longer a boring science and technology popularization of knowledge or learning, it is an education that integrates programming principles, physical science, mathematical knowledge, modeling design and practical ability which combined spatial imagination and logical thinking in order to encourage the creativity and potential of learners with abundant teaching forms (Wang Xueyan & Yang Dongmei , 2017) . Of course, China has also issued a series of policy documents to promote the development of STEM education and robot education. Since 2003, the Ministry of Education has decided to experiment with robot education as a high school elective course in some provinces and cities. In 2004, the Ministry of Education and the Central Electrochemical Education Center also included "computer robots" as a competition project in the national primary and secondary school computer production activities. In 2017, the White Paper on STEM Education in China 2017, drafted by the STEM Education Research Center of the Chinese Academy of Sciences, emphasizes the promotion of the successful STEM education model (Wang Su, 2017) with the guiding principles of "collaboration, cooperation, openness, inclusiveness and innovation". In July 2017, the State Council issued the New Generation of Artificial Intelligence Development Plan, emphasizing the educational application of robots. (State Council, 2017).

3. Instructional design

The instructional design is based on the requirements of the curriculum standards and the characteristics of the teaching objects. To be more precise, instructional design is the process that teachers use modern teaching theories, comply with the characteristics of teaching objects and teachers' own teaching ideas,

experiences and styles, use systematic viewpoints and methods to analyze problems and needs in teaching, determine teaching goal, establish steps to solve problems, and rationally combine and arrange various teaching elements to optimize the teaching effect. The study designed related instructional design based on project-based learning (PBL).

3.1 Participants

The students who participated in the study came from the first-year students of a middle school in Shanghai. Through the test at the beginning of the semester and interviews with the students, it could be founded that they have certain computer application skills. Some students have certain programming skills. However, they weren't familiar with the mode of group contact and project-based learning, which means their self-management ability is poor.

3.2 Experiment procedures

A total of 40 students were recruited the experiment and data were analyzed to explore the influence of robot teaching on their creative tendencies. The teaching content of this project is to make students who had certain grammatical basis for programming to understand the application of full-color LED and display module , so that they can use the connection of hardware modules related to Wonder B and the compilation of code blocks to complete the project designed by the student team at the end of the learning phase.

3.2.1 Display module

Teaching mode	Teacher behavior	Student behavior
Participate	Use the display screen to guide the students to think about which module of the Wonder B hardware can achieve the display effect. The teacher explains the process of "display module"	Associate with the reality of life, discuss and answer relevant questions.
Inquiry (Autonomous research)	1. Through the renderings of 2-3 project examples, let the students guess the actual function of the pea spell display module. 2. After guessing the hardware features, let the students guess where the code area for each hardware function should be. 3. Combine the above considerations and summarize the use of the display module. Describe and summarize the problem on a personal basis.	Trying to understand the hardware features of the display module and the software code through 2-3 instances.
Explanation	Summarize the students' answers, standardize the functions of the display module and code writing, and guide students to understand the programming ideas from finding problems to solving problems.	In the teacher's standardized explanation, check the lack of traps, and promptly put forward record doubts.
Migrate	Take neon lights as an example, combined with the use of the display module, guess what modules are needed. Guide students to use the display module and full-color LED lights to achieve a dual effect of lighting and displaying text.	1. Associate with the reality of life, discuss and answer relevant questions; 2. Try to combine the full-color LED with the display module to complete the corresponding small exercises.

3.2.2 All-round LED lamp

<i>Teaching mode</i>	Teacher behavior	Student behavior
<i>Participate</i>	Use the display screen to guide the students to think about which module of the Wonder B hardware can achieve the display effect.	Associate with the reality of life, discuss and answer relevant questions
<i>Inquiry (Autonomous Research)</i>	The teacher explains the process of "display module" 1. Through the renderings of 2-3 project examples, let the students guess the actual function of the pea spell display module. 2. After guessing the hardware features, let the students guess where the code area for each hardware function should be. 3. Combine the above considerations and summarize the use of the display module. Describe and summarize the problem on a personal basis.	Trying to understand the hardware features of the display module and the software code through 2-3 instances.
<i>Explanation</i>	Summarize the students' answers, standardize the functions of the display module and code writing, and guide students to understand the programming ideas from finding problems to solving problems.	In the teacher's standardized explanation, check the lack of traps, and promptly put forward record doubts.
<i>Migrate</i>	take neon lights as an example, combined with the use of the display module, guess what modules are needed. Guide students to use the display module and full-color LED lights to achieve a dual effect of lighting and displaying text.	1. Associate with the reality of life, discuss and answer relevant questions; 2. Try to combine the full-color LED with the display module to complete the corresponding small exercises.
<i>Explanation</i>	Summarize the students' practice answers and strengthen students' understanding and use of the principle of RGB values.	Search for missing and make up for leaks in teachers' standardized explanations and put forward record doubts in time.
<i>Inquiry (Autonomous Research)</i>	1. Let the students complete the comprehensive exercises independently: flashing lights; 2. Answer questions according to the needs of students; 3. Observe the student's progress.	According to the previous knowledge, group work together to complete the corresponding exercises.
<i>Evaluation</i>	Design an open Q&A questions: Use Wonder B to design a program to solve problems in life or learning.	Students complete their own project design after class; Summarize the key points in the course and promptly feedback doubts.

3.2.3 Programming grammar rules

<i>Teaching mode</i>	Teacher behavior	Student behavior
<i>Participate</i>	Draw out the typical case from the last comprehensive exercise to guide the students to think about the cause of the error.	Discuss and answer questions based on previous exercises
<i>Explanation</i>	Summarize the students' answers, explain the comprehensive exercises, and summarize several types of errors.	Search for missing and make up for leaks in teachers' standardized explanations and put forward record doubts in time.
<i>Migrate</i>	Guide students to think about the grammar rules of programming by looking into the typical errors.	Discuss and answer questions based on previous exercises
<i>Explanation</i>	Summarize the students' answers and standardize the grammar rules of programming.	Search for missing and make up for leaks in teachers' standardized explanations and put forward record doubts in time.
<i>Inquiry</i> (<i>Autonomous Research</i>)	<ol style="list-style-type: none"> 1. Let the students complete the grammar rule corrections independently; 2. Answer questions according to the needs of students; 3. Observe the student's progress. 	<p>The group cooperated to finish the corresponding exercises by using the knowledge they just learned.</p> <ol style="list-style-type: none"> 1. Summarize the key points in the course, and timely feedback doubts. 2. Complete review and inspiration exercises 3. The students who have the ample force can continue to complete the copy exercise.
<i>Evaluation</i>	<ol style="list-style-type: none"> 1. Design a retrospective selective Q&A questions (review exercise): Examine the students' mastery of the knowledge they have learned; 2. Design open Q&A questions (inspired practice): Combine the reality of life and propose a problem that is difficult to solve. 	

3.3 Instruments

Creative tendencies cannot be measured directly by themselves, however, appropriate tools are often used to measure creative tendencies. This scale directly uses the “Williams Creative Tendency Test Table” developed by Williams and revised by Taiwan's Lin Xingtai and Wang Murong. The scale has 50 questions, including four dimensions as adventure, imagination, challenge and curiosity. The options are “Comply Compliant”, “Comparative”, “Uncertain”, “Comparatively Inconsistent” and “Completely Inconsistent”. Subjects make the choice base on their actual situation. The subjects answered according to their actual situation. The scores of the positive questions were 5, 4, 3, 2, 1, and the reverse questions were 5, 4, 3, 2 and 1 respectively. The total score was the total score of the Creativity Tendency Scale. The higher score indicated the stronger creativity tendencies, the lower score indicated the weaker creativity tendencies. The Cronbach’s alpha value of the questionnaire was 0.86.

Robot attitude questionnaire revised from "Computer Attitude Scale (Formal)" developed by Huang Shijie who proposed to divide computer attitude into computer anxiety, computer confidence, computer love, computer obsession, computer use value and computer equipment values.

4. Results

4.1 Descriptive statistics

The following *Figure 1 and Figure 2* are descriptive statistical tree diagrams of the data before and after the adventurous dimension. There are 11 questions in the dimension, and the total score is 55 points. It can be seen from the comparison between the two figures that the students have an anxiety degree before the course of 29, followed by 44, but most of the scores of the pre-test are distributed between 29-30 points, and most of the scores after the test are distributed between 42-44 points.

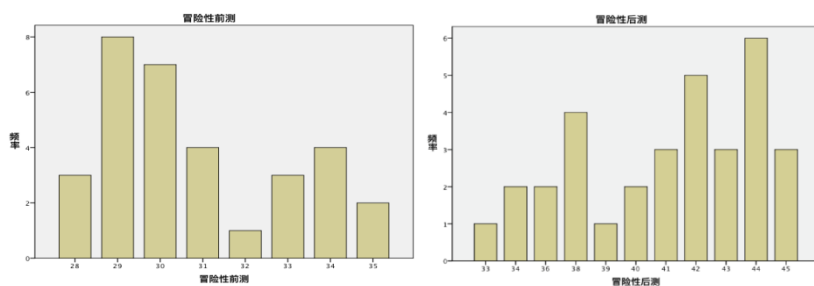


Figure 1 and Figure 2

The following *Figure 3 and Figure 4* are descriptive statistical tree diagrams of the data before and after the curiosity dimension. The total number of dimensions is 14 points, and the total score is 70 points. It can be seen from the comparison between the two figures that the students have an anxiety degree of 36 and 37 before the course. After that, it was 48, but most of the scores were distributed between 35 and 37 points. Most of the scores were distributed between 48 and 55 points.

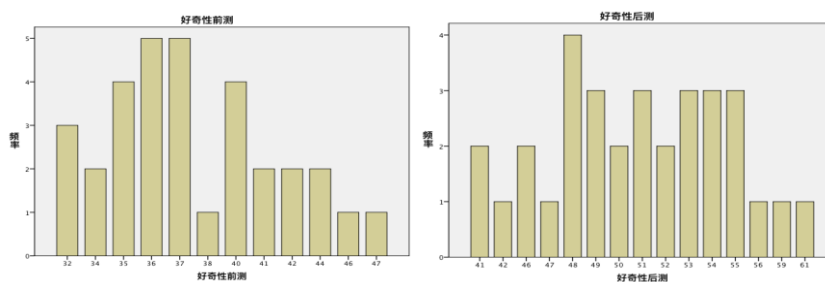


Figure 3 and Figure 4

The following *Figure 5 and Figure 6* are descriptive statistical tree diagrams of the data before and after the imagination dimension. There are 13 questions in the dimension, with a total score of 65. It can be seen from the comparison between the two figures that the students have an anxiety level of 34 before the course, followed by 49, but most of the scores of the pre-test are distributed between 33-34 points, and most of the scores after the test are distributed between 48-50 points.

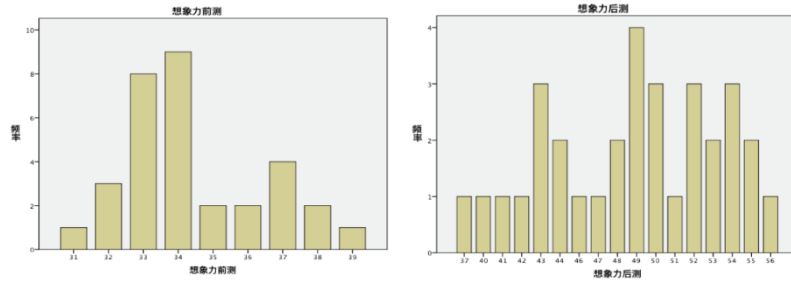


Figure 5 and Figure 6

Figure 7 and Figure 8 below are descriptive statistical tree diagrams of the measured dimensions before and after the challenge. The total score is 12 points, with a total score of 60. It can be seen from the comparison between the two figures that the students have an anxiety degree of 32 before the course, 37, 46 and 47, but most of the scores of the pre-test are distributed between 32-36 points, and most of the scores after the test are distributed between 44-48 points.

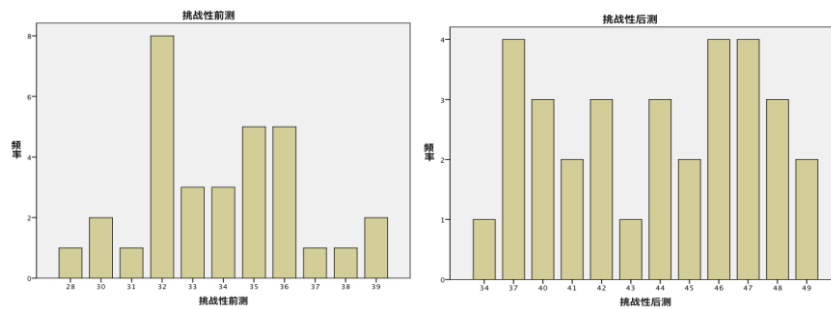


Figure 7 and Figure 8

Figure 9 and Figure 10 below are descriptive statistical tree diagrams of the pre- and post-test data of the Student Creativity Tendency Scale. It can be seen from the descriptive statistical bar graphs measured before and after. The pre-test data is mostly concentrated in 130-140. Between the latter and the measured data, most of them are concentrated between 160 and 200. There has been a significant change in the concentration trend, and the post-test scores tend to be concentrated and stable.

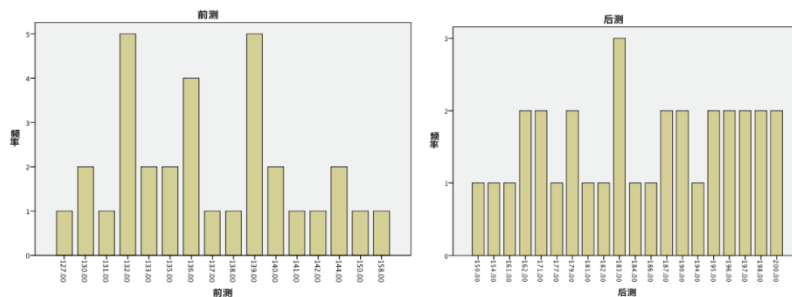


Figure 9 and Figure 10

4.2 Independent sample T test

As can be seen from Table 1 below, the levels of adventurous, imagination, challenge and curiosity are significantly different before and after the robot course in the STEM context.

Table 1

Creativity tends to each dimension independent sample T test

Dimension		S.D.	S.D.	t
Adventurous	Pretest	30.84	2.172	.000***
	Post-test	40.78	3.462	
Curiosity	Pretest	37.94	3.951	.000***
	Post-test	50.66	4.653	
Imagination	Pretest	34.44	2.015	.000***
	Post-test	48.53	4.971	
Challenge	Pretest	33.88	2.612	.000***
	Post-test	43.41	4.079	

It can be seen from Table 2 below that there is a significant difference in the overall level of creativity of students in the STEM context before and after the implementation of the robot course.

Table 2

Creativity tends to independent sample T test results

	S.D.	S.D.	t
Pretest	137.09	6.198	.000***
Post-test	183.38	13.865	

5. Discussion and conclusions

According to the results of Williams Creativity Tendency Scale, there are significant differences in creativity tendencies of students between post and pretest. As for the four dimensions of the scale, there are significant differences in risk-taking, curiosity, imagination and challenge, and there are also significant differences.

In the adventurous dimension, students have significant differences at this level before and after the implementation of the teaching. The reason we found is that students were afraid to start building or writing because of the unfamiliarity with hardware and software and the high cost of supporting facilities. Through the implementation of teaching, gradually have a certain understanding of the software and hardware related to the robot, student started to build it base on their own imagination, and tried to make some physical objects that are different or even non-existent from the construction manual, and they can use the steering gear and the controller to make the object move. In this process, the students' adventurous has changed before and after teaching.

In the curiosity dimension, students have significant differences at this level before and after the implementation of the teaching. Moreover, we found that most of the students in the class were very curious, and only a few students were not very curious during the course. Through analysis, it is found that the dimension of curiosity is influenced by many factors. Different individuals present curiosity in different situations. For example, teachers' expectations will affect students' curiosity, and individual students' cognitive level will also affect curiosity. Wait.

In the dimension of imagination, students have significant differences at this level before and after the implementation of the teaching. Through classroom observation, we found that at the beginning of the implementation of the course, students do not understand what are, the physical objects built in this course and the initial imagination score is low. The imagination is built on existing things. If you don't understand or unfamiliar with those things, it is hard for you to have the image. As the course on going, students pass multiple tasks. During the phrase of completion with diversity hardware devices, students would have more space to explore. At the end of the course, students' imagination be improved.

In the challenge dimension, students have significant differences at this level before and after the implementation of the teaching. The participants have basic knowledge about robot who have stereotypes about lead-in parts and thought it would be boring as usual. With the course carried out, students found the content is more challenging.

6. Limitation and future study

In this study, a one-semester robot course was held in a middle school in Shanghai (one section per week, 1.5 hours per session, totaling 10 sessions). The robot attitude questionnaire and the Williams Creativity Tendency Questionnaire were distributed to the students. There is a significant difference in the attitudes of students before and after the implementation of the teaching (there is no significant difference in the sub-dimension of “belief”), and there are significant differences in the tendency of creativity. Nowadays, there are more and more demand for innovative and comprehensive talents. Robot education can provide an effective way to cultivate innovative and comprehensive talents. Therefore, schools are encouraged to provide certain robot education courses when starting primary and secondary school courses. It provides a platform for developing students' creative tendencies.

We acknowledged that this study has some limitations like it is a quasi-experimental study with no control group to verify the validity of the instructional design. The conclusion of the study has no complete promotion significance but has certain reference value. Besides, the sample size is small with only 32 people and only 4 girls. So it cannot be related to the analysis of gender differences. Finally, this study does not do research based on knowledge content, then we will try to do these in later research further research and development.

References

- Wang Su (2017). Interpretation of the 2017 White Paper on Chinese STEM Education. *Modern Education*, 07, 4-7.
- Wang Xueyan, Yang Dongmei (2017). The integration of LEGO robot education and experience teaching method. *Education and Teaching Forum*, 24, 132-134.
- Zhang Lifang (2015). STEM-based Arduino robot teaching project design research. Nanjing Normal University.
- Zhang Guomin, Zhang Jianping (2008). Research on the status quo and countermeasures of robot education in basic education in China. *Modern Educational Technology*, 05, 92-94.
- State Council (2017). A new generation of artificial intelligence development planning. Retrieved December 20th, 2018, from http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
- Alimisis, D., & Kynigos, C. (2009). Constructionism and robot in education. *Teacher Education on Robotic-Enhanced Constructivist Pedagogical Methods*, 11-26.
- Eguchi, A. (2014, July). Robot as a learning tool for education transformation. In *Proceeding of 4th International Workshop Teaching Robot, Teaching with Robot & 5th International Conference Robot in Education Padova (Italy)*.
- Mataric, MJ (2004, March). Robot education for all ages. In *Proc. AAI Spring Symposium on Accessible, Hands-on AI and Robot Education*.
- Scaradozzi, D., Sorbi, L., Pedale, A., Valzano, M., & Vergine, C. (2015). Teaching robot at the primary school: an innovative approach. *Procedia-Social and Behavioral Sciences*, 174, 3838-3846.

Research on the Impact of e-Schoolbag on Students' Development: from the Perspective of Personal Learning Skills

Chu LIU^{a*}, Jueqi GUAN^b & Youqun REN^c

^a*Department of Education and Information Technology, East China Normal University, China*

^b*Department of Educational Technology, Zhejiang Normal University, China*

^c*The Institute of Curriculum and Instruction, East China Normal University, China*

*stellalc@126.com

Abstract: With the continuous development and deepening of educational informatization, society has increasingly recognized the application of digital learning media and platform in the classroom. As one of the important means of modern educational informationization, e-Schoolbag has exerted great influence on both students and teachers. In order to understand the effect of e-Schoolbag on student's growth from the perspective of personal learning and thinking skills, this study chooses 26688 students in city S district M as the research object. The analysis is from the independent learning, creative thinking, reflective learning, cooperative learning, Self-management and effective participation of these five aspects. At the same time, the study also compared the influence of different learning period and gender. The results show that the use of e-Schoolbag has different effects on the five dimensions of students. In the group of primary school students who use e-Schoolbags, gender difference also affects the influence of e-Schoolbags on students' personal learning and thinking skills.

Keywords: Education informatization, E-Schoolbag, student growth, personal learning and thinking skills

1. Introduction

In 1999, the e-Schoolbag was first used in Singapore and has received widespread attention. With the development of technology, digital educational equipment is going to be an indispensable part of students' learning life. From the blackboard in the traditional classroom, the projection screen, the electronic whiteboard, to the current each students equipped with learning electronic terminal in 1:1 digital learning environment, various of information technologies have provided a rich form of information representation through the development of practical products with leading value, which changed learners' perceptions of things and their learning behaviors (Zhiting & Jueqi, 2014).

In November 2000, the ministry of education of China promulgated the "notice on the implementation of the Notice on the implementation of the "school-school communication" project in primary and secondary schools". In 2009, it began the experimental plan about the e-Schoolbag in Taiwan area and formally implemented in Taipei in 2011. As an important terminal product of educational informatization and digitalization, e-Schoolbags began to enter the vision of schools and were repeatedly piloted in the voice of mixed praise and criticism and endless controversy (Youmei & Huibin, 2014). Ten years after the issue of project, in 2010, Hong Kong put forward the e-Schoolbag learning experimental plan. In the same year, Shanghai municipal government issued the outline of Shanghai medium - and long-term education reform and development plan (2010-2020), which clearly pointed out that digital curriculum environment construction and learning mode reform experiment should be carried out. Improve students' information literacy and innovation ability, guide students to master the ability of using modern information technology to acquire knowledge and it is necessary to promote the development of e-Schoolbag and cloud computing auxiliary teaching, and promote students to use information technology to enrich in-class and out-class learning and research(the ministry of education of the people's republic of China, 2010). Shanghai Hongkou district took the lead in responding to the call and entered into the development mode with regional units and overall

promotion by government departments in the new round of development 10 years(Xiaohua, 2014). In 2017, the information technology curriculum standards for ordinary high schools clearly pointed out that digital learning and innovation is one of the four core qualities of information technology discipline. In April 2018, the ministry of education promulgated the "action plan on education informatization 2.0", which means that China's education informatization has stepped into the 2.0 era from the 1.0 era and has higher requirements and expectations for the development of education informatization. Another 10 years have passed. Under the support and guidance of relevant policies, schools in different regions in China have launched pilot projects of e-Schoolbag and achieved good results. At the same time, there are still many scholars' influences on students and teachers brought by e-Schoolbag into the classroom, such as students' classroom teaching behavior pattern (Xianlong, Xiaobing, Yonghe, & Xinhua,2013), students' academic performance and learning experience(Jueqi, Riezebos, Xiaobing, & Zhiting, 2015), teacher development (Hui, Yanjun, &Shaoqing, 2016) and other aspects. As the most critical link in the application process of e-Schoolbag, students' feelings and growth can best reflect the application effect. Therefore, from the perspective of students' growth, this study attempts to explore the impact of e-Schoolbag application on students' personal learning and thinking skills, and further analyzes the impact of different type of students, hoping to provide a reference for relevant researchers.

2. Research Design

2.1 Research Purposes

The purpose of this survey is to find out whether the personal learning and thinking skills are improved compared to non-experimental students after using the e-Schoolbag. The results of this survey also provide reliable data and conclusions for subsequent research.

2.2 Survey Methods and Questionnaire

The survey used Guan Jueqi's self-reported questionnaire for personal learning and thinking skills (Jueqi, 2017), and used online research methods. The questionnaire consisted of 27 questions and set the level of "very agree, mostly agree, neutral, mostly disagree, and very disagree". Combined with the key research questions, the questionnaire content is divided into five dimensions for survey students, namely independent learning, creative thinking, reflective learning, cooperative learning, self-management and effective participation.

The independent learning dimension includes five statements. They are: "I can clearly identify the problem to be answered/solved, or the task to be completed", "I will solve the problem or task in a step-by-step way as planned", "I can think about the problem according to different viewpoints", "I can obtain the information to help solve the problem/complete the task" and "I can discuss the evidence in detail to support the result". The reliability analysis by SPSS 23.0 shows that the Cronbach α coefficient of this dimension questionnaire is 0.951, it is much higher than 0.7, which means that the internal consistency and the reliability of the questionnaire items are pretty good and the reliability of the questionnaire.

Creative thinking involves "I can generate ideas and explore possibilities for problems or tasks", "I will continue to ask questions to expand my ideas", "I will discuss and analyze my ideas with others", "I will question my hypothesis and other people's hypothesis" and "I will try to choose a new problem-solving method". The reliability analysis by SPSS 23.0 shows that the Cronbach α coefficient of this dimension questionnaire is 0.905, it is much higher than 0.7, which means that the internal consistency and the reliability of the questionnaire items are pretty good and the reliability of the questionnaire.

Reflective learning includes "I will evaluate myself and others and discover learning opportunities.", "In the process of problem-solving/completing, I will summarize my progress in time and act according to the results.", "When I receive feedback such as praise and criticism, I will be active processing.", "I will sum up experience and guide the next study." and "I will communicate my learning with others in different ways.". The reliability analysis by SPSS 23.0 shows that the Cronbach α

coefficient of this dimension questionnaire is 0.945, it is much higher than 0.7, which means that the internal consistency and the reliability of the questionnaire items are pretty good and the reliability of the questionnaire.

Cooperative learning includes “I will cooperate with others for a common purpose.”, “I will lead the discussion to get results.”, “I can put forward the ideas of solving problems with others”, “I will be responsible for myself and others.”, “I will take the initiative to care for and help others." and "I will provide help and feedback to others.". The reliability analysis by SPSS 23.0 shows that the Cronbach α coefficient of this dimension questionnaire is 0.951, it is much higher than 0.7, which means that the internal consistency and the reliability of the questionnaire items are pretty good and the reliability of the questionnaire.

Self-management and effective participation include “around the goal, I can actively participate in learning and perseverance”, “I can actively seek advice and support when needed”, “I can propose methods for problems/tasks and break them down into specific implementations. Steps, "I will constantly improve myself to benefit myself and others", "I will express my opinion to influence others" and "I will support different opinions or ideas than me.” The reliability analysis by SPSS 23.0 shows that the Cronbach α coefficient of this dimension questionnaire is 0.941, it is much higher than 0.7, which means that the internal consistency and the reliability of the questionnaire items are pretty good and the reliability of the questionnaire.

At the same time, the SPSS 23.0 analysis found that the KMO value is 0.905, which is greater than 0.5 and is greater than 0.9, indicating that the questionnaire has good structural validity. As shown in Table 1, the value of sig. indicates that the differences are significant.

Table 1

Questionnaire validity analysis

Test of KMO and Bartlett		
Sample enough Kaiser-Meyer-Olkin metrics		.982
Bartlett's sphericity test	Approximate Chi-square	973173.341
	df	496
	Sig.	.000**

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

2.3 Participants

26,688 questionnaires were collected and the effective questionnaires were 26,570. The school stage spans primary, middle and high schools. Among them, there are 14,026 students who have participated in the e-Schoolbag pilot project, and 12,544 students who have not used e-Schoolbag (Table 2).

Table 2

Sample basic information statistics

	Learning period	Male	Female	total
students who have participated in the e-Schoolbag pilot project	Primary	5515	5244	14026
	Middle	1248	1292	
	High	368	359	
students who have not participated in the e-Schoolbag pilot project		6520	6024	12544
Total				26570

3. Analysis of the Survey Results

3.1 Analysis of Differences between Male and Female Students

In order to understand the differences between the male and female students' personal learning and thinking skills in different learning stages, the study gets the scores corresponding to the dimensions to obtain the average value of the students under each topic. The higher the mean value of the students is in this dimension, the ability is stronger.

3.1.1 Overall Analysis of the Differences between Male and Female Students

By using SPSS 23.0 for independent sample t-test, the differences between the male and female students in the e-Schoolbag experimental site were obtained. The results of the data analysis were shown in table 3. The T-test results show that there are no significant differences between male and female students in the use of e-Schoolbag in three modules. They are independent learning, creative thinking, self-management and effective participation. However, there are significant differences in reflective learning and cooperative learning of these two modules. There are significant differences in the reflective learning module, that is, females are more able to rethink their self-reflection in learning than male, and more effectively carry out cooperative learning. Through self-reflection and teamwork, they can improve their learning effects and exercise their personal thinking skills.

Table 3

A Comparison of Male and Female Students in E-Schoolbag Pilot Project

Students' personal learning and thinking skills	Gender	N	Mean	SD	p-value
independent learning	Male	7131	4.2028	.76825	.130
	Female	6895	4.2219	.72630	
creative thinking	Male	7131	4.1041	.74664	.432
	Female	6895	4.0944	.71825	
Reflective learning	Male	7131	4.1579	.74929	.014
	Female	6895	4.1883	.70880	
Cooperative learning	Male	7131	4.2129	.72170	.012
	Female	6895	4.2429	.68492	
self-management	Male	7131	4.1702	.72888	.108
	Female	6895	4.1896	.69631	

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

3.1.2 Analysis of the Differences between Male and Female Students in Primary School

By using SPSS 23.0 for independent sample T-test, the differences between the male and female students in the primary school of e-Schoolbag experimental sites were answered in the questionnaire. The results of the data analysis were shown in table 4. The results of the T-test show that there are no significant differences between the male and female students in the primary school by using e-Schoolbag when it comes to independent learning, creative thinking, self-management and effective participation. However, there are significant differences in the modules in reflective learning and cooperative learning. The data shows that female students in primary school are more able to rethink and introspect after class than male. At the same time, they also pay more attention to teamwork and learning in daily learning, and promote their learning and development through collective wisdom.

Table 4

A Comparison of Male and Female Students in Primary School E-Schoolbag Pilot Project

Students' personal learning and thinking skills	Gender	N	Mean	SD	p-value
independent learning	Male	5515	4.1287	.75980	.136
	Female	5244	4.1502	.73237	
creative thinking	Male	5515	4.0257	.73286	.565
	Female	5244	4.0177	.71812	
Reflective learning	Male	5515	4.0839	.73736	.010
	Female	5244	4.1201	.70822	
Cooperative learning	Male	5515	4.1460	.70919	.018
	Female	5244	4.1778	.68638	
self-management	Male	5515	4.1035	.71614	.075
	Female	5244	4.1278	.69608	

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

3.1.3 Analysis of the Differences between Male and Female Students in Middle School

By using SPSS 23.0 for independent sample T-test, the differences between the male and female students in the middle school of e-Schoolbag experimental sites were answered in the questionnaire. The results of the data analysis were shown in table 5. The results of the T-test showed that there were no significant differences between the male and female students in the five dimensions of independent learning, creative thinking, reflective learning, cooperative learning, self-management and effective participation.

Table 5

A Comparison of Male and Female Students in Middle School E-Schoolbag Pilot Project

Students' personal learning and thinking skills	Gender	N	Mean	SD	p-value
independent learning	Male	1248	4.4303	.77777	-0.966
	Female	1292	4.4576	.64172	
creative thinking	Male	1248	4.3577	.76374	-0.577
	Female	1292	4.3420	.65862	
Reflective learning	Male	1248	4.3936	.76194	-0.529
	Female	1292	4.4085	.65731	
Cooperative learning	Male	1248	4.4197	.75026	-1.265
	Female	1292	4.4545	.63035	
self-management	Male	1248	4.3841	.75931	-0.265
	Female	1292	4.3915	.65125	

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

3.1.4 Analysis of the Differences between Male and Female Students in High School

By using SPSS 23.0 for independent sample T-test, the differences between the male and female students in the high school of e-Schoolbag experimental sites were answered in the questionnaire. The results of the data analysis were shown in table 6. The results of the T-test showed that there were no significant differences between the male and female students in the high school using the e-Schoolbag in the five modules of independent learning, creative thinking, reflective learning, cooperative learning, self-management and effective participation.

Table 6

A Comparison of Male and Female Students in High School E-Schoolbag Pilot Project

Students' personal learning and thinking skills	Gender	N	Mean	SD	p-value
independent learning	Male	368	4.4026	.64041	.714
	Female	359	4.3294	.73304	
creative thinking	Male	368	4.3605	.64334	.644
	Female	359	4.2975	.68091	
Reflective learning	Male	368	4.3605	.65809	.232
	Female	359	4.3378	.67435	
Cooperative learning	Male	368	4.4079	.63735	.817
	Female	359	4.3291	.66789	
self-management	Male	368	4.3575	.64504	.624
	Female	359	4.2969	.67024	

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

3.2 A Comparison of Students' Personal Learning and Thinking Skills of the Impact of e-Schoolbag

The study uses SPSS 23.0 to analyze the data of students whether or not to use e-Schoolbag. The results of the data analysis were shown in table 7. The results of T-test show that there are extremely significant differences in five modules independent learning, creative thinking, reflective learning, cooperative learning, self-management and effective participation between students use e-Schoolbag and not. That is, students who use e-Schoolbag to learn have higher independent learning ability than those who do not use e-Schoolbag (Mean = 4.2122>4.0420); they have stronger creative thinking skills (Mean = 4.0993>3.9616); stronger reflective learning ability (Mean = 4.1728>4.0207); stronger cooperative learning ability (Mean = 4.2277>4.0966); stronger self-management and effective participation (Mean = 4.1797>4.0418).

Table 7

A Comparison between experimental and non-experimental E-Schoolbag Pilot Project

Students' personal learning and thinking skills	School type	N	Mean	SD	p-value
independent learning	Experimental	14026	4.2122	.74796	.000
	Non-experimental	12544	4.0420	.74267	
creative thinking	Experimental	14026	4.0993	.73281	.000
	Non-experimental	12544	3.9616	.70635	

Reflective learning	Experimental	14026	4.1728	.72980	.000
	Non-experimental	12544	4.0207	.71715	
Cooperative learning	Experimental	14026	4.2277	.70399	.000
	Non-experimental	12544	4.0966	.68590	
self-management	Experimental	14026	4.1797	.71310	.000
	Non-experimental	12544	4.0418	.69599	

Note: * .p<0.05, ** .p<0.01, *** .p<0.001

4. Research Conclusion

4.1 *the Effect of E-Schoolbag to the Different Gender of Different Learning Period Students' Personal Learning and Thinking Skills*

Through the analysis of the effect of e-Schoolbag to the different gender of different learning period students' personal learning and thinking skills, we can get that there are no significant differences in the three modules of independent learning, creative thinking, self-management and effective participation for students who use e-Schoolbag to learn. However, there are significant differences between the two modules of reflective learning and cooperative learning. There are significant differences between reflective learning and cooperative learning and the differences between the two modules are shown as the ability of female students to be higher than male. However, only in the primary school stage, there are significant differences between the male and female in the reflective learning module and the cooperative learning module. There are no differences in other modules. There are no differences between the other learning period, including middle school and high school. On the one hand, due to the increase of learning pressure in middle and high school, students pay more attention to the memory of knowledge in the use of e-Schoolbag and focus on improving their academic performance, while ignoring the improvement of their own thinking. The lower the learning period, the broader the thinking of students and the use of e-Schoolbag can jump out of the inherent framework, which will produce different effects. On the other hand, students with less experience in e-Schoolbag have a more positive attitude. At the beginning of the use of e-Schoolbag, due to the novelty, students have a positive attitude towards e-Schoolbag. After using the e-Schoolbag for a long time, the slow update speed of the e-Schoolbag restricts the enthusiasm of the students, and the students with less experience have a more positive attitude towards the e-Schoolbag (Lichun & Jianli, 2015).

4.2 *the Effect of E-Schoolbag to Students' Personal Learning and Thinking Skills*

Through the analysis of students whether to use e-Schoolbag or not, based on the fact that there are just several inevitable differences between the students, such as teachers, the level of the schools, we found that students who use e-Schoolbag have a more effective improvement in their personal learning and thinking skills. It can be seen through the extremely significant differences in independent learning, creative thinking, reflective learning, cooperative learning, self-management, and effective participation. That is, overall, compared with the group that did not use the e-Schoolbag, the students who used the e-Schoolbag showed higher independent learning ability, stronger creative thinking ability, stronger reflective learning ability, stronger cooperative learning ability, stronger self-management and effective participation consciousness. It can be seen that the application of e-Schoolbag can not only feel the advantages brought by external factors but also feel the growth of students' thinking.

Limitation and Outlook

This study about e-Schoolbag is based on the example of all pilot students in one of districts of Shanghai and investigates and analyzes the changes in personal learning and thinking skills in the

process of participating in the e-Schoolbag project. However, the research still has some shortcomings. First of all, the students' personal learning and thinking skills are measured only through questionnaires. There is no analysis and discussion on the deeper cognitive rules of students. Secondly, due to the large amount of experimenters, there are many uncontrollable interference factors, such as students' level of knowledge, the teaching style and level of application of e-Schoolbag of teachers. I hope that the greatest improvement will be made in the in-depth study in the future.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References

- Hui Tong, Yanjun Yang, & Shaoqing Guo. (2016). A Review and Reflection on the Research Development of e-Schoolbag Application Effect Evaluation. *Journal of Distance Education*, 34(1), 99-112.
- Jueqi Guan. (2017). Study on the digital learning ability of primary and middle school students in the electronic schoolbag environment. *East China Normal University*.
- Jueqi Guan, Riezebos Peter, Xiaobing Su, & Zhiting Zhu. (2015). The influence of electronic schoolbag on students' learning experience and achievement -- a case study of mathematics application in Minhang district primary school in Shanghai. *China Educational Technology*, (9), 56-62.
- Lichun Zhang, & Jianli Jiao. (2015). Students' Attitude and Understanding of e-Schoolbag. *Open Education Research*, 21(2), 98-105.
- The ministry of education of the people's republic of China. the outline of Shanghai medium - and long-term education reform and development plan(2010-2020) [EB/OL]. <http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s4604/201010/110458.html>, 2019-05-20.
- Xianlong Xu, Xiaobing Su, Yonghe Wu, & Xinhua Wang. (2013). Analysis of classroom teaching behavior mode for electronic schoolbag application. *Modern Distance Education Research*, (2), 84-91+106.
- Xiaohua Yu. (2014). The U.S. iPad project and its implications for Chinese electronic backpacks. *Open Education Research*, 20(02), 46-55.
- Youmei Wang & Huibin Chen. (2014). Research hotspot and development trend of electronic schoolbag in China in recent ten years -- knowledge map analysis based on common word matrix. *China Educational Technology*, (5), 4-10.
- Zhiting Zhu & Jueqi Guan. (2014). Technological forces in educational change. *China Educational Technology*, (1), 1-9.

Influence of Financial Course on Eighth Grade Students' Financial Concepts, Math Motivation, Math Anxiety in Taiwan

Yu-Ching HSU, Mei-Wen NIAN, Chang-Hsuan YANG, Yuan-Hsuan Lee, Jiun-Yu WU*

Yu-Ching Hsu, National Chiao Tung University, Taiwan

Mei-Wen Nian, National Chiao Tung University, Taiwan

Chang-Hsuan Yang, National Chiao Tung University, Taiwan

Yuan-Hsuan Lee, National Tsing Hua University, Taiwan

Jiun-Yu Wu, National Chiao Tung University, Taiwan

**jiunyu.rms@gmail.com*

Abstract: In the generation of educational reform, we have always wondering what kind of important ability has been taught in the curriculum? Financial literacy is an essential life skill (OECD, 2017). No matter what kinds of occupation will a person have, he or she must know how to manage his or her asset. In this research, we explored that if students' math motivation, math anxiety, and financial concept would change or not after they received the hybrid financial courses in 4 weeks. We used math motivation and math anxiety as grouping variables, then use *k*-mean clustering to separate students into two groups. Two-way mixed-design ANOVAs were conducted to test the mean differences and change of math motivation, math anxiety, and financial concept across time points between different groups. We found only the significant change of math anxiety from students after they received the financial courses, while math motivation and financial concept kept constant.

Keywords: Math motivation, math anxiety, financial literacy

1. Introduction

What should the curriculum in junior high school be? What should students learn in school? There is an important issue needs to be mentioned first—after graduate from high school, no matter what kinds of occupation will a person have, he or she must know how to manage his or her own properties. Therefore, financial literacy is an essential life skill (OECD, 2017). However, the contemporary awareness of how to cultivate children's financial concepts has not yet matured in Taiwan. Thus, how to educate financial concepts is an important issue. Furthermore, what is the influence of financial education on student learning achievement? It is also an issue that we urgently need to explore.

Skagerlund, Lind, Strömbäck, Tinghög & Västfjäll (2018) found that math anxiety can predict financial literacy of adults. Therefore, the researchers in this study believe that the exploration of financial education on math motivation and anxiety should be completed first because the financial concepts can be regarded as an application of mathematics in daily life. For students who have serious anxiety about mathematics, expecting to reduce their anxiety through a financial course, which was based on the mathematics in daily life. The connection between motivation, anxiety, and mathematical achievement has been explored for a long time (McDonough & Ramirez, 2018; Wang et al., 2015). Although the promotion of the financial concepts is also a matter in recent years, the researchers in this study believe that the effect of financial education combined with the mathematics curriculum is a matter of great concern. If there is a way to promote students' math motivation and reduce math anxiety while promoting financial education. Furthermore, it will also increase the students' financial concepts or the possibility of improving mathematics achievements. Therefore, the research question is how different math motivation, math anxiety, and financial concepts are before and after the financial courses for students with different types of math motivations and math anxieties?

2. Literature Review

2.1 *Financial Literacy and its emerging importance*

Faced with the diversification of financial products, people learn how to manage their own money is an important skill. The OECD pointed out in the 2014 report that Financial literacy is an essential life skill. The Office of Financial Education (OFE) in the United States advocates that schools are the best places to provide financial education (OFE, 2002). OECD also pointed out that the sooner you get into financial education, the better (OECD, 2005). In recent years, countries have attached importance to financial education and have put forward proposals for financial education. The United States proposed the National Standards in Personal Finance with Benchmarks, Applications and Glossary for K-12 Classrooms (OFE, 2002). Australia also proposed the National Consumer and Financial Literacy Framework (NCFLF, Consumer and Financial Literacy Working Party, 2005).

After the global financial crisis, it once again caused international awareness of financial concepts. Since 2006, the Financial Supervisory Commission of Taiwan has actively entered the campus and community for financial knowledge promotion activities, and established the Risk Management and Insurance Education Promotion Portal (<https://rm.ib.gov.tw/Pages/SiteMap.aspx>) to promote domestic financial education. Also in 2012, Basic Learning Framework for Financial Education was proposed and clearly pointed out the financial education priorities and learning objectives of Citizenship Education. Both foreign countries and Taiwan have actively promoted financial education in recent years. For this reason, this study hopes to improve the financial concepts and the importance of it.

2.2 *Math Motivation, Math Anxiety and Achievement*

Motivation has a positive impact on learning achievement, while anxiety negatively affects learning achievement (Wang, Shakeshaft, Schofield, & Malanchini, 2018). It is often taken by both to consider the impact on students' performance. In addition, some scholars have explored the impact of learning anxiety on future behaviors in daily life (Skagerlund et al., 2018).

McDonough & Ramirez (2018) pointed out that if students have higher math anxiety and math self-concept, they will initiate their own defense mechanism, which will make it easier for students to forget the learning content and affect the learning outcomes. For students with high anxiety in mathematics, the achievements of mathematics usually do not perform well. Thus, those students also evade mathematics classes, indicating that math anxiety may affect mathematics achievements in different ways. Wang et al. (2018) found that people who have high anxiety about mathematics exams spend more time on studying. However, Wang et al. (2015) found that the relationship between math anxiety and math performance varies with the math motivation of students. For students with high math motivation, math anxiety and math performance have an inverted-U relation, while students with low math motivation have a negative linear relation between math anxiety and mathematics.

The relation between motivation, anxiety and learning outcomes is complex. Scholars often discuss motivation and anxiety at the same time to predict achievement (Wang et al., 2015). Wang et al. (2018) used math motivation and math anxiety to do the cluster analysis for exploring the differences in motivation and anxiety between different levels of learning time and outcome in math. Skagerlund et al. (2018) explored the relation between adult financial literacy and math anxiety. They found that math anxiety negatively affects financial literacy, so people may also evade financial behavior because of math anxiety.

Skagerlund et al. found that the impact of financial literacy on math anxiety. Although there are many studies that indicate the impact of motivation and anxiety on learning, there is no research indicates the influence of math motivation, math anxiety on the financial concepts for the junior high school students. Therefore, this study expects to use math motivation and math anxiety for grouping and discuss whether our financial courses can effectively enhance students' learning motivation, or change students' learning anxiety and improve students' financial concepts.

3. Method

In this study, the financial courses were used as an intervention for a four-week study to explore the changes of Taiwanese junior high school students in math motivation, math anxiety, and financial concepts. In order to explore the influence of financial courses on students with different levels of motivation and anxiety, this study further used math motivation and math anxiety to divide students into two groups. We expected that after the financial course, students' math motivation and financial concepts will increase, math anxiety will decrease, and students with different combinations of motivation and anxiety will have different financial concepts.

3.1 Participants

The participants of this study were 48 eighth grade students (22 males and 26 females) in Hsinchu, Taiwan. Those participants used one hour of mathematics alternative learning period per week to learn financial concept from an instructor for four weeks.

3.2 Experimental Design and Procedure

This study is mainly concerned with the changes in student's math motivation (MM), math anxiety (MA), and financial concept (FC) before and after the financial class. In the first week, the participants completed the questionnaires of math motivation (MM_pre), math anxiety (MA_pre), and financial concept (FC_pre). Next, the instructor introduced various financial tools (i.g. demand deposit, time deposit, and stock) and brought the students financial concepts. During the second and third week, the students used a public website of stock simulation to conduct investment actions (<https://www.cmoney.tw/vt/>). At the same time, the instructor brought the concept of math and finance.



Figure 1. The public website of stock simulation (<https://www.cmoney.tw/vt/>).

In the last week, the instructor summarized the simulation of these two weeks and carried out the post-test of math motivation (MM_post), math anxiety (MA_post), and financial concept (FC_post) to the students. Finally, the research teams merged the student's math motivation, math anxiety, and financial concept to do the analysis.

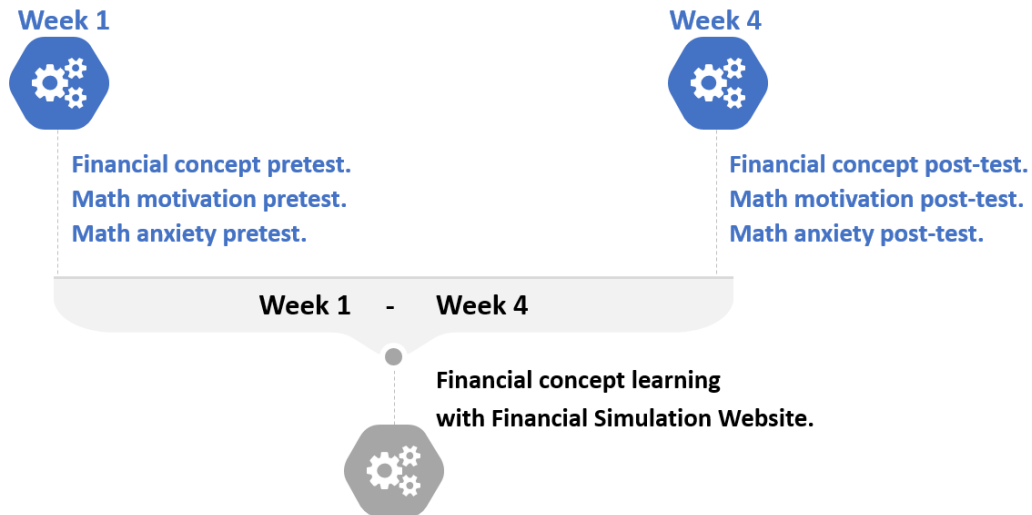


Figure 2. The procedure of financial course for four weeks.

3.3 Measure and Instrument

3.3.1 Math Motivation

The questionnaire of math motivation was based on Motivational Strategies in Calculus Learning Questionnaire for Taiwan Technology University Students (MSCLQ-TTU, (Hsin, Lin, & Yeh, 2005)). This study rewrote the calculus as mathematics. In addition, this study used 18 items in the motivation part of MSCLQ-TTU. Each item is Likert 5-point scale. Finally, the researchers summed up all the scores of the math motivation questionnaire and converted it into z-score. The higher the score is, the higher the motivation is.

3.3.2 Math Anxiety

The questionnaire of math anxiety was based on Modified Abbreviated Math Anxiety Scale (mAMAS, (Carey, Hill, Devine, & Szűcs, 2017)). There are 9 items. Each item is 5-point Likert scale. Participants assessed the level of anxiety experienced by each event in the math class (1 = low anxiety to 5 = high anxiety). The researchers summed up all the scores and converted it into z-score. The higher total score is, the higher anxiety that participants felt in math class is.

3.3.3 Financial conceptions

The questionnaire of financial were designed by researchers, extended from course materials. There are 10 items in the financial concept questionnaire, all of which are the true-false test. For example, *the interest rate of time deposit is always higher than demand deposit*. The researchers calculated the rate of right answers and converted it into z-score. The higher score, the better the concept of finance.

3.4 Statistical Analysis

This study used IBM SPSS Statistics 20 and R & Rstudio v.3.4.4 for statistical analysis. First, the researchers used MM_pre and MA_pre as grouping variables, performed the K-mean analysis and divided students into two groups. Then, the researchers used Two-way mixed-design ANOVA to test the differences between the groups and time in math motivation, math anxiety and financial concepts.

4. Results

4.1 Non-hierarchical Cluster Analysis

Researchers used MM_pre and MA_pre to divide the 48 participants into two groups. Group 1 were made up of 26 students (10M, 16F), and Group 2 were 22 students (12M, 10F). After the examination of t-test, there were no statistically significant differences between group means and sample mean on MM_pre ($G1_{MM_pre}: t(25) = -1.03, p = .314$; $G2_{MM_pre}: t(21) = 1.22, p = .237$). Thus, the two groups' MM_pre are “average motivation(AM)”. In math anxiety, those two groups had statistically significant difference between group means and sample mean on MA_pre ($G1_{MA_pre}: t(25) = 7.51, p < .001$; $G2_{MA_pre}: t(21) = -8.34, p < .001$). Therefore, we claimed that G1 is “high anxiety(HA)”, and G2 is “low anxiety(LA)”. Conclusively, we regarded G1 as “average motivation and high anxiety(AMHA)” group and viewed G2 as “average motivation and low anxiety(AMLA)” group. The z-scores of variables of these two groups are in Table 1.

Table 1

Z-score of Math Motivation, Math Anxiety and Financial Concepts in two groups.

Group	N	MM_pre	MA_pre	MM_post	MA_post	FC_pre	FC_post
G1 (AMHA)	26	-0.18	0.94	-0.20	0.31	0.17	0.07
G2 (AMLA)	22	0.28	-0.88	0.29	-0.56	0.23	-0.30

4.2 Two-way Mixed-design ANOVA

Accordingly, the differences in math motivation, math anxiety and financial concept between the two groups and between time are both noticeable. Thus, we used the following three mixed-design ANOVAs to examine the difference, with group and time as independent variables, math motivation, math anxiety and financial concept as dependent variables.

4.2.1 Math Motivation

We used group and time as I.V., math motivation as D.V., and carried out the two-way mixed-design ANOVA to test the influence of group and time on math motivation. The results showed that the main effects of the group and time were not significant ($F_G(1,46) = 3.426, p = .071, \eta^2 = .057$; $F_T(1,46) = 0.002, p = .969, \eta^2 < .001$). Besides, there were no interaction effect as well ($F_{G \times T}(1,46) = 0.012, p = .914, \eta^2 < .001$). Consequently, the math motivation did not differ between the two groups and there was no significant change and maintained at a certain level before and after the course.

Table 2

Two-way mixed-design ANOVA with math motivation as D.V., group and time as I.V.

Source	SS	df	MS	F	p	η^2
Between groups						
Group	5.312	1	5.312	3.426	.071	.057
Time	0.001	1	0.001	0.002	.969	<.001
Group*Time	0.004	1	0.004	0.012	.914	<.001
Within groups						
between subjects	71.322	46	1.550			
residuals	15.837	46	0.340			
Total	92.476	95				

4.2.2 Math Anxiety

Based on the MA as D.V., group and time as I.V.. The two-way mixed-design ANOVA results showed significant interactions with ($F_{G \times T}(1,46) = 23.16, p < .001, \eta^2 = .120$). Therefore, a simple main effect test was performed. The results showed that under the condition of G1(AMHA), MA_pre was significantly higher than MA_post ($\mu_{MA_pre-MA_post} = 0.630, p < .001$). Under the condition of G2(AMLA), MA_post significantly higher than MA_pre ($\mu_{MA_post-MA_pre} = 0.320, p = .018$). The MA of G1(AMHA) was significantly greater than MA of G2(AMLA) regardless of pre or post test (MA_pre: $\mu_{G1-G2} = 1.83, p < .001$; MA_post: $\mu_{G1-G2} = 0.88, p < .001$).

Further examination of the mean, we found that the MA_post mean difference between G1(AMHA) and sample was not significant ($t(25) = 1.91, p = .067$), but the mean difference was significant between G2(AMLA) and sample ($t(21) = -3.67, p = .001$). That is to say, G1(AMHA) has dropped from high anxiety to average anxiety, while the G2(AMLA) anxiety has increased, but it is still low.

Table 3

Two-way mixed-design ANOVA with math anxiety as D.V., group and time as I.V.

Source	SS	df	MS	F	p	η^2
Between groups						
Group	43.46	1	43.46	60.89	<.001***	.500
Time	0.57	1	0.57	2.47	.123	.013
Group*Time	5.38	1	5.38	23.16	<.001***	.120
Within groups						
between subjects	32.84	46	0.71			
residuals	10.68	46	0.23			
Total	92.93	95				

Table 4

Simple main effect test of mixed-design ANOVA with math anxiety as D.V., group and time as I.V.

Source	SS	df	MS	F	p	Post-hoc
Time at						
G1(AMHA)	5.16	1	5.16	22.43	<.001***	$\mu_{pre-post} = 0.630, p < .001$ ***
G2(AMLA)	1.34	1	1.13	4.91	.032*	$\mu_{post-pre} = 0.320, p = .018$ *
Residuals	10.68	46	0.23			
Group at						
Pre-test	39.71	1	39.71	83.95	<.001***	$\mu_{G1-G2} = 1.83, p < .001$ ***
Post-test	9.132	1	9.13	19.31	<.001***	$\mu_{G1-G2} = 0.88, p < .001$ ***
Residuals	43.52	92	0.47			

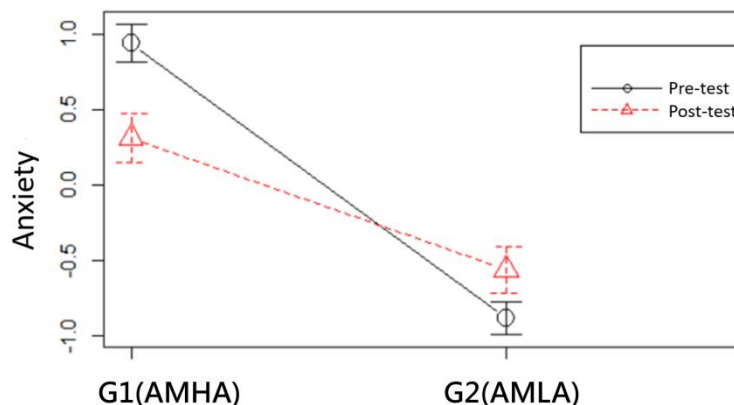


Figure 3. The plot mean of mixed-design ANOVA about MA .

4.2.3 Financial concepts

The mixed-design ANOVA was performed with FC as D.V., group and time as I.V.. As a result, it was found that the spherical test hypothesis was rejected (T: *Mauchly's W* = 0.87, $p = .038$; $T \times G$: *Mauchly's W* = 0.87, $p = .03$), so the following results were reported for the Greenhouse-Geisser correction. The main effects and interaction effect were not significant ($F_G(1,46) = 0.10, p = .001$; $GGe_T = 0.88, p = .107$; $GGe_{T \times G} = 0.88, p = .297$), indicating that the two groups did not have differences on FC. FC maintained a certain level at different times, and there was no interaction between groups and time.

5. Discussion and Conclusion

Researchers used math motivation and math anxiety as grouping variables to divide the 48 participants into two groups. Group 1 was regarded as a group with average math motivation and high math anxiety, Group 2 was a group with average math motivation and low math anxiety. Next, after the financial courses, the researchers explored the changes in students' math motivation, math anxiety and financial concepts.

Researchers found that there is a significant difference between the two groups in math anxiety, but there is no difference in math motivation and financial concepts. Before and after the financial course, students maintained certain math motivation and financial concepts, but anxiety changed after the course. The following are the discussions of the results.

From the perspective of analysis results, Group 1's anxiety has always been higher than Group 2's. It is worth noting that we can find significant changes in math anxiety of the two groups before and after the course. Group 1 had originally high anxiety, but after the financial course, math anxiety dropped to average anxiety. Group 2 had originally low anxiety, although it still remained low anxiety after course, the math anxiety was improved. As Wang et al. (2015) said, it was found that whether the anxiety remained extraordinary high or low will affect students' learning performance. That is to say, the extraordinary high anxiety may make students unable to concentrate on learning, and negative emotions may occur during the study, which may affect the grades. On the other hand, if students have low anxiety, it may represent that students do not pay attention to learning, do not care, or even ignore learning, resulting in poor learning. Therefore, under average anxiety, students can face up to learning but they do not interfere with negative learning emotions.

From the analysis of the financial concepts, there is no difference between two groups before and after the course, which is different from our initial assumptions. It may be because the grouping variables we used are motivation and anxiety about mathematics, not about financial concepts. Mathematics is a discipline for students, but the financial concepts is a life skill. Perhaps for this reason, the results of math motivation and anxiety grouping have little effect on the financial concepts. In addition, we also found that there is no difference between pre-test and post-test in financial concepts. It may be because this study only teaches on stocks and investment risks, and the period is only four weeks. Another possibility is that the number of questions in the test is not enough. Since it is a self-edited test in our study, the content of the topic bias concept is more. But there are fewer items in the judgment of financial behavior and financial affairs. Therefore, it is suggested that future research can increase the number of questions in the financial test and the breadth of the content to test the knowledge of students' financial concepts more comprehensively.

This study used math motivation and math anxiety to group students and examine the differences in math motivation, math anxiety, and financial concepts before and after the implementation of the financial course. The research results show that with such a financial course, it is beneficial to students' changes in math anxiety, from extreme high and low anxiety to the average anxiety. Math motivation and financial concepts also maintain a certain level. It also proves that such a life-oriented learning course, in addition to cultivating students' life skills, understanding financial concepts and investment risks, also helps students' learning anxiety changes in the discipline.

In the future, it is recommended that researchers, in addition to motivation and anxiety, can try to collect more student traits or increase the number of classification groups. Refer to Wang et al. (2018)

for a more detailed motivation and anxiety grouping to explore more combinations of different learning states, as well as a more comprehensive interpretation of learner traits and learning.

References

- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology, 8*.
- Hsin, C.-I., Lin, S. S. J., & Yeh, C.-C. (2005). Pilot analysis of motivational strategies in learning calculus among technology college students' motivational strategies in learning calculus. *Journal of National University Tainan, 39*(2), 65–82.
- McDonough, I. M., & Ramirez, G. (2018). Individual differences in math anxiety and math self-concept promote forgetting in a directed forgetting paradigm. *Learning and Individual Differences, 64*, 33–42. <https://doi.org/10.1016/j.lindif.2018.04.007>
- OECD (2005). Improving financial literacy. *Financial Market Trends, 2005*(2), 111–123. <https://doi.org/10.1787/fmt-v2005-art11-en>
- OECD (2017). *PISA 2015 results: Students' financial literacy* (Vol. IV). Paris, France: OECD Publications. <https://doi.org/10.1787/9789264270282-en>.
- Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G., & Västfjäll, D. (2018). Financial literacy and the role of numeracy—How individuals' attitude and affinity with numbers influence financial literacy. *Journal of Behavioral and Experimental Economics, 74*, 18–25. <https://doi.org/10.1016/j.socec.2018.03.004>
- Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., ... Petrill, S. A. (2015). Is mathematical anxiety always bad for math learning: The role of math motivation. *Psychological Science, 26*(12), 1863–1876. <https://doi.org/10.1177/0956797615602471>
- Wang, Z., Shakeshaft, N., Schofield, K., & Malanchini, M. (2018). Anxiety is not enough to drive me away: A latent profile analysis on math anxiety and math motivation. *PloS One, 13*(2), e0192072. <https://doi.org/10.1371/journal.pone.0192072>

Geneticus Investigatio: A Classroom-Based Technology-Enhanced Learning Environment for Problem-Solving Process Skills in Genetics

Anurag DEEP^{a*}, Sahana MURTHY* & Jayadeva BHAT*

Indian Institute of Technology Bombay, India

^aanuragdeep4949@iitb.ac.in

Abstract: Solving problems in science domains requires learners to integrate concepts across topics, along with problem-solving and inquiry process skills. The complexity of these concepts and skills becomes manifold at the tertiary undergraduate level, and are known to be challenging for learners. To support learners in this process, we have designed and developed Geneticus Investigatio (GI), a technology-enhanced learning environment for semi-structured problems in the context of Mendelian genetics. GI facilitates the integration of concepts and problem-solving process skills, through inquiry-based learning activities interspersed with evaluative question prompts and reflective activities. GI is developed using Google sites and H5P as a web-based learning environment to provide easy access to learners and to enable teachers to adapt the learning activities to different topics. In this paper, we present the theoretical basis and design of GI. We report a quasi-experimental classroom study (N=63) in which we investigated learning of problem-solving skills and perceptions of usability and usefulness of GI. The results indicate higher learning gains of problem-solving process skills with GI, and learner perceptions that certain activities in GI are helpful for learning concepts and process skills.

Keywords: Problem-solving process skill, Question Prompts, Mendelian Genetics

1. Introduction

Problem-solving is an integral part of the undergraduate science curriculum in various topics. These problems can range from well-structured to ill-structured (Fernandes & Simon, 1999). An in-between type is a semi-structured type of problem, in which students are required to connect concepts across topics and use a variety of problem solving and inquiry skills. Such problems are common in genetics, which is a compulsory foundational course of undergraduate bioscience and medical learners. In these problems, learners are required to identify and justify the patterns of inheritance behind various biological phenomena. These problems may follow various theoretical frameworks such as Mendelian or non-Mendelian inheritance patterns, and encompass a variety of concepts related to the breeding context of plants and animals. In addition to identifying and applying appropriate theories and concepts, such problems require the formulation of hypotheses, identification of variables, making predictions and testing predicted outcomes against the experimental ones. Thus a learner has to simultaneously work with concepts across multiple topics, problem-solving skills, and inquiry skills. Experts solve these kinds of problems by addressing all these complexities, starting with analyzing the problem in parts and generating a possible hypothesis for an explanation. This is followed by designing the experiment, interpreting the results according to the theoretical framework and the hypothesis. The difference between expected and observed results is calculated using statistical or other methods, based on which a decision is to be made as to whether the results are acceptable or not. This process is often unclear to novice learners, especially why these steps are required (Kim & Hannafin, 2011), leading to mechanical application of the problem-solving steps (Karagoz & Cakir 2011). Thus, learners need to explicitly learn problem-solving skills within the context of genetics experiments and integrate them along with the relevant theoretical concepts.

Research suggests utilizing the affordances of the technology-enhanced learning (TEL) environments in developing these skills by providing overall structure to the learning activities, immediate and personalized feedback, reflective question prompts, and so on. Several inquiry-based learning environments focus on developing such skills, for example, WISE (Slotta 2002), Geniverse

(Concord consortium 2010). Most of these TEL environments are for topics in high-school and middle school science.

Our proposed TEL environment Geneticus Investigatio (GI) is designed for college-level biology undergraduates with a focus on the solving problems by applying genetics concepts, understanding of basics of statistics with integrated science process skills. GI emphasizes integration of concepts across the required topics, along with reflective and evaluative question prompts and scaffolds. These question prompts engage learners in the understanding of concepts across the multiple topics to solve the problem, and on reflection of actions performed while doing the learning activities. Through the learning activities in GI, learners perform steps of problem-solving like identifying hypotheses, designing of experiments, comparing predicted and observed results, and accepting or rejecting the hypothesis.

One design consideration in the development of GI was to make it easily accessible to learners in college classrooms, and to make it adaptable to different topics if instructors wished to add or edit content, learning activities or problems in other topics. GI has thus been developed using Google sites and H5P (Jouble, 2013) as a web-based learning environment which is known to be convenient to access digital content by students and teachers alike (Yin et al., 2017). GI is browser based and works on laptops, tablets, or mobile phones. Students can use it easily in classroom settings or anywhere else and only need a device and wireless connection, and teachers can adapt it quickly since it does not require advance computer knowledge.

In this paper, we describe the theoretical basis and design of GI, and report a quasi-experimental classroom study with 63 undergraduate bioscience learners in the context of Mendelian genetics. The two research questions which were the focus of this study are:

- Do students who interact with GI develop problem-solving process skills?
- What are user perceptions of usability and usefulness of GI?

2. Background Research

One of the principal goals of science education has been training learners in solving problems of various types. Researchers and theorists have made remarkable progress in identifying and characterizing problem solving. Some of them include identifying students' difficulties in diverse contexts, proposing problem-solving phases, and associated learning activities (Xun & Land 2004). A variation of this problem-solving activity includes hypothesis generation, testing, and revision. This variation is essential in the context of understanding or doing science by the learners, which is the core practice in science education. It helps in establishing the feasibility/correctness of a hypothesis, eliminate candidate hypothesis/set of results, and compare predicted/observed results. It also allows learners to develop an in-depth understanding of the subject (Cooper, Hanmer, & Cerbin, 2006).

Decades of research into science inquiry learning has given us insights into nature of learning and challenges, design of learning environments to support students (Bransford, Brown, & Cocking, 2000), and guidelines and principles for scaffolds (Quintana et al, 2004). Affordances of TEL environments have been used to develop scientific inquiry skills similar to problem-solving skills. Among them are WISE (Slotta 2002), Go-Labs (Jong et. al. 2014), Apple Tree (Chen et al. 2013), and Geniverse (Concord consortium 2010). These environments are to be used either online or can be downloaded. Most of this research has been in topics for the middle and high school levels.

In undergraduate science learning, a significant difficulty that has been reported in genetics problem-solving is rote application of problem-solving process steps without a comprehensive conceptual understanding of these steps (Karagoz & Cakir 2011). In typical undergraduate curricula, students encounter the required concepts and skills for solving such problems across across different courses, for example, knowledge of genetics concepts in a basic or advanced genetics course, concepts of statistics in their bio-statistics course and application of process skills in practical labs. Thus students lack an integrated perspective while solving problems. So there is a need for developing learners' problem-solving skills, especially for open problems. This skill is especially crucial for undergraduate bio-science learners in the context of genetics as it deals with multiple underlying reasons for a biological observation. To pinpoint the specific reason, learners have to generate, test, and revise the hypothesis. An example from the context of genetics is that learners are required to identify and justify the patterns of inheritance behind various biological phenomena. To identify these inheritance patterns,

they have to solve problems which are either cause-effect problems (closed problems) or effect-cause problems (open problems) (Orcajo & Aznar 2005).

Solving such open-ended problems becomes a daunting task for them. They require scaffolds at various places during the problem-solving process. These scaffolds can be in the form of feedback, access to domain concepts, etc. An example of feedback could be in the way of guiding questions for reflection or identification of possible mistakes. To do that, technology affordances is widely used. Some of the existing learning environments which meet some of these requirements are Genetics with Jean (Thompson & McGill 2017) an affective tutoring system to teach the concepts of genetics. Another case-based laboratory simulation was built for learning core concepts and skills in medical genetics (Makransky et al. 2016). Some of the interactive affordance from these environments could meet our need for the learning environment. But there is a need of affordances that could help in providing overall structure to the sequence of learning activities along with dynamic and personalized feedback in various forms. An example of this is reflective question prompts for evaluating conceptual understanding while watching an interactive video. Another example is the interactive learning activities like drag and drop activity to engage learners while interacting with the content.

3. *Geneticus Investigatio*: A Problem-solving Learning Environment for Genetics

The pedagogical basis of GI gets inputs from guidelines in existing research in inquiry learning and problem solving, as well as from interviews we conducted with instructors teaching undergraduate level genetics. In literature, it has been documented that students have difficulties in identifying parts of the hypothesis, designing experiments, predicting the result and concluding about the hypothesis. Besides these, various topics in genetics, especially the topic of Mendelian genetics was perceived to be difficult by the students (Bahar 1999). We conducted a study with the undergraduate instructors where they solved an open-ended problem related to Mendelian genetics. They validated known students' difficulties. Besides this, the integration of concepts of genetics, statistical methods, and science process skills was also stated as difficult for the students. So there was a need to facilitate the integration of principles and concepts from multiple topics along with problem-solving skills, experimental abilities, and statistical methods.

A preliminary version of GI was developed earlier, which was primarily based on agent-based modeling and simulations (Deep, Murthy & Bhat 2016). In this environment, learners identified properties and behaviors of agents along with identifying rules governing the interaction between these agents. They then executed their model and compared their output with that of the expert model. A pilot study with this version of GI revealed that learners had difficulty in identifying the dependent and independent variable and designing a suitable experiment to test the hypothesis. Besides these, they were not able to make reasoned predictions which needed significant application of concepts of genetics along with understanding of concepts of statistics like Chi-square test, degree of freedom, calculating and comparing chi-square value with the critical value. These and other limitations necessitated the need for a revised version of GI which contained stronger supports for students, and a more accessible way for teachers to include the supports in the learning environment.

Figure 1 shows the key features of the GI. The overall sequence of learning activities in GI for the integration of domain, process skills, and statistics has been adapted from the steps of inquiry learning (Pedaste et al., 2015). These activities are interspersed with evaluative and reflective prompts for promoting metacognition during problem-solving (Xun & Land 2004). GI also require the learners to arrange the sequence of activities for experimenting as done in an authentic scenario.

3.1 *Pedagogical Design and Learning Activities*

The overall pedagogy of the revised version of GI contained learning activities focused on integration, and evaluative question prompts. An essential aspect was the reflective summarising activity along with scaffolds. They helped in implementing the following:

- *Integration of domain concepts, problem-solving process skills and statistical tools:*

It provided the overall structure to the learning environment and ensured that students should be able to integrate process skills along with the knowledge of genetics and statistics. Students began by defining the problem and identifying a suitable hypothesis to test (Pedaste et al., 2015). It requires the learner to understand the context, for example, the scientific phenomenon which is to be explained. To

explain the context, he/she select a hypothesis from the given set of hypotheses. In this, learning activities requires students to state the reason behind the selection of a particular hypothesis, state the assumptions they will make while testing this hypothesis along with declaring the dependent and independent variable. The system displays hypothesis and drag and drop activity for identifying the variables (Figure 2).

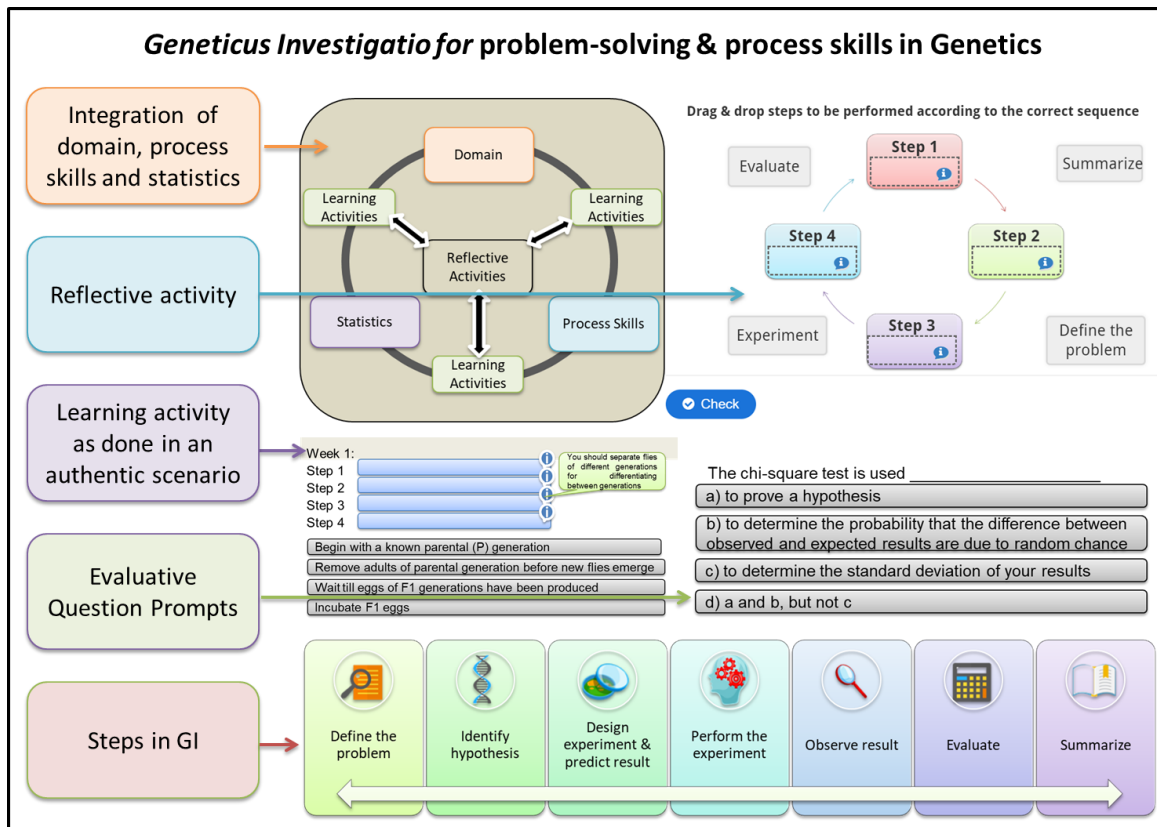


Figure 1. Key features of *Genetics Investigatio*

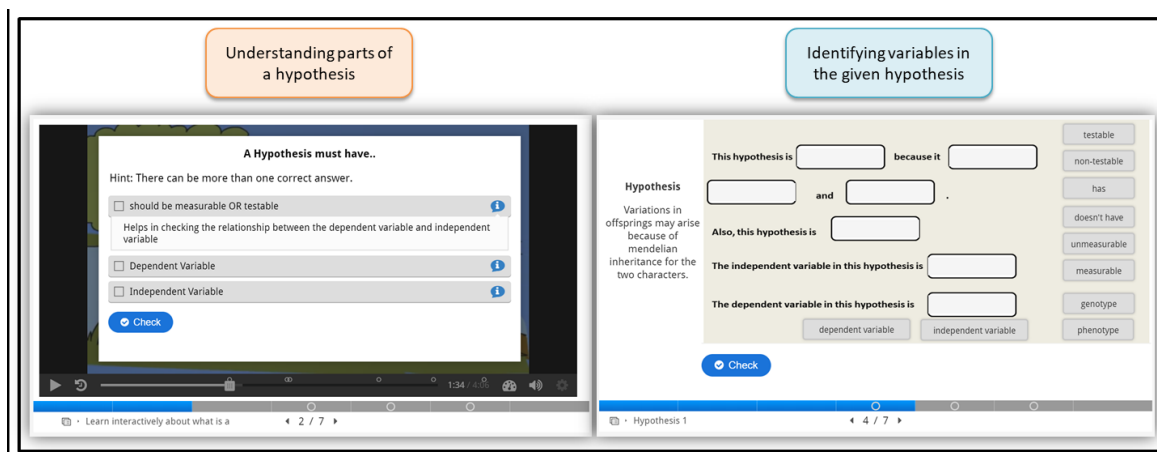


Figure 2. Learning activities in *GI* for selecting hypothesis

It was followed by the step of testing hypothesis, which included designing an experiment and reasoning from the hypothesis to predict the result. In this, learning activities requires learners to decide about the cross made, design the steps of breeding experiment, and calculate the predicted value (Figure 3). The system displays activities related to determining the cross made and calculating the ratio by providing editing boxes. These editing boxes are for stating laws of inheritance, creating Punnett

square and the calculating ratio of offspring. Besides this, learners interactively design steps of breeding experiments and watch a lab demo video of an actual experiment done in practical labs.

Figure 3. Learning activities in GI for testing hypothesis

The last step was to revise the hypothesis if required by comparing the result of the expected and observed values. The last learning objective has two goals, namely "Evaluating" and "Summarizing." Once the experiment is designed, and results are collected, it has to be statistically compared with the predicted outcome and come to a conclusion. In the evaluate phase, learners learn interactively about the Chi-square test, calculate the Chi-Square value, compare it with the critical value, and conclude based on critical value (Figure 4). The system displays interactive video which has reflective question prompts related to what, why, and how of chi-square and calculate the chi-square value by providing the functionality of editing boxes. Besides this learner reflect on the steps to be done while solving similar problems through the drag and drop activity.

Figure 4. Learning activities in GI for revising hypothesis

- **Evaluative Question Prompts:**

The integration of concepts, skills and use of statistics was interspersed with evaluative question prompts to reinforce the understanding of conceptual knowledge (A hypothesis must have) or to strengthen the application of conceptual knowledge (Did you think about the following while selecting the hypothesis?) (Xun & Land 2004).

- **Reflective activities:**

Students reflected upon the steps and sub-steps which has to be done while solving a similar scenario and the learning activity required them to arrange them in the correct order.

- **Scaffolds:**

Learners were provided with immediate feedback throughout the learning activities. Along with the feedback, hints were provided to scaffold learners in the problem-solving process. Learners were asked

to state their reasoning explicitly in many places, which ensured that they should take an informed decision during the interaction. In addition to that, additional resources related to concepts of genetics and statistics were provided in the form of video, pdf, and solved examples which can be accessed by the learners anytime during the problem-solving process.

These functionalities are incorporated at different steps within the GI. The interactions in the GI required the learners to navigate back and forth with interspersed drag and drop activities in the majority of the learning activities. This approach creates a seamless transition from guided problem solving as done in the traditional classroom to a personalized web-based learning environment to foster and practice problem-solving skill without limiting the solution space which is also aligned to the student's curriculum. The user interface of GI is designed and implemented with Google sites, which is an open-source toolkit. The learning activities of GI have been designed in H5P, which is an open source free HTML5 toolkit to develop interactive contents. H5P supports the creation of interactive learning activities where learners can interact with artifacts available in the environment. The users can access GI through standard web user interfaces through any device. Besides this, group level learning behaviors can be accessed in real time from the Google analytics platform, which is helpful for the teachers to provide real-time feedback.

4. Study Design

4.1 Participants & Context of Study

The participants of this study were sixty-three undergraduate students of Bio-Science at one of the colleges of Mumbai University in India. The participants were randomly assigned into control or experimental group. In this study, we chose the di-hybrid cross following the Mendelian inheritance as the context covered in the learning material. Problem-solving in this topic requires the students to generate, test and revise hypothesis along with connecting concepts related to basics of Mendelian genetics, understand and decide about the appropriate statistical calculation along with inference.

This topic is suitable to be implemented in GI because students are required to identify the underlying reason for biological phenomena, e.g. identifying the inheritance pattern of characters in *Drosophila*. In this topic, there could be multiple underlying reasons like incomplete dominance, co-dominance, etc. Associated with the given task, we expect that the students will have an understanding of the concepts of Mendelian genetics. They have learned it as part of their high school curriculum followed by an introductory genetics course in the first semester of their undergraduate degree. However, they were never made or asked to do open-ended problem solving related to the current topic. In the learning environment, students were provided with the additional resources related to concepts of Mendelian genetics.

4.2 Experimental Setting

This study was conducted as a part of a problem-solving workshop for Bio-science learners. It was conducted in a supervised setting using the GI learning environment for capturing data of the learning gains on problem-solving and their perception of usability of GI. The study had five steps, as presented in the figure 5.

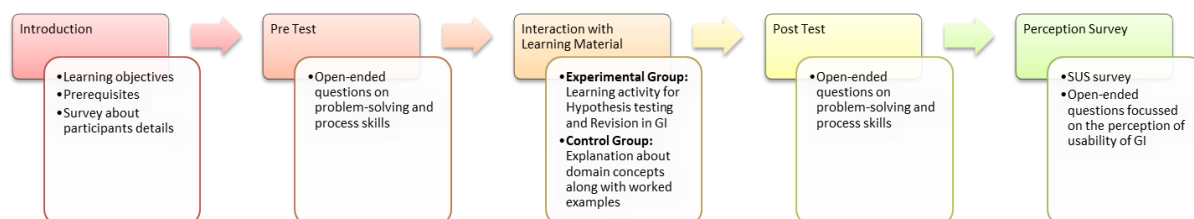


Figure 5. Procedure of Study

The first step required the participants to fill the registration form. It mentioned the learning objective of the workshop, the pre-requisites for the workshop, and questions related to the participant's academic details. The pre-test followed it on the day of the workshop. The pre and post-test questions were validated and checked for its reliability within the research team and subject matter experts. It had

six tiers of questions (Table 1). Once the participants attempted the first three questions, they were given the second sheet, which had the remaining questions. They were also provided with the additional resource which had the chi-square formula and table of the critical value of the chi-square distribution. On an average participant took 40 minutes to solve the pre-test.

Table 1: *Sample pre and post-test questions*

Q. No.	Targeted Problem-solving & Process Skills	Relevant part of question statement
1	Identify parts of hypothesis	Why do you think that it is a reasonable hypothesis? Which physical quantities or variables you will measure in this hypothesis?
2	Calculate the predicted value	Calculate the number and type of different progenies if total number of progenies after second generation is 320.
3	Justification for predicted value	State the assumptions, if any, that you are making while predicting the result.
4	Statistical comparison	Does your expected result match the observed result?
5	Decision of hypothesis	What will you do to determine if your hypothesis is correct or not based on the results?
6	Process to solve similar problem	Describe the sequence of steps which you will perform while solving a similar scenario with another organism.

The pre-test was followed by interaction with the learning material. This is where the experimental and the control group have a different activity. In the experimental group, participants interacted with GI, in which they do learning activities of selection, testing, and revision of hypothesis. On the other hand, the control group went through the learning material related to basic concepts of genetics, the importance of model organism, hypothesis formation and how to calculate the chi-square test and compare with the critical value. They also went through the worked examples. The main difference in the learning materials of the two groups were the features of evaluative question prompts with customized feedback, reflection activity, and the drag and drop activities. These learning materials for the control group were in the form of video, pdf, and Google slides, which were organized as Google website. These videos were the same as in the experimental group but did not contain the scaffolds and prompt which were present in GI.

After that, participants of both the groups took the post-test, which was similar to the pre-test. The workshop concluded with the last activity of filling the perception survey. It was implemented through the Google form with an aim for understanding participants' perception of usability and usefulness of GI. We took a traditional survey instrument for testing the usability of GI. For usability testing, we used the 10-item System Usability Scale (SUS) (Brooke, 1996) widely used for assessing the usability of a wide variety of learning environments. We asked additional open-ended questions in the survey to capture participant's perception of gross usefulness and usability of the GI. Target statements of these open-ended questions were as follows:

Q1: What features of the GI did you find most useful?

Q2: After interacting with GI, I learned something which I consider to be valuable. GI is valuable for

Q3: How do you plan to use the knowledge you obtained from this online workshop in other topics/subject or anywhere else? Please explain briefly.

5. Data Analysis and Results

5.1 The Effect of GI on Students' Learning Performance

We calculated improvements in learning by evaluating their pre and post-test based on adapted scientific ability rubric. The rubric items correspond to the problem-solving skill as given in Table 2. The inter-rater reliability was high (Cohen's Kappa: 0.774, p-value<0.001). We calculated the average mean value and standard deviation of the scores on sub-skills. On average, the final group score is increasing. Furthermore, we did a statistical test to see if the changes were significant. We did a paired-sample t-test, for both the control and experimental groups. The difference between the average

of the post-test score minus pre-test score and μ_0 is statistically significant (p -value: 0.000) for the experimental group. The observed standardized effect size is large (0.99). We also did an independent sample t -test on normalized gain.

Table 2: Rubric Item-wise Statistics of the Pre and Post-test Scores for Control and Experimental Groups

Problem-solving & Process Skills	Group	Pre-Test: Mean (SD)	Post-Test: Mean (SD)	Normalized gain	Paired t-test: Sig. p-value	Independent t-test: Sig. p-value
Identify parts of hypothesis	Control	1.03 (0.78)	0.86 (0.74)	-0.08	0.13	0.26
	Experimental	1.38 (0.6)	1.38 (0.65)	0	1.00	
Calculate the predicted value	Control	0.69 (1.20)	1.10 (1.05)	0.17	0.12	0.09
	Experimental	1.09 (1.00)	1.76 (0.85)	0.35	0.00	
Justification for predicted value	Control	0.34 (0.55)	0.07 (0.26)	-0.10	0.03	0.00
	Experimental	0.44 (0.75)	0.65 (0.81)	0.08	0.11	
Statistical comparison	Control	0.31 (0.47)	1.03 (0.98)	0.26	0.00	0.52
	Experimental	0.97 (0.9)	1.71 (0.94)	0.36	0.00	
Decision of hypothesis	Control	0 (0)	0.52 (0.83)	0.17	0.00	0.79
	Experimental	0.18 (0.63)	0.71 (1.00)	0.18	0.01	
Process to solve similar problem	Control	0.69 (0.71)	0.66 (0.77)	-0.01	0.81	0.00
	Experimental	0.53 (0.66)	1.41 (0.78)	0.35	0.00	
Total	Control	3.07 (2.55)	4.24 (3.27)	0.07	0.05	0.00
	Experimental	4.59 (3.06)	7.62 (3.46)	0.22	0.00	

5.2 The Perception of Usability and Usefulness of GI

We performed thematic analysis for analyzing the response to the three open-ended questions about gross usefulness and usability of GI. The result of our analysis is summarized in Table 3. Participants found the interactive video, question prompts for reflection, drag and drop learning activity and understanding of domain as useful features of GI. Analysis of the result of the two questions related to the usefulness of GI reveals that GI helps in learning of the skill of hypothesis testing and revision and learning of genetics concepts. The SUS survey responses were used to calculate the SUS score as per standard method (Brooke, 1996). The SUS score came to 63.35, indicating the product is usable.

Table 3: Themes and Respective Sample Excerpts

Q1. What features of the GI did you find most useful?			
S. No.	Theme	Meaning	Instance of responses from participants artefacts
A	Interactive video	The content of the video and interspersed reflective questions	"I liked the videos and the question answer format in between the videos"
B	Question prompts	Multiple choice questions for the purpose of reflections and embedded hints	"...the hints provided to solve the questions...."
C	Drag and drop activity	Drag and drop activity related to the steps of experimental processes.	"...the way it taught each and every detail about the experiment..."
D	Domain of genetics	GI helped in understanding of concepts related to genetics	"the genetic analysis was very useful"

Perception of Usability of GI			
Q2. After interacting with GI, I learned something which I consider to be valuable. GI is valuable for			
Q3. How do you plan to use the knowledge you obtained from this online workshop in other topic/subject or anywhere else? Please explain briefly			
S. No.	Theme (Learning of)	Meaning	Instance of responses from participants artefacts
A	Problem solving and process skills	Identifying the instances and applying the series of steps	"I plan to use it whenever I have problems in solving genetics....like first setting a hypothesis, comparing the expected and observed values and finally concluding the hypothesis"
B	Domain	Understanding and applying the concepts of genetics	"knowing the genetics concepts in more easy and interactive way and applying the learned knowledge"
C	Statistical concepts	Application of statistical concepts in a particular scenario	"solving the mathematical sums online" "can be used in biostatistics"
D	In advance studies	Apply the knowledge learnt in other topics in bio-sciences and in advanced studies	"as I aspire to complete my masters in genetics I found it really helpful and yes it cleared my basics"

6. Discussion and Future Work

Overall the experimental group demonstrated high learning gain in the application of problem-solving skills as compared to the control group. The most likely explanations for this observation would be that the conjectured design features of GI were useful for the learning of problem-solving skills. It is supported by the fact that the participants were able to identify and state the interactive design features as found in the thematic analysis of the feedback questions. Some of the design features which are worth mentioning are interactive video, reflective and evaluative question prompts and the drag and drop activity.

Rubric wise analysis of learning gains reveals that interaction with both the control and experimental learning material resulted in higher effectiveness in teaching the application of procedural steps like calculation of the predicted value and making the statistical comparison. Stated differently, the participants seemed to understand the application of procedural steps. This is not entirely unexpected as they are used to the kind of teaching method in which the teacher, demonstrates the steps and the students mechanically apply those steps in similar problems. In GI, instead of the teacher these participants watched the interactive video explaining steps of Chi-square calculation. In contrast to the learning gain of procedural steps, the learning gain of process skills, e.g. justification for the predicted value and decision for the hypothesis was not significant for both the groups. This result is following the findings reported in the existing literature that learning of skills requires multiple interactions over some time (Kim & Hannafin, 2011). Based on the result, we conjecture that multiple interactions with GI and across contexts will lead to significant learning gain.

High learning gain in the experimental group for learner's response to the last question which was about steps required to solve similar problems revealed that GI helped them to identify and reflect on steps and sub-steps performed for similar contexts. This result is worth discussing as this high learning gain in this question could be attributed to the summarizing activity. In this activity, since participants had to reflect upon overall learning activity and its sub-activities, they were able to abstract the steps of problem-solving. The drag and drop activity in GI provided them the flexibility of sequentially arranging the steps and access the hints as and when required. We conjecture that because of this, they were able to demonstrate the skill explicitly in the answers of the post-test questions. Along with that, thematic analysis of open-ended question responses revealed that GI is valuable for the learning of process skills of hypothesis testing and revision. Besides this, they also perceive that interaction with GI will help in better understanding of genetics and statistical concepts and will help in advance studies. GI was marked as a usable product based on the SUS score. Repeated use of the tool is likely to boost their confidence in interacting with the tool. The web-based learning environments make learning flexible, portable, and attractive (Hashemi, Azizinezhad, Najafi, & Nesari, 2011).

Our next development work includes analysis of the users' feedback to modify or implement additional functionalities in the learning environment. We would also like to validate the conjecture of cognitive process performed by the students while interacting with GI through eye-tracking and identify the pedagogical and interface design changes needed as part of our future work.

Acknowledgements

We would like to thank Kavya Alse, Rumana Pathan and Ajit Mhatre of IDP-ET, IIT-Bombay for helping in the implementation of this research study. We thank Prof. Moitreyee Saha of VPM's B. N. Bandodkar College of Science, Thane, Mumbai for helping us with this research study.

References

- Bahar, M., Johnstone, A. H., Hansell, M. H. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, 33(2).
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Chen W., Looi C., Xie W., & Wen Y., (2013) Empowering argumentation in the science classroom with a complex CSCL environment. In *Proceedings of the 21st International Conference on Computers in Education*. Indonesia: Asia-Pacific Society for Computers in Education.
- Concord Consortium. (2010). Geniverse (Accessed: 16 May. 2019). Geniverse [Online]. Available: <https://geniverse-lab.concord.org>
- Cooper, S., Hanmer, D., & Cerbin, B. (2006). Problem-solving modules in large introductory biology lectures. *The American Biology Teacher*, 68(9), 524-530.
- Deep, A., Murthy, S., & Bhat, P. J. (2016, December). Geneticus Investigation: An Agent-Based Modeling System for Teaching-Learning Hypothetico-Deductive Reasoning in Mendelian genetics. In *2016 IEEE Eighth International Conference on Technology for Education (T4E)* (pp. 164-171).
- De Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: the Go-Lab federation of online labs. *Smart Learning Environments*, 1(1), 3.
- Fernandes, R., & Simon, H. A. (1999). A study of how individuals solve complex and ill-structured problems. *Policy Sciences*, 32(3), 225-245.
- Hashemi, M., Azizinezhad, M., Najafi, V., & Nesari, A. J. (2011). What is mobile learning? Challenges and capabilities. *Procedia-Social and Behavioral Sciences*, 30, 2477-2481.
- Ibáñez Orcajo, T., & Martínez Aznar, M. (2005). Solving problems in genetics II: Conceptual restructuring. *International Journal of Science Education*, 27(12), 1495-1519.
- Jouble (2013). h5p.org (Accessed: 20 May. 2019). H5P [Online]. Available: <https://h5p.org/>
- Karagoz, M., & Çakir, M. (2011). Problem Solving in Genetics: Conceptual and Procedural Difficulties. *Educational Sciences: Theory and Practice*, 11(3), 1668-1674.
- Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education*, 56(2), 403-417.
- Makransky, G., Bonde, M. T., Wulff, J. S., Wandall, J., Hood, M., Creed, P. A., ... & Nørremølle, A. (2016). Simulation based virtual learning environment in medical genetics counseling: an example of bridging the gap between theory and practice in medical education. *BMC medical education*, 16(1), 98.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., ... & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational research review*, 14, 47-61.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., ... & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences*, 13(3), 337-386.
- Slotta, J. (2002). Designing the web-based inquiry science environment (WISE). *Educational technology*, 42(5), 15-20.
- Thompson, N., & McGill, T. J. (2017). Genetics with Jean: the design, development and evaluation of an affective tutoring system. *Educational Technology Research and Development*, 65(2), 279-299.
- Xun, G. E., & Land, S. M. (2004). A conceptual framework for scaffolding III-structured problem-solving processes using question prompts and peer interactions. *Educational technology research and development*, 52(2), 5-22.
- Yin, C. J., Uosaki, N., Chu, H. C., Hwang, G. J., Hwang, J. J., Hatono, I., & Tabata, Y. (2017, January). Learning behavioral pattern analysis based on students' logs in reading digital books. In *Proceedings of the 25th International Conference on Computers in Education* (pp. 549-557).

Augmented Reality in Education: Three Unique Characteristics from a User's Perspective

Jule M. KRÜGER^{a*}, Alexander BUCHHOLZ^a & Daniel BODEMER^a

^a*Media-based Knowledge Construction, University of Duisburg-Essen, Germany*

*jule.krueger@uni-due.de

Abstract: In this paper, three technological characteristics of augmented reality (AR) are reframed from a perceptual, user's perspective and discussed concerning their potential for education and in the context of research on technology-supported learning. The first characteristic, *contextuality*, describes that users of AR can experience the real world and virtual elements simultaneously. The second characteristic, *interactivity*, includes the possibilities to interact with AR through the manipulation of both real objects and virtual properties, which offers novel possibilities for interaction. The third characteristic, *spatiality*, focusses on the linking of virtual objects to specific points in space and the more realistic three-dimensionality that AR visualizations offer. It is proposed that these three characteristics can provide a way to structure the broad research landscape of AR in education and form a basis for future research projects. Two studies are presented and linked to the three characteristics. In the first study, the comparison of a desktop simulation and an AR simulation in an individual learning setting is linked to the characteristics of interactivity and spatiality. In the second study, the contextuality of AR is systematically varied and exploited to present group awareness information about other learners next to these learners instead of separated from them. The results of the studies are discussed in the context of the three characteristics and the paper concludes that there are a lot of different educational settings in which AR could be beneficial. The classification of and systematic variation in research based on the three characteristics may form a basis to systematize educational AR research. Furthermore, the results of this research and the three characteristics themselves can inform the design of AR applications to support learning.

Keywords: Augmented reality, Contextuality, Interactivity, Spatiality, Technology-supported education, Multimedia learning

1. Introduction and background

During the past centuries, augmented reality (AR) has turned from a technological vision of the future, which could often be found in science fiction movies, to a technological achievement of the present, which can now be created by the smart technological devices we have in our pockets. This development concerning the access to the necessary technology creates novel opportunities for applying AR in different fields. One area that many recent studies concerning AR focus on is education (Cipresso, Giglioli, Raya, & Riva, 2018). Education may also be one of the most promising areas for applying AR and there is an increasing number of studies that focus on the opportunities that AR as a way of visualizing information has to offer for both individual and collaborative learning settings (Akçayır & Akçayır, 2017; Phon, Ali, & Halim, 2014; Radu, 2014; Wu, Lee, Chang, & Liang, 2013). In most of these studies, advantages of AR in comparison to more traditional learning settings are examined. Positive effects that have been found when using AR in education are enhanced learning performance and motivation, higher enjoyment and engagement, more positive attitudes towards the learning material, and a better collaboration between learners (Akçayır & Akçayır, 2017; Bower, Howe, McCredie, Robinson, & Grover, 2014; Chen, Liu, Cheng, & Huang, 2017; Dunleavy & Dede, 2014; Phon et al., 2014; Radu, 2014; Saidin, Halim, & Yahaya, 2015; Wu et al., 2013). Challenges that were discovered are for example technical limitations, the use of the application being too complicated and mentally overloading, the amount of time that has to be invested to develop the applications, and pedagogical issues when trying to integrate AR into the classroom (Akçayır & Akçayır, 2017; Bower et al., 2014; Dunleavy & Dede, 2014; Radu, 2014).

Over the years, various definitions of AR have been used in different areas of research. A rather general definition describes AR as “technology which overlays virtual objects (augmented components) into the real world” (Akçayır & Akçayır, 2017, p. 1) and in earlier definitions, AR is often linked to head-mounted displays, which were the preferred display devices before smartphones and tablets were available (Azuma, 1997). One of the most commonly used definitions by Azuma (1997) defines AR as systems with three characteristics: (1) combination of the real world and virtual elements, (2) real-time interactivity, and (3) registration in 3D. The definition is used in papers by Azuma (1997) and Azuma et al. (2001), which are the two most cited papers in AR as of 2016 (Cipresso et al., 2018). This underlines the importance of this definition and those three characteristics in AR research. In the current paper, Azuma’s definition is employed because of its use in the educational field (e.g., Bower et al., 2014; Radu, 2014), its scope (not too broad or too narrow), and its independence of a technological device. In addition to different definitions, there have also been various attempts to classify AR applications and technologies (see Normand, Servières, & Moreau, 2012 for an overview). In the most known taxonomy, the Reality-Virtuality Continuum, AR is placed between the two extremes of real and virtual environment, leaning towards the side of the real environment (Milgram & Kishino, 1994). A newer taxonomy by Normand et al. (2012) classifies AR applications based on four axes, namely tracking (degrees of freedom and accuracy), augmentation type (optical see-through, video see-through, spatial augmentation), temporal base (past, present, future, time independent) and rendering modalities (beyond visual augmentation). With this taxonomy, AR applications can be classified depending on their goal and independent of the technology or the device used (Normand et al., 2012).

While the different definitions and taxonomies are often used independently of the research area, in the educational AR literature there have been attempts to connect AR to different learning theories and pedagogical approaches. Bower et al. (2014) and Dunleavy and Dede (2014) connect AR to situated and constructivist learning by assessing that learning with AR can take place at a relevant location and a deeper learning can occur with the support of AR. Game-based learning, in which immersion in the learning material is important, and inquiry-based learning, in which a scientific data gathering process is enacted, are also mentioned in connection to AR (Bower et al., 2014). In a review of the usage of learning theories to support the design of educational AR applications, Sommerauer and Müller (2018) mainly found that Mayer’s multimedia principles from his Cognitive Theory of Multimedia Learning (Mayer, 2009), situated learning, game-based learning and simulations, and experiential learning were used in studies. Based on their findings, they furthermore developed a design framework that can be used for designing educational AR applications (Sommerauer & Müller, 2018). While research on AR in education has been conducted for some time now, it is still not completely obvious how exactly AR is better for supporting learning than other learning technologies like tablet-based simulations or desktop learning environments. One key affordance of AR that Bower et al. (2014) mention is that with AR, students can rescale virtual objects of all sizes in order to better understand them. It is, however, not evident, how this is better than executing the same action on a tablet or desktop screen. Affordances of AR that are mentioned by Wu et al. (2013) and might also be true for technologies other than AR (for example a normal smartphone app), are ubiquity and situatedness, the visualization of the invisible and the bridging of formal and informal learning. Although it is evident that these affordances all have the potential to support learning, it is not completely clear how exactly AR as a form of visualizing information plays a unique role in them. That is why, in the remainder of this paper, we aim to present and discuss three characteristics of AR that have been identified to be important factors in supporting learning. We describe how they are in this specific way only found in AR and not in other learning technologies, and thus reveal unique values that AR has for education, as proposed by Wu et al. (2013). Furthermore, we suggest that these three characteristics might provide a structure and a focus for educational AR research, to examine when and how the implementation of AR is most beneficial for education. This may help to develop a systematic research agenda for the use of AR in education scenarios and thus also support instructors and designers in developing effective AR-based learning experiences for various target groups and learning objectives in formal and informal learning settings. After the introduction and discussion of the three characteristics from a user’s perspective in the next section, two studies that have been conducted on AR-supported learning are presented and discussed in the context of the characteristics. These studies exemplify how the three characteristics can be used for classifying and planning empirical research. A conclusion for the three characteristics and future research is drawn at the end of the paper.

2. Three Characteristics of AR from a User's Perspective

As stated by Hugues, Fuchs, and Nannipieri (2011), augmenting reality in itself is not possible, so that in AR a person's perception of reality is augmented. Therefore, we chose to look at the characteristics that AR possesses from a perceptual, user's perspective. In order to do this, we considered the three characteristics in the aforementioned definition of AR by Azuma (1997): (1) combination of the real world and virtual elements, (2) real-time interactivity, and (3) registration in 3D. The technology that delivers the AR experience to the user must possess these properties. In order to reframe the characteristics from a user's perspective, we looked at how they affect the user's experience of AR and propose three characteristics of the experience of using AR that cannot be found in this specific form in other technologies: contextuality, interactivity, and spatiality. In the following paragraphs, these three characteristics are described and their value for technology-enhanced learning is discussed. Also, four interesting research areas are given for each of the characteristics: two concerning individual learning, and two concerning collaborative learning. Table 1 shows an overview of the three characteristics.

2.1 Contextuality

In Azuma's (1997) definition, the first characteristic is that real world and virtual elements are combined in AR. From a technological perspective this means that virtual and real elements are displayed simultaneously. The displaying device must be context sensitive and aware of its location to show the user the digital content that is relevant at that place in that moment (Dunleavy & Dede, 2014). When looking at this characteristic from a perceptual, user's perspective, this means that the user perceives the displayed virtual elements (e.g., objects, pictures, text) in the context of the real world around them (e.g., physical objects, other learners). In contrast to virtual reality, the context is not completely covered by the virtual elements, and in contrast to information on a screen, the virtual element and the context are not separated from each other (Rekimoto & Nagao, 1995). With this, novel opportunities and challenges to link the context and the virtual elements appear. Therefore, the first AR-specific characteristics reframed from a user's perspective is "contextuality".

Concerning the benefits that contextuality has for learning, it can be said that with AR it is possible to situate learning in a relevant context, which may increase the authenticity and ground students in reality (Wu et al., 2013). Even though it may also be possible to look up information that is relevant to the place where the user is at that moment with mobile devices, in AR the possibility to overlay visual virtual information over the environment gives additional potential for "perfectly situated scaffolding" (Bower et al., 2014, p. 6). Here, the relationship between the real world and the virtual information is closer than when just looking at relevant information on a mobile device. Bower et al. (2014) call the ability to contextually overlay information onto the real world one of the key pedagogical affordances of AR and Dunleavy and Dede (2014) state that embedding learning within relevant environments is very likely to enhance learning. In scientific literature, there is furthermore a connection made between the contextuality of AR and Mayer's (2009) multimedia principles of spatial and temporal contiguity (Akçayır & Akçayır, 2017; Radu, 2014). Through contextuality, instructional information can be made available at the right place and time and can this way be situated inside the real world. This implements the contiguity principles, which state that information that belongs together should be presented in an integrated way and at the same time (Mayer, 2009) in order to avoid split attention and thus increased cognitive load (Ayres & Sweller, 2014). When working in a collaborative learning setting, the contextuality of AR can also be beneficial. In co-located collaboration, contextuality means that because the virtual elements do not occlude other learners and the context, virtual information can be added to face-to-face collaborative learning settings. Learners can then perceive virtual information, the other learners, and the context around them at the same time. Here, it must be considered that a complex interplay between the three elements takes place, which might have an influence on the collaboration between the learners and their references to learning material or other external artifacts (see Bodemer, Janssen, & Schnaubert, 2018; Stahl, 2006). In general, through the characteristic of contextuality, AR has the potential to apply some of the multimedia principles onto the real world and support especially the situating of learning in a relevant environment. This provides interesting opportunities for applying AR to support learning both inside and outside the classroom.

Different questions concerning the contextuality of AR that still need to be answered through empirical research are, for example: (a) Do people indeed learn better when they are in a relevant context than when they are not and which (cognitive, motivational, and emotional) factors play a role in this?, (b) How closely must the context and the virtual information be thematically related for the overlaying of information to be beneficial?, (c) How does the interplay between learners, contexts, and virtual material have an influence on the interactions between two or more learners learning collaboratively?, (d) What are the advantages and challenges of placing group awareness information (see Bodemer et al., 2018) about other learners directly next to the respective learner? Concerning this last question, a study is presented later in this paper (study 2).

2.2 *Interactivity*

The second characteristic that Azuma (1997) mentions in his definition of AR is that AR elements are interactive in real time. From a technological perspective this means that the elements must be programmed to react to input that the user or – in a collaborative setting – the users give.

From a perceptual, user's perspective this entails that users experience the virtual elements reacting to their and other learners' actions. In turn, all users can react to the element's actions. In AR, virtual elements have two interactive sides. Because virtual objects in AR are placed inside the real world, they lend themselves to natural and intuitive interaction that is not possible with screen-bound virtual objects (e.g., "real" touching, gesture-based interaction). On the other hand, users can manipulate the virtual AR objects in other ways than purely physical objects (e.g., input of new data to change simulations, control through input devices) and can receive realistic and immediate feedback upon their input. This way, the interactive capabilities of real and virtual elements are combined in AR. Billingham and Dünser (2012), for example, state that in AR books, different forms of interaction are possible, like turning real pages to change the virtual scenery or tilting and rotating the pages to view the virtual elements from different angles. Hence, users can interact with the digital content by manipulating real objects, using a tangible interface metaphor. Therefore, a second AR-specific characteristic reframed from a user's perspective is its "interactivity".

Concerning the benefits that interactivity has for learning, it was found that even the most intuitive form of interaction with an object (i.e., perspective changing by walking around it) can be advantageous for learning (Holmes, Newcombe, & Shipley, 2018). Following embodied cognition theory, whole-body interaction with AR learning material can also lead to better learning outcomes (Johnson-Glenberg & Megowan-Romanowicz, 2017). Concerning collaborative learning settings, it can be said that in AR all learners can interact with the virtual elements in the same way and can watch how other learners interact with them. With other learning technologies, one person controls the mouse and keyboard and others watch, or everybody uses their own device to collaborate online. In AR, learners and their actions can directly be linked to each other, which may support the forming of a mental model of the other learners and thus group awareness. In general, AR's interactivity provides interesting new ways to interact with learning material, supporting learning in different settings.

Questions that still need to be answered with empirical research concerning the interactivity of AR are for example: (a) How does AR-based interaction (using a tangible interface metaphor in which interaction with an AR marker in the real world leads to manipulation of virtual objects) have a different effect on learning especially the connections between objects in comparison to a more familiar touch-based interaction with virtual objects?, (b) How must interaction with the material be designed to evoke higher order thinking processes?, (c) What influence does the collaborative interaction with the AR material have on the interaction between learners?, (d) How does watching other group members interact with the material support understanding and for example grounding processes in the group?

2.3 *Spatiality*

The third characteristic mentioned in the definition is that virtual elements must be registered (i.e., placed) inside the 3D real world (Azuma, 1997). From a technological perspective this means that the real world must be tracked continuously, so that the virtual element can be pinned to a specific point in space. Also, the spatial specifics like the dimensionality of the element itself need to be defined. From a perceptual, user's perspective this means that the virtual elements should seem to exist in the same space as the real world. When virtual objects are placed inside the 3D real world, they can appear

to have more spatial depth than virtual objects shown purely on flat screens. Pseudo-spatial visualizations are possible when using monocular depth cues on AR flat screens, while even true spatial visualizations can be created with the aid of binocular disparity when using AR glasses (Jeřábek, Rambousek, & Wildová, 2015). The third AR-specific characteristic reframed from a user’s perspective, is thus its “spatiality”.

Concerning the benefits of spatiality in educational settings it can be said that physical 3D objects were found to be better for learning than 3D computer models (Preece, Williams, Lam, & Weller, 2013). When looking at the spatial properties of 3D AR models, they lie between physical and computer models, so that they may also be more beneficial for learning than normal computer models. Advantages concerning the mental load of participants using a 3D visualization to learn a visual motor task over using a 2D visualization could also be found (Dan & Reiner, 2017). AR might be especially useful for learning the spatial structure of 3D material (Radu, 2014) and subjects with a spatial component are learned more effectively with AR (Billinghurst & Dünser, 2012). In collaborative learning settings, an example of how the fixation of an AR object to a point in space can be used is through knowledge sharing by tagging and annotating objects (Specht, Ternier, & Greller, 2011). The objects over which the learners collaborate or which they create collaboratively can also be three-dimensional and fixed to one point in space. This may offer various advantages over working together on two-dimensional screen-based material. In general, it can be said that learners may especially benefit from AR’s spatiality when learning about spatial structures and relationships.

Questions that arise and should be answered through empirical research are for example: (a) Is using a three-dimensional AR object as beneficial for learning spatial structures as real objects are, in comparison to screen-based objects?, (b) How much does the use of stereoscopic AR glasses in comparison to screen-based monoscopic AR influence the spatial perception of an object and what are the advantages concerning the spatial understanding the user acquires about it?, (c) Does the collaborative creation of a three-dimensional artefact lead to better learning than the creation of a two-dimensional artefact?, (d) How exactly does using a whole room as a space to learn in together instead of a shared screen influence the interaction with the material and between the learners?

Table 1

Three Characteristics of AR from a User’s Perspective

Azuma’s characteristic	User perspective characteristic	Description
Combination of the real world and virtual elements	Contextuality	<ul style="list-style-type: none"> ▪ users perceive virtual elements simultaneously with real world (including other users) around it ▪ users do not perceive virtual elements and context (including other users) separately
Real-time interactivity	Interactivity	<ul style="list-style-type: none"> ▪ users experience virtual elements reacting to them and other users, and experience themselves and other users reacting to actions of the elements ▪ interactive properties of physical AND virtual elements
Registration in 3D	Spatiality	<ul style="list-style-type: none"> ▪ virtual elements placed inside the 3D real world appear as if they were really there ▪ virtual elements appear more spatial than if shown on screen

2.4 Interplay of the three characteristics

The three characteristics of AR and their advantages for educational settings are not only interesting on their own, but also in their combination into one experience. Moving around a virtual AR object and looking at it from all perspectives, for example, concerns both interactivity and spatiality of AR. When the object stays in one place, it reacts to the user’s movement (interactivity), which is possible because the object is fixed to a point in 3D space (spatiality). The authenticity of an experience can also be influenced by all three characteristics. Authenticity can imply the placement of a virtual object in a relevant, authentic environment (contextuality). It can also refer to the authenticity of the object itself, including its 3D presentation (spatiality). Furthermore, authenticity may imply authentic interaction

with the virtual object (interactivity). An authentic virtual object placed in a relevant, authentic real-world environment and with authentic interactive properties, may provide the most authentic experience for learners.

This shows that the three characteristics cannot always be considered separately but can interact with each other. It is important to examine them through experimental research both separately and in interaction, to get an overarching picture of how AR can be used best in educational settings. In the following sections, we present two experimental studies that we executed concerning the use of AR in different educational settings: study 1 as an example of considering different characteristics (interactivity and spatiality) in an individual setting, study 2 as an approach of systematically varying one of the proposed characteristics (contextuality) in a collaborative setting. This way, two quite different ways of using the characteristics to structure and design empirical research are presented.

3. Study 1: Interactivity and spatiality in an individual setting

The first experimental study is based on research about learning with computer simulations. Using computer simulations paired with inquiry-based learning instructions like scientific discovery learning proved to be valuable in many ways for the learner to comprehend complex concepts in research contexts and practical applications (de Jong, 1991; de Jong & van Joolingen, 1998). AR applications for learning purposes can also be understood as (interactive) computer simulations or visualizations, but research about learning with AR applications rarely explored the fact that traditional and AR simulations share common concepts but differ in various aspects. It is unclear whether the learning benefits in working with AR applications found in these studies were due to the AR aspect of the application or because the learning material was a simulation or interactive visualization instead of traditional paper and text. The aim of this study was to compare a traditional (tablet-based) computer simulation with an AR version of the application with regards to their effects on conceptual knowledge, cognitive load, motivation, and spatial abilities of the learners. Although the study was not planned based on the three proposed characteristics, when comparing the AR and non-AR applications used, it shows that interactivity and spatiality differ between them. Concerning interactivity, it can be said that while in AR the users moved around the simulation and interacted with a real object (AR marker) to manipulate it, in the traditional simulation they used touch-based drag-and-drop on a tablet. Spatiality differed in the two applications in that virtual AR objects appear to be more spatial because the user has the reference of the real world, while this is not the case in a normal screen-based simulation.

3.1 Method

For this study, two almost identical computer simulations were developed and compared in an experimental laboratory setting: a normal computer simulation of a power plant on a tablet, and an AR simulation with AR markers and the tablet as a video-see-through display for AR elements. The two simulations differed regarding their interactivity and spatiality as described in the previous paragraph. During the experiment the participants (N = 56) followed a scientific discovery-based learning script with the goal of comprehending the underlying concept of power plants by building their own, changing the composition of the plant components, and first hypothesizing and then observing the outcome. The participants were randomly assigned to use either the traditional (nt = 28) or the AR simulation (nAR = 28). It was hypothesized that after the interaction with the material, participants have equivalent conceptual knowledge and cognitive load during the learning process as well as improved spatial abilities and motivation when learning with the AR simulation compared to the traditional simulation. Based on this, three TOST equivalence tests and five t-tests were executed to analyze the data.

3.2 Results

The equivalence tests were all executed for the equivalence bounds Cohen's $d = +/-0.67$, based on the smallest detectable effect with this sample size. The hypothesis that conceptual knowledge was equivalent in the two simulations could be supported ($Mt = 12.79$, $SDt = 3.06$; $MAR = 12.64$, $SDAR = 2.84$), 90% CI for d [-0.40;0.49], lower bound, $t(54) = 2.69$, $p = .005$, upper bound, $t(54) = -2.33$, p

= .012. An equivalence of intrinsic cognitive load in the simulations was also found ($M_t = 4.88$, $SD_t = 1.95$; $MAR = 5.02$, $SDAR = 1.98$), 90% CI for d [-0.52;0.38], lower bound, $t(54) = 2.25$, $p = .014$, upper bound, $t(54) = -2.76$, $p = .004$. For extraneous cognitive load, equivalence in the simulations could not be concluded ($M_t = 1.21$, $SD_t = 1.35$; $MAR = 1.58$, $SDAR = 1.42$), 90% CI for d [-0.72;0.18], lower bound, $t(54) = 1.50$, $p = .070$, upper bound, $t(54) = -3.52$, $p < .001$. Based on three t-tests, no significant differences were found between the groups for these three variables.

The hypothesis concerning the difference in the resulting spatial abilities was not supported, as no significant difference between the traditional ($M_t = 7.96$, $SD_t = 4.15$) and the AR simulation ($MAR = 9.07$, $SDAR = 5.00$) was found, $t(54) = -0.90$, $p = .371$, $d = -0.25$. Contrary to expectations, motivation did also not differ between the two forms of simulation: intrinsic motivation ($M_t = 5.51$, $SD_t = 1.18$; $MAR = 5.78$, $SDAR = 0.94$), $t(54) = 0.94$, $p = .352$, $d = 0.26$, identified regulation ($M_t = 4.61$, $SD_t = 1.30$; $MAR = 5.14$, $SDAR = 0.99$), $t(54) = -1.73$, $p = .089$, $d = -0.47$, external regulation ($M_t = 4.69$, $SD_t = 0.98$; $MAR = 4.46$, $SDAR = 1.00$), $t(54) = 0.88$, $p = .382$, $d = 0.24$, and amotivation ($M_t = 2.58$, $SD_t = 1.18$; $MAR = 2.19$, $SDAR = 1.18$), $t(54) = 1.25$, $p = .217$, $d = 0.34$.

3.3 Discussion

The results of this study indicate that just transferring a desktop simulation into an AR simulation and thus manipulating interactivity and spatiality together might not be enough to be more beneficial for the learner regarding conceptual knowledge, motivation, cognitive load and spatial abilities. After using the application, the participants learning with the AR simulation had equal conceptual knowledge and intrinsic cognitive load and nearly equal extraneous cognitive load as the participants using the traditional simulation. The groups did not differ in motivational aspects and spatial abilities. Still, this experiment can serve as an initial study to find out more about how the three characteristics influence learning. In this study, both interactivity and spatiality were manipulated in the applications. To find out more about the specific benefits the two characteristics and their interaction have on learning processes and outcomes, more systematic studies are necessary in which interactivity and spatiality are varied separately. Furthermore, AR offers other possibilities than the ones varied in this study. A procedural simulation or visualization where the learner can use the application directly in the environment where the knowledge domain is registered (based on the characteristic of contextuality) might be more beneficial to the learner regarding learning outcomes and learning related variables. This also requires more research in the form of an experiment with systematically manipulated predictor variables.

4. Study 2: Contextuality in a collaborative setting

A further experimental study that was systematically planned and executed based on one of the three characteristics has focused on how to use the potential of AR's contextuality in a collaborative setting. Due to contextuality, the user can perceive virtual information, other learners, and the environment simultaneously. This way, virtual information can be shown exactly at the right time and place. As suggested by Radu (2014), this characteristic can be connected to Mayer's (2009) multimedia principles of spatial and temporal contiguity which state that information that belongs to each other should be presented at the same time and close to each other, preventing the splitting of attention and decreasing extraneous cognitive load. In computer-supported collaborative learning (CSCL), group awareness tools (GATs) can be used to support collaborative learning processes (Bodemer et al., 2018). As GATs provide contextual information about the social learning environment, it is crucial that they do not divert attention from germane learning activities. When group awareness (GA) information about other learners is visualized in face-to-face collaborative settings, this information is often printed out or shown on a screen, which means that the given information is separated from the context in which it is relevant (i.e., the collaboration with the other person) due to the medium that delivers it. This could especially be a problem in bigger groups of learners, because the correct GA information must still be connected to the right person. AR's unique characteristic of contextuality provides the opportunity to show GA information directly next to the corresponding person. Similar to the work of Holstein, Hong, Tegene, McLaren, & Alevan (2018), where teachers were provided with real-time information about their students' learning process through augmented reality glasses, this GA information could be

presented directly over or next to the corresponding student. In this study, the systematic variation in the two conditions was thus based on contextuality so that in the AR condition the information and the context were integrated, while in the non-AR condition they were separated from each other. The aim of the study was to find out whether placing information about people directly next to them in comparison to placing it further away has an influence on cognitive load and retention of the information.

4.1 Method

To compare the visualization of GA information next to people and further away from them, we used pictures instead of a real implementation in AR to investigate the characteristic of contextuality in a controlled laboratory setting. In the study, the participants ($N = 38$) worked on tasks in which they had to form study groups of the people shown to them in pictures based on the GA information given about them. The participants were randomly assigned to one of two conditions: GA information visualized directly next to the corresponding person in the picture (AR mockup; $n_{AR} = 18$) or GA information shown separately below the pictures ($n_{nonAR} = 20$). In the different tasks given to the participants, the number of people shown to them was varied between two and ten people to see if an effect of the proximity of the information differs with a differing number of people. The two independent variables were thus the proximity of the information to the people (between-subject) and the number of people displayed (within-subject). It was hypothesized that these two factors and their interaction influence the cognitive load of the participants as measured continuously through a secondary reaction task and the efficiency in executing the task as measured by their time spent on the task. Furthermore, it was expected that the proximity of the information influences the participants' self-reported extraneous cognitive load and their recall of the GA information. Two mixed-design ANOVAs and two independent samples t-tests were used to analyze the data based on these hypotheses.

4.2 Results

The hypothesis that the proximity of the information has an influence on the continuously measured cognitive load could be supported with a significantly slower reaction time (ms) in the group where picture and information were shown further apart ($M_{nonAR} = 2197.83$, $SD_{nonAR} = 1863.29$; $M_{AR} = 1153.49$, $SD_{AR} = 744.64$), $F(1,36) = 4.93$, $p = .033$, $\eta^2 = 0.12$. The same pattern was found for the time spent on the task, where the group with the separate information presentation needed more time (s) to solve the tasks than the group with the integrated visualization ($M_{nonAR} = 62.86$, $SD_{nonAR} = 19.76$; $M_{AR} = 50.93$, $SD_{AR} = 11.11$), $F(1,36) = 5.11$, $p = .030$, $\eta^2 = 0.12$. Concerning the within-subject factor (number of people), it can be said that even though more people shown generally meant both a longer time spent on the task, $F(1.24, 44.44) = 21.11$, $p < .001$, $\eta^2 = 0.37$, and a longer reaction time in the secondary task, $F(2.531, 91.13) = 4.933$, $p = .005$, $\eta^2 = 0.12$, this pattern was not found for all pairwise comparisons. No significant interaction effect was found for reaction time, $F(2.53, 91.13) = 1.47$, $p = .233$, $\eta^2 = 0.04$, or time spent on the task, $F(1.24, 44.44) = 1.25$, $p = .279$, $\eta^2 = 0.03$.

Concerning the variables that were not measured for every single task, no significant difference was found in either self-reported extraneous cognitive load ($M_{AR} = 4.07$, $SD_{AR} = 1.68$; $M_{nonAR} = 4.25$, $SD_{nonAR} = 1.67$), $t(36) = -0.32$, $p = .748$, $d = -0.11$, or recall of the GA information between the two groups ($M_{AR} = 2.39$, $SD_{AR} = 1.29$; $M_{nonAR} = 2.20$, $SD_{nonAR} = 1.11$), $t(36) = 0.49$, $p = .630$, $d = 0.16$.

4.3 Discussion

In this study, in which contextuality was varied systematically, significant differences between the two groups concerning the reaction time in a secondary task and the time on task were found. The participants in the AR mockup group needed less time for solving the tasks and reacted faster on the secondary task, which shows that they were more efficient and less cognitively occupied in their task of forming study groups based on the information about the people. However, these results could not be supported by the results in the self-reported cognitive load and recall of the information, which did not differ between the groups. A confounding variable that might have led to the differences in the timings between the groups was that the participants from the non-AR group had to scroll down on the pages with the tasks, while the others did not. In a future study, this factor must be held stable between the

groups. Also, other objective measures for cognitive load, which should not be influenced by scrolling (e.g., eye-tracking metrics), might be used to compare the two forms of visualization in a future study. A factor that may have led to less differences between the groups is that the tasks could be solved without even looking at the pictures of the people. This way, the participants might not even have made the connection between the people and the information. Split attention only happens when one part of the material is not understandable without the other (Ayres & Sweller, 2014). This was not the case here and an adapted study design should be considered for future studies.

5. Conclusion

In this paper, three characteristics of AR are reframed from a user's perspective and discussed in relation to their potential for supporting individual and collaborative learning. It is proposed that these three characteristics can be used as a basis for researching AR in educational settings and two studies which have been executed with the three characteristics in mind are presented.

The two studies differed considerably in their usage of the characteristics. In study 1, the experimental manipulation can be classified into two of the characteristics, namely interactivity and spatiality. Concerning this study, we conclude that to get a more complete picture, follow-up studies are necessary in which the two characteristics are varied separately and systematically. This way, their influence on learning processes and outcomes can be determined. In study 2, a systematic experimental variation based on the characteristic of contextuality took place and positive effects on efficiency and cognitive load could be found. Due to confounding variables, the results of the study should be interpreted with caution. Follow-up studies that control for these factors are needed to confirm the results concerning the increased efficiency and decreased cognitive load in the setting.

While contextuality, interactivity, and spatiality all seem to be important for using AR in educational settings, more systematic empirical research concerning their potentials, their impact and their interplay is necessary. Based on the two presented studies, which initialized the research on AR in education at our lab, more empirical studies with systematic variations based on the three characteristics are currently conducted and planned, such as two experimental studies that intend to systematically disentangle the characteristics of interactivity and spatiality.

AR-supported learning experiences have the potential to be applied in different settings and with various goals, which can also be seen in the differences between the two presented studies. Thus, systematic AR-related research findings can enrich the design of formal and informal educational environments for individual and social learning of diverse students. In order to provide a structuring basis for this heterogeneous research field, the three characteristics contextuality, interactivity, and spatiality are proposed to serve as common denominators for the users' experience of AR in a wide range of learning settings.

Acknowledgements

We would like to thank Sophie-Marie Zentarra for collecting the data of the second study in the course of her Bachelor's Thesis.

References

- Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, 1–11. <https://doi.org/10.1016/j.edurev.2016.11.002>
- Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 206–226). Cambridge: Cambridge University Press.
- Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355–385. <https://doi.org/10.1162/pres.1997.6.4.355>
- Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6), 34–47. <https://doi.org/10.1109/38.963459>

- Billinghurst, M., & Dünser, A. (2012). Augmented reality in the classroom. *Computer*, 45(7), 56–63. <https://doi.org/10.1109/MC.2012.111>
- Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge/Taylor & Francis.
- Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented Reality in education—Cases, places and potentials. *Educational Media International*, 51(1), 1–15. <https://doi.org/10.1080/09523987.2014.889400>
- Chen, P., Liu, X., Cheng, W., & Huang, R. (2017). A review of using augmented reality in education from 2011 to 2016. In E. Popescu, Kinshuk, M. K. Khribi, R. Huang, M. Jemni, N.-S. Chen, & D. G. Sampson (Eds.), *Innovations in Smart Learning, Lecture Notes in Educational Technology* (pp. 13–18). Singapore: Springer Singapore.
- Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02086>
- Dan, A., & Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology*, 122, 75–84. <https://doi.org/10.1016/j.ijpsycho.2016.08.013>
- de Jong, T. (1991). Learning and instruction with computer simulations. *Education and Computing*, 6(3), 217–229. [https://doi.org/10.1016/0167-9287\(91\)80002-F](https://doi.org/10.1016/0167-9287(91)80002-F)
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201. <https://doi.org/10.3102/00346543068002179>
- Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., pp. 735–745). New York, NY: Springer New York.
- Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, 178, 7–25. <https://doi.org/10.1016/j.cognition.2018.05.003>
- Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. *LAK'18: 8th International Conference on Learning Analytics and Knowledge*, 79–88. <https://doi.org/10.1145/3170358.3170377>
- Hugues, O., Fuchs, P., & Nannipieri, O. (2011). New augmented reality taxonomy: Technologies and features of augmented environment. In B. Furht (Ed.), *Handbook of Augmented Reality* (pp. 47–63). New York, NY: Springer New York.
- Jeřábek, T., Rambousek, V., & Wildová, R. (2015). Perceptual specifics and categorisation of augmented reality systems. *Procedia - Social and Behavioral Sciences*, 191, 1740–1744. <https://doi.org/10.1016/j.sbspro.2015.04.419>
- Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications*, 2(1), 24. <https://doi.org/10.1186/s41235-017-0060-9>
- Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press.
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12), 1321–1329.
- Normand, J., Servières, M., & Moreau, G. (2012). A new typology of augmented reality applications. *Augmented Human*, 1–8. <https://doi.org/10.1145/2160125.2160143>
- Phon, D. N. E., Ali, M. B., & Halim, N. D. A. (2014). Collaborative augmented reality in education: A review. *2014 International Conference on Teaching and Learning in Computing and Engineering*, 78–83. <https://doi.org/10.1109/LaTiCE.2014.23>
- Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). “Let’s Get Physical”: Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education*, 6(4), 216–224. <https://doi.org/10.1002/ase.1345>
- Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6), 1533–1543. <https://doi.org/10.1007/s00779-013-0747-y>
- Rekimoto, J., & Nagao, K. (1995). The world through the computer: Computer augmented interaction with real world environments. *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology - UIST '95*, 29–36. <https://doi.org/10.1145/215585.215639>
- Saidin, N. F., Halim, N. D. A., & Yahaya, N. (2015). A review of research on augmented reality in education: Advantages and applications. *International Education Studies*, 8(13), 1–8. <https://doi.org/10.5539/ies.v8n13p1>
- Sommerauer, P., & Müller, O. (2018). Augmented reality for teaching and learning—A literature review on theoretical and empirical foundations. *Twenty-Sixth European Conference on Information Systems (ECIS2018)*. Retrieved from https://aisel.aisnet.org/ecis2018_rp/31

- Specht, M., Ternier, S., & Greller, W. (2011). Dimensions of mobile augmented reality for learning: A first inventory. *Journal of the Research for Educational Technology*, 7(1), 117–127.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, Mass: MIT Press.
- Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41–49. <https://doi.org/10.1016/j.compedu.2012.10.024>

Actor-Network Theory Approach Using M-Learning Technologies in the Public Senior High School as Pedagogy

Mamerto T. GONEDA ^{a*}, Francis Arlando L. ATIENZA ^b, John Paul P. MIRANDA ^c Rolando R. FAJARDO ^d, Dominic C. CABAUTAN ^e

^a *Claro M. Recto High School, DepEd Manila, University of the East – Manila, Philippines*

^b *Pamantasan ng Lungsod ng Manila, University of the East – Manila, Philippines*

^c *Don Honorio Ventura State University, University of the East – Manila, Philippines*

^d *Far Eastern University, University of the East - Manila, Philippines*

^e *Isabela State University, University of the East - Manila, Philippines*

^a mamerto.goneda001@deped.gov.ph, ^b fcatiENZA@yahoo.com, ^c johnpaul.p.miranda@gmail.com,
^d rfajardo@feu.edu.ph, ^e d15cabautan@gmail.com

Abstract: The researchers adopted the phases of Actor-Network Theory approach for the public senior high school teachers with the use of M-Learning technologies as pedagogy since most of them lack access in using different learning materials and textbooks. With this, the teachers find other ways on how to provide learning materials for their students. The study used free and open source educational software to support and sustain the needs of the students and one of the M-learning open source tools is Edmodo. Also, the teacher supplements the needs of the learners with the use of some technological devices such as mobile phone, tablet and desktop computers. The result of the study showed that it is moderately accepted in terms of using the free and open source Edmodo application. The respondents moderately accepted the use of such devices as an educational pedagogical strategy. Moreover, the satisfaction of ANT approach with the use of the m-learning technologies in the public senior high school students in terms of m-learning materials, activities and assessment was highly accepted with a total mean of 4.56. Lastly, the level of agreement of ANT approach as pedagogy in terms of functionality, efficiency, usability, availability, and reliability was strongly agreed by the students.

Keywords: M-learning, M-technologies, Actor-Network Theory (ANT)

1. Introduction

Over the past years, the teaching methodology in the Philippines in public high school has evolved tremendously that is geared and focused to the very needs of our foremost clientele-the students. The teaching pedagogy, taken as an academic discipline that studies how knowledge and skills are exchanged in an educational context has taken much consideration and leads to the betterment and quality of our secondary education.

With this, a lot of researches, pedagogical enhancement and technological infusion have been used in order to uplift and sustain the quality of education in public secondary schools.

1.1 Statement of the Problem

This study sought to adopt the Actor-Network Theory approach using M-Learning technologies in the public senior high school as pedagogy.

Specifically, it aims to answer of the following;

- 1.1.1 What are the challenges encountered by the participants in a traditional teaching approach in providing learning to the public senior high school students?

- 1.1.2 What is the level of acceptance of ANT approach using M-Learning technologies in the public senior high school as pedagogical strategy in terms of:
 - 1.1.2.1 Software Technology,
 - 1.1.2.2 Hardware Technology?
- 1.1.3 What is the level of satisfaction of ANT approach using M-Learning technologies in the public senior high school in facilitating the performance of the learners in terms of:
 - 1.1.3.1 M-learning materials;
 - 1.1.3.2 M-learning activities; and
 - 1.1.3.3 M-learning assessment?
- 1.1.4 What is the level of agreement of ANT approach using M-Learning technologies in the public senior high school as pedagogical strategy in terms of:
 - 1.1.4.1 Functionality;
 - 1.1.4.2 Efficiency;
 - 1.1.4.3 Usability;
 - 1.1.4.4 Availability;
 - 1.1.4.5 Reliability?

1.2 Actor-Network Theory

Adopting actor-network theory in the Philippines education system has a purpose. The main goal of the education in the Philippines is to give the best quality education for all the learners. Carroll et. al. 2012 stated that the fundamental aim of ANT is to explore how network are built or assembled and maintained to achieve a specific objective. ANT provides the ability to uncover the chain of actions or influences from various actors which are carried out to deliver a specific action and outcome.

Actor-Network Theory (ANT) is a part of social theory that highlights on relationship of actors. Social theory is seen as a translator and analyzer tool for researcher to discover the cause and solution of a problem. These theories also help researcher to variously explain and analyze how social action, social processes, and social structures work (Salamat et al. 2011, VI-87). Also, these theories may help the different public senior high school how to support the needs of the actors in the network.

Fenwick and Edwards' (2010) book on Actor-Network Theory in Education suggested that ANT could provide a useful 'lens' to look through at the literature under investigation. Parallels can be drawn to Fenwick and Edwards' account of Bigum's (1998) classifications of the discourses used to position information technologies as technologies for learning. Using the mobile technologies in the education might help to lessen the problem of the teachers and students in the school. Adopting the phases of ANT approach (McBride 2000) may guide the network the implementation of ANT with integration of the technology in the school.

The difference between ANT and other network theories is that in the former, actors and entities can be either human or nonhuman. Their properties are also being dependent on their relationship in a network. ANT treats all elements of the system equally in understanding the relationship among actors (Salamat et al. 2011, VI-84)

Bigum's analysis draws on the perspective of ANT. This theoretical framework uses "the metaphor of heterogeneous network, a way of suggesting that society, organization, agents and machines are all effects generated in pattered networks of diverse (not simply human) materials" (Law 1992, 380).

ANT's main feature is it focuses on inanimate entities and their effect on social processes. An actor is thus defined as the "source of an action regardless of its status as a human or non-human", this is a radical notion in that it contests that inanimate things (e.g. such as technology) can also have agency (Prout 1996 and Law 2002). An actor can however only act in combination with other actors and in constellations that give the actor the possibility to act (Law 2002). Thus, inherent to ANT is a move away from the idea that technology impacts on humans as an external force, to the view that technology emerged from social interests (e.g. economic, professional) and hat it thus has the potential to shape social interaction (Prout 1996).

2. Methods

2.1 Research Methodology

The research method used in this study was descriptive method where data collection was done through survey questionnaire and interview. The respondents answered questions diligently based from the use of M-Learning tools inside the classroom. The questionnaire also asked respondents on their experience on the usage of this technology in different platforms like mobile phones, tablets, and laptops.

2.2 Research Instruments

The following instruments were used by the researchers:

2.2.1 Survey Questionnaire

Below is the Likert scale that the researchers used in the questionnaire of this study “Actor-Network Theory Approach Using M-Learning Technologies in the Public Senior High School as Pedagogy.”

Table 1

Level of acceptance of the respondents using m-learning technologies in the public senior high school with Actor-Network Theory as pedagogy.

Likert Scale	Range	Verbal Interpretation
5	4.51 – 5.00	Highly accepted
4	3.51 – 4.50	Moderately accepted
3	2.51 – 3.50	Accepted
2	1.51 – 2.50	Minimal accepted
1	1.00 – 1.50	Not accepted

Table 2

Level of satisfaction of the respondents using m-learning technologies in the public senior high school in gauging the performance of the learners.

Likert Scale	Range	Verbal Interpretation
5	4.51 – 5.00	Highly satisfied
4	3.51 – 4.50	Moderately satisfied
3	2.51 – 3.50	Satisfied
2	1.51 – 2.50	Minimal satisfied
1	1.00 – 1.50	Not satisfied

Table 3

Level of agreement of the respondents using m-learning technologies in the public senior high school with Actor-Network Theory as pedagogical strategy in terms of functionality, efficiency, usability, availability, and reliability.

Likert Scale	Range	Verbal Interpretation
5	4.51 – 5.00	Strongly agree
4	3.51 – 4.50	Agree
3	2.51 – 3.50	Neutral
2	1.51 – 2.50	Disagree
1	1.00 – 1.50	Strongly disagree

2.3 Statistical Treatment of Data

The researchers used a measure of central tendency to obtain shorthand of the entire data and indirectly described the population where the responses were gathered. Specifically, the weighted mean was employed to determine the average value of response the respondents. The statistical formula for the weighted mean is denoted by:

$$Wm = \frac{TWF}{N}$$

Where:

Wm = Weighted mean

TWF = Total of the products of the weight multiplied by their corresponding frequencies

N = Number of rater or total frequency

2.4 Actor-Network Theory Approach

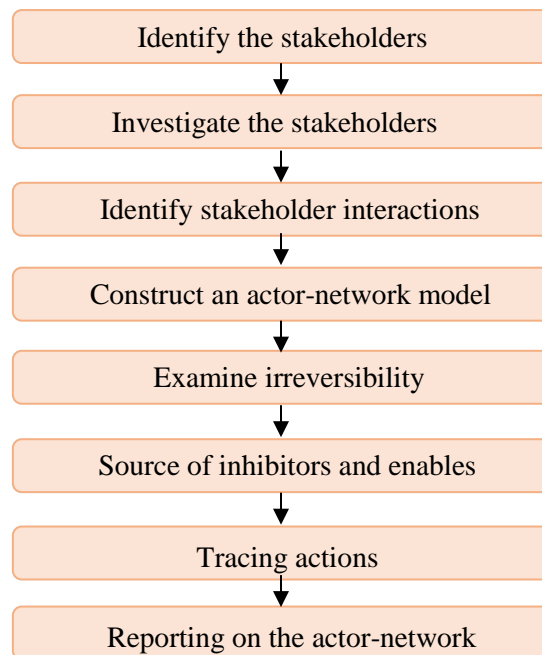


Figure 1: Phases of Adopting the ANT approach (adopted from McBride, 2000)

2.5 Open-Source M-Learning



Figure 3: Image of Edmodo, a free and open-source application
(Source: www.edmodo.com)

2.6 M-Technologies



Figure 4: Image of m-technologies such mobile phones, tablets, laptop
(Source: www.google.com/search?q=mobile+technologies&tbm=isch&ved)

3. Results

There were 50 respondents who were selected to answer the survey questionnaire. 31 of them were male and 19 were female. The respondents were the senior high school teachers from the 3 selected DepEd public secondary schools within the National Capital Region which are Claro M. Recto High School, Lagro High School and Camarin High School.

In terms of technology usage profile of the respondents, the usage of the devices such as mobile phone got 96% or 48 in frequency, tablets got 56% or 28 in frequency and laptop got 34% with 17 in frequency. Therefore, the mobile phone usage ranked first and most of the teachers have been using these devices for more than 3 years with 72%, 2 to 3 years got 22% and less than 1 year for 6% only.

One of the challenges of the senior high public teachers who used traditional teaching approach was taking much time in preparing instructional materials. This resulted to 82% or 41 of them experienced the said concern. On the other hand, 68% of the teachers said that there is a limited teaching material like textbooks and other learning materials used inside the classroom. Also, 76% mentioned of spoon feeding in lecture and discussion which they really wanted to avoid. Lastly, 72% or 36 of them said they wanted to avoid teacher-centered inside the classroom.

The level of acceptance of ANT approach using m-learning technologies in the public senior high school as pedagogical strategy of the teacher resulted to moderately accepted. This means that the open-source Edmodo application for the m-learning using the devices such as mobile phone, tablets and laptops were really accepted by the respondents as a pedagogical strategy in education. The network which is the technology and devices used of the teachers contributed to good and proper learning of the students. Furthermore, the actors who were the students learned and achieved good benefits from the non-human materials.

With regards to the satisfaction felt by the public teachers, the result showed that high satisfaction was seen. The use of m-learning materials was highly satisfied with a mean score of 4.562, while the m-learning activities had a mean score of 4.435 which was moderately satisfied. The m-learning assessment had a result of total mean response of 4.524 which was highly satisfied.

Lastly, the level of agreement of the actor-network theory approach using m-learning technologies in the public senior high school as pedagogy in terms of functionality, efficiency, usability, availability and reliability has resulted to a strongly agree. The use of Edmodo application and using technology devices like mobile phone, tablets and laptops were very helpful for the teachers as their pedagogical strategy in education.

4. Discussion/Conclusion

In summary, the actors or humans play a big factor in managing the non-human entities in education. The non-human entities have a big role in our education especially in the public senior high school where technology shapes the human entities that needs support, engagement and answer to the needs of the teachers and most especially the students.

The integration of technology in education as a pedagogical strategy of the teacher helped them to achieve their goal of supplementing their teaching strategy properly to their students. It is an effective strategy for the senior high teachers so that no learners will be left behind or disengaged during class lectures and discussions.

Lastly, the researchers of the study proved that using the m-technologies such as mobile phone, tablets, and laptops was highly accepted in classroom discussions. It bridged the gap of unavailable learning materials, teacher-centeredness and students non-participatory as major challenges faced by the teachers in their day-to-day chore inside the classroom.

5. Recommendation

The following are the recommendations to be considered:

- a. Other open-source application can be used by the public high school teachers using the m-technologies such as mobile phone, tablets, 2n1 laptops, etc. as educational pedagogical strategy.
- b. Learners can be able to access the lesson from the teacher anytime and anywhere.
- c. The phases of adopting the ANT approach (adopted by McBride, 2000) can be used as a guide for possible implementation in primary or collegiate school setting.
- d. Future researchers will use this research as basis for future studies that will improve the study more especially in the Philippine education system both public and private schools.

Acknowledgements

This research was supported by our professor Dr. Rex Bringula of the University of the East, Graduate School where he gave guidance and ideas in the completion of this study. The researchers also thanked the teachers of the 3 different public secondary high school teachers for their participation in answering honestly the questionnaire where results were extracted and evaluated. Lastly, this study is also possible with the help, critique and support of fellow Doctor in Information Technology students from the University of the East.

References

- Bigum, C. (1998). Solutions in search of education problems: Speaking for computer in schools. *Educational Policy* 12, no, 5:595-601.
- Fenwick, T., Edwards, R. (2010). *Actor-network theory in education*. London: Routledge.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. New York, NY: Oxford University Press.
- Law, J. (1992). Notes on the theory of the actor-network: *Ordering strategy and heterogeneity*, August 1992, Vol. 5, Issue 4, pp. 379-393
- Law, J. Mol A. (2002). *Complexities – social studies of knowledge practices* Duke University Press
- Prout A. (1996). Actor-network theory, technology and medical sociology: an illustrative analysis of the metered dose inhaler. *Sociology of health and illness* 1996, 18:198-219
- Salamat, M.A.B., Hassan, S.B. (2011). *An Actor-Network Theory (ANT) approach to Malaysian e-participation framework*. 2011 International Conference on Social Science Humanity, PERD Vol. 5, ACSIT Press, Singapore

The Pilot Implementation using an Adapted Technology Acceptance Model to Evaluate an Innovative Use of Smartphone for Scientific Investigation Programme in Tertiary Education

Siew Wei THO^{a*}, Tien Tien LEE^a, Sadiyah BAHAROM^a, Faridah Lisa Supian^a, Nurul Syafiqah Yap ABDULLAH^a & Nurul Aina Syakirah ZAINAL ABIDIN^a

^a Faculty of Science & Mathematics, Sultan Idris Education University, Malaysia
*thosw@fsmt.upsi.edu.my

Abstract: This study reports research on the pilot implementation of a new educational programme integrating smartphone technology with scientific investigation through an innovative method of mobile technologies enhanced learning. Smartphone technology has the potential to serve as an excellent tool for students to conduct scientific investigation and learn scientific concepts. A newly-developed educational programme was first pilot and carried out by 37 pre-service science teachers in an education university. Evaluation employed a quantitative method with additional qualitative data determined using a adapted Technology Acceptance Model (TAM) questionnaire survey and open-ended questions to collect data on students' perceptions and implementation issues. The survey findings showed that student teachers agreed with the suitability of educational values with the use of smartphones for performing science investigation. Negative comments and suggestions for refinement were also identified. Thus, we have refined our programme for further applications in secondary schools.

Keywords: smartphone technology; scientific investigation; mobile technologies enhanced learning

1. Introduction

The educational system in the 21st century focuses on the 4Cs, namely critical thinking, creativity, communication and collaboration. Educators need to produce students and pre-service teachers able to think critically and creatively, and with the ability to communicate and collaborate using modern technology. Communication and collaboration can be implemented easily and effectively via technology. Recently, study through mobile learning has become a contemporary research topic which “can offer new opportunities for learning that extend within and beyond the traditional teacher-led classroom” (Sharples, Arnedillo-Sánchez, Milrad, & Vavoula, 2009, p. 233). Moreover, Chan et al. (2006) suggested that seamless learning with mobile devices may enable learning experiences in a number of various learning environments. Therefore, the implementation of mobile learning or to be specific smartphones in the teaching and learning offers a lot of benefits to both the teachers and students. Mobile learning utilizes the learner's spare time from any place at any time (Motiwalla, 2007), and conduct scientific experiments (Vogt & Kuhn, 2012). However, the use of smartphones also leads to certain challenges such as problems with connectivity, battery life, and network issues. All the challenges need to be overcome for a media-enhanced learning environment that can engage students in the learning process. In this study, smartphone technology can be feasibly used as a handy tool to carry out certain scientific investigations. Thus, this study aims to present research findings on educational values and the acceptance of science learning among student teachers after a scientific investigation of the innovative use of smartphone devices.

2. Problem Statements

Recently, numerous findings on mobile technology research and development have revealed that these technologies play a crucial role in science and technology (Sharples et al., 2009; Hwang, Tsai, Chu, Kinshuk, & Chen, 2012). However, most published papers focus on technical issues of development and lack innovative application and evaluation of students' motivation in science learning as well as acceptance of smartphone-based science learning. Thus, little evidence has been found to verify whether this technology can be applied or used to facilitate and assist student's acceptance in science learning. As a result, smartphone devices are important for experiential and real-time interactive science learning environments in which students can perform, observe, respond and share selected experiments (Hwang, et al., 2012; Tho, Lee, & Baharom, 2018). Such a hands-on learning environment via mobile technology enables science students to easily perform real-time experiments almost anytime and anywhere. It will also overcome students' boredom by the conventional or "Cookbook" experiment with manual data collection and graph plotting. As a result, this innovative learning environment can be applied to overcome problems related to limited class time, "Cookbook" experiment and accessibility.

3. Research Questions

- What are the perceptions of the student teachers in terms of the adapted Technology Acceptance Model (TAM) for this educational programme?
- What are the views and major problems encountered of the student teachers about performing scientific investigation through the innovative use of smartphones?
- How can student teachers' opinions be used to improve this educational programme for performing scientific investigation through the innovative use of smartphones?

4. Research Methodology

4.1 Research Design

Quantitative method with qualitative data with two research designs: design and development and also survey design had been used in this study. In other words, this study consists of both development and evaluation of an educational programme. For development, the student worksheets and teacher guide manual were designed and prepared by the researchers using Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model as well as funds acquired through the university research grant. Concerning the scientific investigation, the strengths of the experimental operations include the smartphone itself with the help of free applications (apps) software such as Physics Toolbox Suite apps (created by Vieyra Software, 2016) which can be easily downloaded from the Google Play Store. As a consequence, a test-run of the two programmes was conducted, with consecutive refinement of the learning materials and activities.

4.2 Sample

Two groups of semester 5 student teachers taking science education programme at a faculty of an education university were selected using single-stage cluster sampling. A total of 37 student teachers were involved and surveyed in this study where 21 respondents from Programme 1, and 16 respondents from Programme 2. They were asked to participate in the evaluation process of this educational programme as the student teachers' views and suggestions were highly important, because they studied various science education, technology, and teacher training courses. Furthermore, this study gave them the opportunity to practice the innovative use of Smartphone for scientific investigation, and this idea could be applied in their future teaching and learning at school. To comply with the educational research code, student teachers were required to complete and sign a consent form for voluntary participation in this research.

4.3 Instrument

A research instrument was adapted, and used, namely post-survey questionnaire with open-ended questions was adapted and modified from Technology Acceptance Model (TAM) (Davis, 1989; Teo, Wong, & Chai, 2008; Tho & Yeung, 2016). Adapted TAM is used to seek a concern about how a user accepts and uses a certain technology (Davis, 1989; Teo, Wong, & Chai, 2008). In fact, both perceived usefulness and perceived ease of use are expected to lead to positive attitudes and intentions to accept and practice on particular information technology (Davis, 1989). The questionnaire instrument included survey items on four categories of adapted TAM including perceived usefulness (four items); perceived ease of use (four items); behavioural intention (two items); and attitudes (four items) which were based on a Likert scale from 1-strongly disagree to 4-strongly agree. Open-ended questions were used to collect student teachers' learning experiences (Tho & Yeung, 2016). To ensure content validity and reliability, the questionnaire and open-ended questions were reviewed and criticized by a panel of research experts who examined the content and language used.

4.4 The Flow of Educational Programme and Data Collection

In this educational programme, an innovative way to apply the smartphone through technology-enhanced learning in science education had been developed using ADDIE model. The related scientific investigation via smartphone devices can be performed primarily depends on the smartphone built-in sensors with the help of free application software. There are three different stages in this programme as highlighted in Figure 1 a), namely the briefing, designing and performing simple science investigation and presentation. Figure 1 b) briefly outlines the science activities with the help of smartphone device in the elevator by facilitator and student teacher. After sharing and consolidation, the related post-survey was administered. Then, we refined the design and development of the research based on the results obtained.

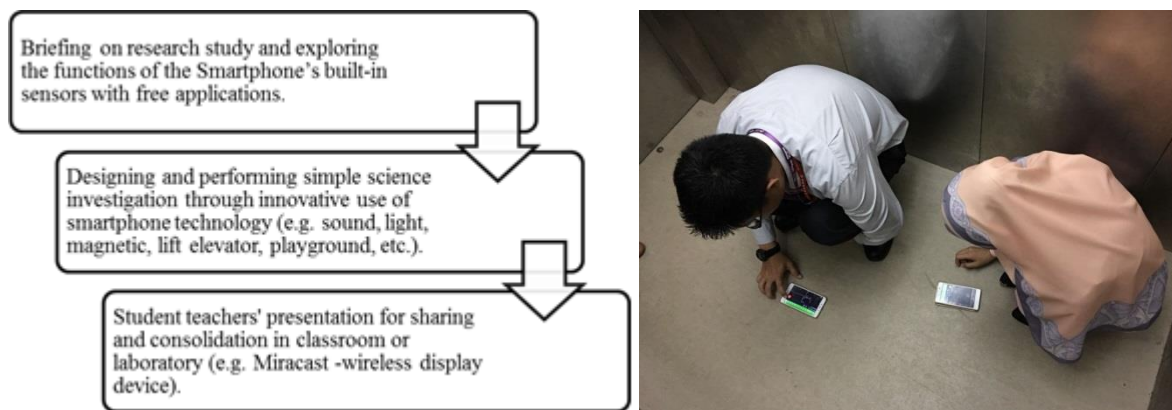


Figure 1. a) Schematic Diagram of Programme; b) Determine the Apparent Weight in a Lift Elevator

4.5 Data Analysis

For quantitative data analysis, the data from the survey was input into the statistical analysis software. Then, the mean scores and standard deviation were calculated. Furthermore, the independent samples t test was used to analysis the survey data. The independent samples t test was applied to examine the differences between two groups of programme. The significant difference was accepted or rejected in any category at the .05 significance level.

For qualitative data analysis, the open-ended data was input for coding into the Qualitative Data Analysis (QDA) Miner Lite free software, and themes according to open-ended questions comprising advantages and disadvantages (First part), knowledge learning, problems encountered, improvement suggestions and topic recommendations (Second part).

5. Findings and Discussion

5.1 Educational Programme Through Innovative Use of Smartphone

For the designing simple scientific investigation in the second part of the educational programme, several meaningful indoor or outdoor science activities had been designed by student teachers under the supervision of facilitators. Several examples of science activities included investigate the relationship between light intensity and distance (indoor); investigate the relationship between magnetic field strength and distance (indoor); investigate the apparent weight in the elevator (outdoor); determine the velocity of sound in air (indoor); Sound pollution (outdoor).

5.2 Survey and Open-ended Data

For the four categories on perceived usefulness, perceived ease of use, behavioural intention and attitudes according to the educational programme, the mean scores (Likert scale 1-strongly disagree to 4-strongly agree) range between 3.14 to 3.54, and the mean score for each category was above the scale of 3 for agreeing with the survey items. The Cronbach's alpha reliability coefficient of the survey was 0.75. Table 1 shows the mean scores obtained from the student teachers' survey and it was acceptable to adopt the classical test theory for the present data analysis. In Table 1, the student teachers from Programme 1 constantly rated higher in the survey items compared to student teacher from Programme 2, except for the behavioural intention category. However, there were no significant differences found by the independent samples t test between these two programmes. Overall, the findings have revealed that student teachers agreed with the educational values inspiring this educational programme through innovative use of Smartphone. In addition, the lowest overall mean score given by student teachers is perceived ease of use category. It was anticipated that the student teachers need more guidance and time to familiarise themselves with this educational programme, as the technique is new to them.

Table 1

Mean scores (with SD in Brackets) of Student Teachers' Response on Survey Items and N is the Number of Participants

Category	Programme 1 (N=21)	Programme 2 (N=16)	Overall (N=37)
Perceived Usefulness	3.53 (.35)	3.47 (.54)	3.51 (.43)
Perceived Ease of Use	3.21 (.39)	3.14 (.55)	3.18 (.46)
Behavioural Intention	3.29 (.51)	3.44 (.63)	3.35 (.56)
Attitudes	3.54 (.32)	3.52 (.51)	3.53 (.41)

In the first part of the open-ended questions, the data collected from student teachers were analysed for positive and negative themes as shown in Table 2. In general, those qualitative data are consistent with the findings obtained by previous researchers who indicated some advantages and disadvantages of using smartphone for educational research and developmental work. Hence, this educational program involving the innovative use of smartphone has the potential to improve their practical skill, with the intention of preparing student teachers to apply this approach in their future career in education. The second part of the open-ended questions is the student teachers feedbacks on their opinions on studying the educational programme were collected in Table 3. Student teachers found these are interesting, saving time and providing accurate results. They were able to perform the scientific investigation anyplace. These responses were consistent with the survey results and the first part of the open-ended questions.

Table 2

Two Open-ended Questions to Assess Student Teachers' Opinions Concerning Advantages and Disadvantages in General on Using Smartphone to Perform Science Investigation (N=37)

Open-ended Question	Insightful Comments*
Advantages	Easy to conduct scientific investigation (14); Performing scientific investigation at anywhere (11); Interesting (8); Saves time (8); Performing scientific investigation anytime (7); Low cost (7); Provides accurate result (4); Use apps without Internet connection (4); Able to create more experiments (3); Able to gain knowledge (3)
Disadvantages	Smartphone with limited built-in sensors (20); Less guiding notes (9); Battery may drain too fast (6); Expensive to buy smartphone with more sensors (5); Require Internet to download apps (5); Hanging/lagging problem (3); Too depending to smartphone (3); Disruption by other apps (2); Incompatible apps (2); Tend to do other thing like surfing Internet, Facebook, etc. (1)

* The comments sorted from most common to least common (with frequencies given in brackets).

The findings indicated most of the student teachers did not experience difficulties. But, they still faced disruption from limited built-in sensors, noise disruption for sound experiment, lack of understanding of raw data and insufficient guiding notes. Therefore, they suggested that participants obtain enough guidance from the facilitators in future implementations. Furthermore, some negative points were unavoidable such as lagging and battery problem. Two insightful suggestions were collected: purchasing more built-in sensors smartphone and creating more interesting activities related to physics topics. Thus these data are consistent with the research and development in physics topic obtained by earlier researchers, such as Tho et al, (2018); Vogt and Kuhn (2012).

Thus, it is noted that a number of negative comments and suggestions were also listed for educational programme refinement. Those negative points of the programme are inevitable nevertheless it can be improved or resolved with suitable refinement in the design and development of the educational programme. Generally, comments from the open-ended questions were consistent with the first part of the open-ended questions and survey results. Hence, the student teachers acknowledged and accepted the values of the educational programme through the innovative use of smartphones.

Table 3

Four Open-ended Questions in The Second Part to Assess Student Teachers Opinions Concerning the Educational Programme (N=37)

Open-ended Question	Particular Perceptive Comments*
What have you learned from the scientific investigation through the use of Smartphone?	Learn to use smartphone for scientific investigation (15); Easy to conduct scientific investigation (15); Fun and interesting (9); Saves time (3); Provides accurate results (3); Performing scientific investigation anywhere (1)
Have you encountered any problems concerning the programme while doing the scientific investigation? If so, briefly describe the problem(s).	No problem encountered (14); smartphone with limited built-in sensors (9); Disruption by the surrounding environment (4); Problem with data analysis (4); Less guiding notes (3); Hanging/lagging problem (3); Battery may drain too fast (1); More apps need to be installed (1); Can't perform all kind of experiments (1)
If you were given another chance to do the scientific investigation again, please suggest some ways for improvement?	Give more guide of scientific investigation (14); Purchase a smartphone with more sensors (6); Create more interesting activities and experiments (5); Update android version and apps (4); Provide high speed Internet (3); Reduce disruption by surrounding environment (2); Smartphones need to be fully charged (1)
Please suggest a feasible science topic or activity that is possible to apply the scientific investigation through the use of Smartphone.	Related to force and motion including velocity, acceleration, free fall, momentum (11); Related to sound and wave experiment (10); Light experiment (7); Magnetic experiment (6); Rotational motion (1); Thermodynamic experiment (1); Measuring height using barometer sensor (1)

* The comments are sorted from most common to least common (with frequencies given in brackets).

6. Educational Implications, Refinement, Conclusion and Future Work

The findings of this study illustrated that the new educational programme through innovative use of smartphone has a significant potential for application of outdoor or indoor science practices in the university level as a complement to scientific investigation. The significance requirements for applying mobile technologies into education are widely discussed and supported by previous research studies (Motiwalla, 2007; Sun, Looi, Wu, & Xie, 2015). Based on the analysis, there are three key implications of this study for the future development and school-based implementation of this new educational

programme. First, the refined educational programme could be used for empowering and supplementing science teaching and learning practices beyond the limitations of normal lectures in university. Second, open-ended data revealed that specific consideration need to be given to the educational programme training and guidance particularly using smartphone for collecting experimental data (Kirschner, Sweller, & Clark, 2006; Tho & Yeung, 2016). This is likely caused by implementing of new programme which is different from their normal learning process and therefore a more comprehensive guide in procedures need to be established before implementing this to secondary classroom. Lastly, the refined programme was successfully integrated to the secondary science education including several feasible topics in science subjects. To date, there have been rather limited indoor and outdoor science investigations for secondary level particularly in Malaysia. Thus, student teachers can apply this approach in their future teaching career in schools.

Based on the analysis of survey and open-ended data collected, we refined the programme by 1) preparing several reasonable price of China smartphone products with more built-in sensors during future implementation; 2) preparing high volume of power bank for charging smartphone devices; 3) preparing and giving clearer educational programme operating guidelines in the form of manual and video clip; 4) modifying present programmes and creating more exemplars for each built-in sensor; 5) integrating with the secondary science or physics syllabus with the aim that the student teachers can apply this programme in their future learning and teaching in schools. In conclusion, the findings of this study illustrated that the new educational programme through innovative use of smartphone has a significant potential for application of outdoor or indoor science practices in the tertiary level.

Acknowledgements

We are very grateful to Sultan Idris Education University for financial support (University Grant Research Code: 2018-0047-107-01 and 2016-0105-107-01).

References

- Chan, T. W., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., ... & Soloway, E. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1(1), 3-29.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- Hwang, G. J., Tsai, C. C., Chu, H. C., Kinshuk, & Chen, C. Y. (2012). A context-aware ubiquitous learning approach to conducting scientific inquiry activities in a science park. *Australasian Journal of Educational Technology*, 28(5), 931-947.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
- Motiwalla, L. F. (2007). Mobile learning: A framework and evaluation. *Computers & Education*, 49(3), 581-596.
- Sharples, M., Arnedillo-Sánchez, I., Milrad, M., & Vavoula, G. (2009). Mobile learning. In *Technology-enhanced learning* (pp. 233-249). Springer, Dordrecht.
- Sun, D., Looi, C. K., Wu, L., & Xie, W. (2015). The innovative immersion of mobile learning into a science curriculum in Singapore: An exploratory study. *Research in Science Education*, 1-27.
- Teo, T., Luan, W. S., & Sing, C. C. (2008). A cross-cultural examination of the intention to use technology between Singaporean and Malaysian pre-service teachers: an application of the Technology Acceptance Model (TAM). *Journal of Educational Technology & Society*, 11(4), 265-280.
- Tho, S. W., Lee, T. T., & Baharom, S. (2018). Scientific investigation through innovative use of mobile devices in the playground. *Teaching Science*, 64(2), 25-29.
- Tho, S. W., & Yeung, Y. Y. (2016). Technology-enhanced science learning through remote laboratory: System design and pilot implementation in tertiary education. *Australasian Journal of Educational Technology*, 32(3), 96-111.
- Vieyra Software (2016). Physics Toolbox Suite [Mobile application software]. Retrieved from <https://www.vieyrasoftware.net/>
- Vogt, P., & Kuhn, J. (2012). Analyzing free fall with a smartphone acceleration sensor. *The Physics Teacher* 50(3), 182-183.

Proposal for Deviceless Learning Environments Instead of Environments Using Smart Devices

Kozo MIZUTANI^{a*}

^a*Department of Information and Electronic Engineering, School of Science and Engineering,
Teikyo University, Japan*

*mizutani@ics.teikyo-u.ac.jp

Abstract: Smart devices such as smartphones and tablets have become indispensable in educational settings in the 21st century. However, to both ensure the educational effect and reduce the cost of using these devices, we should overcome the issues identified by earlier, pioneering studies. Accordingly, this paper proposes a deviceless learning environment called “Followable Learning Environment” (FLE). Although FLE is based on projection mapping techniques that project a user interface (UI) on desks, the UI is dynamically projected on a student’s desk or hands according to their behavior by using the area of projection as the student’s movable area. Since the UIs of several students are projected by one projector, we need not separately prepare the devices and their management for each student. In addition, using FLE, we can realize new functions that are impossible using smart devices. This paper describes the background of this research and explains the concept of FLE through the development of a prototype system.

Keywords: Learning Environment, Smart Devices, Augmented Reality, Dynamic Projection Mapping, One-to-One Computing

1. Introduction

Today, smart devices such as smartphones and tablets have become indispensable in educational settings. Although many pioneering studies have focused on the pedagogical effects of using smart devices in educational settings, many studies also mention issues such as the preparation and management of these devices. Another consideration pertaining to the use of these devices is the resulting increase in cost of educational services.

This paper proposes a new type of learning environment called the “Followable Learning Environment” (FLE) as a substitute to the use of smart devices in educational settings. Although FLE is based on projection mapping techniques that project a user interface (UI) on a desk, the UI can be dynamically projected on a student’s desk or hands according to their behavior by using the area of projection as the student’s movable area. In addition, we can realize new functions that are impossible using smart devices, such as a changing display size function that changes the display size from small, for individual students, to large, for collaborative learning, and overlap projection functions to promote students’ understanding. Accordingly, this paper describes the background of this research and explains the concept of FLE through the development of a prototype system.

2. Background

Many studies have revealed that the use of tablets enhances the learning motivation and understanding of students. For example, a survey on 6,057 students and 302 teachers in Quebec, Canada, revealed the positive effects of tablet use, such as an increase in student motivation, ease of information access, the quality of presentations, within-classroom collaboration, and creativity (Karsenti et al. 2013). A study by Alvarez et al. (2011) showed that tablets strengthen collective discourse capabilities and facilitate a richer and more natural body language, on collaborative learning in undergraduate course. Many studies have examined the utilization of smart devices in education in Japan, as well. In addition, in 2013, the Japanese government published an IT strategy to provide one smart device to one student (one-to-one computing). Although the utilization of smart devices in educational settings has undisputed benefits, it

generates some problems that remain unresolved to date. This study focuses on the problems of utilizing smart devices in school settings as following.

2.1 Initial Costs and Operational Costs

The problems associated with high costs are always a major consideration. Regarding the use of smart devices, both the costs of preparing the devices and costs of software licenses, staff training, provision of services, provision of technical support, and so on should be considered. With respect to the preparation of devices, although some schools request the parents of students to buy devices in some cases, the schools provide their students with devices, such as 1-1 computing initiatives, in some other cases. For successful 1-1 computing, high-quality infrastructure and readily available technical support are necessary (Valiente 2010). As a factor of increasing costs, there might be a problem for 1-1 computing itself. Since the main user of a device is a student, administrators cannot completely control the use of such devices. Although various Mobile Device Management (MDM) tools are currently available, it has limitations for management scope as well, it is difficult problem that to cope with both security and functionality. A similar problem occurs in corporate usage, as well. To manage issues related to both security and functionality, a thin client solution, such as the Virtual Desktop Infrastructure (VDI), is introduced in corporate use. On other hand, in educational settings, particularly the utilization of IT in primary school, the introduction of 1-1 computing using tablets seems to have attained global focus.

2.2 Distraction

Another consideration is the problem of being distracted in class. This problem is an important in improving the cost–benefit relationship. The Quebec survey mentioned in Section 2 reported that the greatest challenge faced by teachers is the provision of tablets as a distraction for students. A survey conducted by Ditzler et al. (2016) also mentioned the problem of distraction.

Because students always want to use a device, they might not be able to concentrate on the learning activity even when the activity does not need such a device. In this case, although we should instruct students to turn off the device, place it in a bag, and so on, teachers might hesitate to use devices depending on the instructional model since these devices impair convenience. Similar problems also be mentioned in the survey by Roblin et al. (2018).

2.3 Problems in Primary and Junior High Schools

In addition, under the assumption that tablet devices are used in primary and junior high schools, this study also focuses problems as follows.

- ✓ *Device Loss/Breakage:* Students often inadvertently lose/break their devices. Although, such accidents can be compensated by the provision of insurance, they do increase the overall cost. A case study of Henderson et al (2012) mentioned that additional costs had also incurred through the purchase of protective gear for devices, such as cases and screen protectors.
- ✓ *Batteries:* Currently, devices have sufficient battery capacity for one-day use. However, students often forget to charge batteries. Moreover, they might use the devices for entertainment during recess hours, which also requires batteries.
- ✓ *Small Desks:* Although it might be a problem unique to Japan, students' desks in classrooms are relatively small in size. In typical Japanese primary schools, a desk has a breadth of approximately 600 mm and length of 400 mm. Since the desk is small in size, students often drop objects placed on the desk; this might be one reason for the breakage of devices.

These problems can be solved by preparing spare devices. However, the problems should be solved smoothly without interrupting the progress of lessons. For this purpose, sufficient numbers of spare devices and well-trained technical staff should be made available, which causes an increase in overall cost.

A major trend that has emerged in recent years is 1-1 computing. However, although this concept has been implemented to solve the aforementioned problems, it seems to that, to date, no study has attempted to fundamentally solve these problems. Accordingly, this study proposes a new concept to solve these problems.

3. Concept of the Followable Learning Environment

The author proposes the application of a new type of UI called Followable User Interface (FUI) (Yamaguchi et al. 2015). In the FUI environment, the UI is projected on a user's hand using sensors placed on the ceiling of a room that detect hand behavior (Figure 1). One UI is projected by using the partial area of one projector. Therefore, the area of the projector is considered the area in which the user can move his or her hands. In addition, several UIs of one user can be projected from one projector at the same time, in theory. Users' experience of this environment is similar to that of using conventional UIs without smart devices. The size of area which the FUI environment can projects UI using two projectors and two sensors in the current experimental system is only about 4m × 3m. In the future, to spread the area until room size by using several sensors and projectors is planned in this study.

In this study, the concept of FUI is applied to a classroom for solving the aforementioned problems, as shown in Figure 2. By using a sensor located on the ceiling, the system detects the desks and gestures of multiple students. One projector projects each student's UI at the same time. By increasing the number of sensors and projectors, it is possible to provide UIs to the entire classroom. This study calls this concept based on FUI as "Followable Learning Environment (FLE)".

Instead of projecting the UI on a user's hand, it is projected on a student's desk. Since the method is deviceless, additional devices need not be prepared for each student. An application instance of UI, which is projected on the desks, is executed on the FLE system similar to a thin client system. It is expected that this method reduces management costs compared to the use of smart devices.

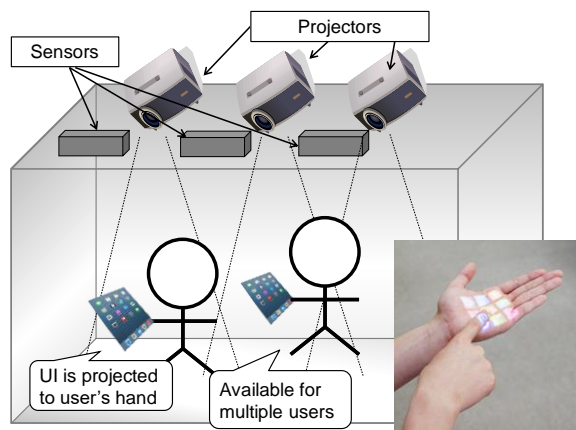


Figure 1. The Concept of the Followable User Interface.

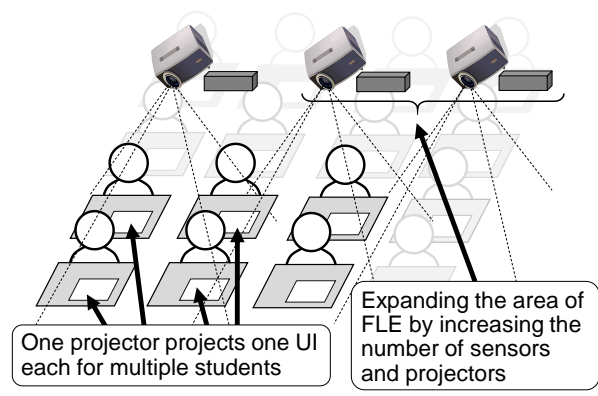


Figure 2. The Concept of Followable Learning Environment.

4. Development of a Prototype System and Experimental Results

Realizing a system based on the FLE concept is a challenging task. For this purpose, we should develop elemental technologies for instance, to integrate sensor data, detect desk positions and shapes, detect gestures, and perform high-precision projection mapping, as well as applications that utilize these technologies. Accordingly, as the first step, this study developed a prototype system that performs desk detection and UI projection using one sensor and one projector to explore the feasibility of FLE.

Figure 3 depicts the abstract of the prototype system. The system uses a depth sensor, which is located near the ceiling such that it is directed toward the floor. The captured depth data are sent to a workstation. In the workstation, the system program detects desks, decides UI positions, and creates UIs. The UIs are projected from a projector connected to the workstation. Similar to the sensor, the projector is placed near the ceiling such that it is directed toward the floor.

The prototype system was evaluated using actual desks. Figure 3 depicts the arrangement image of the overall system. Figure 4 depicts the arrangement image of the overall system. Since the projected UIs are pseudo UIs, we cannot operate them. The experiment used prepared desks of dimensions 600 mm × 500 mm which are nearly size of used in primary schools. The top plate height from the floor was 700 mm. The sensor and projector were placed at a height of 2500 mm and 2200 mm from the floor. Figure 5 depicts some UI projection examples of different cases of using several desks: (a) arrangement in the vertical direction, (b) arrangement of four desks, and (c, d) arrangement for group work. Since UIs were projected using a part of the projection area of one projector according to the FLE concept, UIs could be projected on each desk even after changing the arrangement of desks.

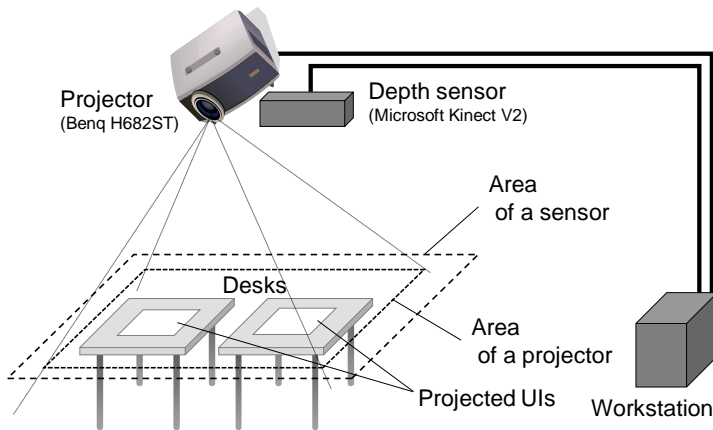
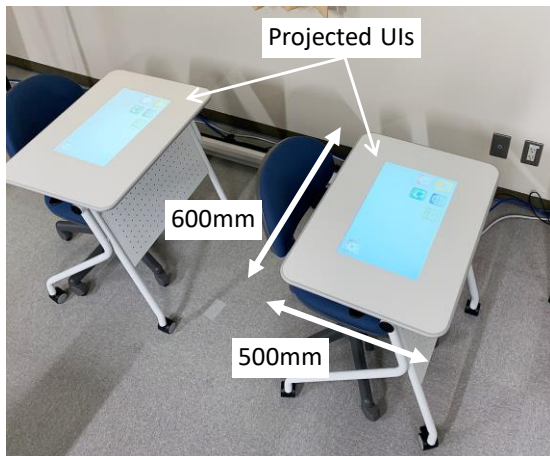


Figure 3. Abstract of the Prototype System.



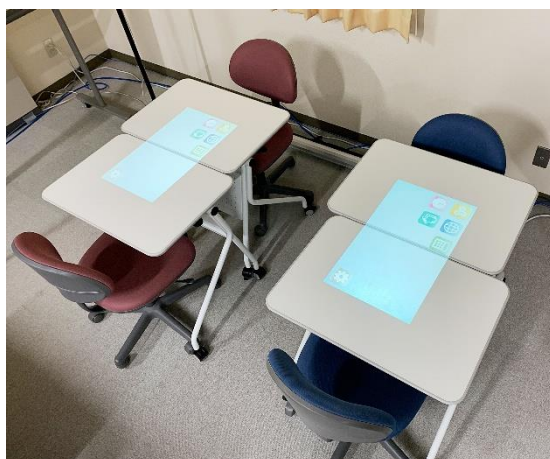
Figure 4. Arrangement of the Overall System.



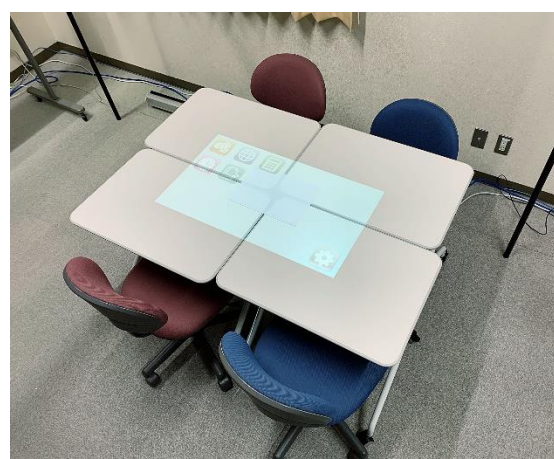
(a) An arrangement of two desks



(b) An arrangement of four desks



(c) Two desks and two groups



(d) Four desks and one group

Figure 5. Examples of the UI Projections Provided by the Prototype System.

5. Consideration

Since the prototype system performs desk detection and pseudo UI projection alone, we cannot evaluate the extent to which the aforementioned problems could be overcome by the prototype system based on the results of our experiment; however, this study considers the FLE concept from different perspectives.

5.1 Feasibility

The prototype system used only one sensor and one projector. However, to practically realize FLE, functions such as the recognition of gestures for UI operation, application software for educational use, an integration mechanism to integrate multiple sensors and projectors, system management tools, and so on, should be implemented. These are challenging tasks because the realization of each function requires the consideration of various factors.

In the experiments, the sensor covered an area of approximately 2560 mm × 2030 mm, and the projector covered approximately 2000 mm × 1120 mm (when zooming to a small size). Within a classroom, it is not realistic to place many projectors on the ceiling. However, projector vendors have proposed products which are integrated with the functions of both a ceiling light and a projector. By popularizing such products, instead of utilizing ceiling lights by themselves, the feasibility of FLE implementation can be increased.

In the experiment, the size of a UI projected on a desk was 350 mm × 215 mm, and the resolution is 369 pixels × 208 pixels, which is not sufficient for reading e-textbooks, web pages, and so on. In FLE, the resolution of each UI is depended to the resolution and projection angle of the projector. A simple method to increase the resolution of each UI is to apply high resolutions; however, it depends on the number of desks covered by the projector. If a projector with high resolution and a wide projection angle is used, it is possible to increase the resolution of each UI and the number of desks simultaneously. Although the projector's resolution is increased to mainly improve image reproducibility, such as in recent movie appreciation use, the implementation of FLE gives projector technologies new needs.

5.2 Effect of Solving Problems

The results of the experiment cannot explain the effect of improving all the problems described in section 2. However, since the system's architecture is similar to that of a thin client system, it is expected to reduce operational costs. There is no need to worry about charging the batteries or loss of devices, as well. Although there is the risk that projectors might break down, students can use FLE by temporarily moving their desks to the areas of other projectors until the faulty projector is replaced.

Regarding the distraction of students, the condition will be improved using an FLE because a teacher can completely control his or her students' UIs similar to a thin client system. It is easy to turn each UI on/off. When they do not need UIs, the students can use their desks effectively since no devices to take up desk space, unlike in non-FLE settings.

5.3 Future Potential

As mentioned in the Feasibility section, realizing FLE is a challenging task, particularly since we have to wait for the implementation of products that can integrate high-resolution, wide-angle projectors and a ceiling light. However, if the concept of FLE is realized, it is expected to provide functions that cannot be realized using 1-1 computing, such as the following:

- ✓ Optimization of the Arrangement of Items on a Desk:
Since the position and size of UIs can be adjusted, students can optimize the arrangement of items on their desks. Theoretically, since FLE can project non-square UIs, it might be able to use the top of a desk more efficiently. These functions are not possible with the use of today's smart devices.

- ✓ Collaborative Learning:
In collaborative learning, large display tools, such as interactive whiteboards, are often used. In 1-1 computing initiatives, these displays are generally prepared in addition to the devices required by each student. However, this causes a rise in costs. Further, when preparing such displays for different groups (since small displays reduce the efficiency of collaboration tasks), it is desirable to change the size of displays according to the number of members in each group. However, this is impossible when using conventional techniques. As shown in Figure 5, FLE can provide UIs optimized for group work without additional investment. Theoretically, using FLE, several large UIs and personal UIs can be projected at the same time on the area provided by adjoining desks. Hence, FLE is expected to easily realize collaborative environments.
- ✓ Application of Projection-Based Augmented Reality and Learning Analytics:
Since FLE uses projectors, it can incorporate the concept of projection-based augmented reality (AR). Many studies apply AR to learning experiences, which is an advantage of AR, and report effects such as improvements in learning gains, enhancement of motivation, facilitation of interaction, and so on (Bacca et al. 2014). FLE enables the augmentation of several desks at the same time by projecting visual effects using a part of the projection area of one projector.
Moreover, since FLE uses sensors, we can capture the behavior of each student for learning analytics. It is expected to realize applications such as encouraging students by detecting a lazy attitude, making a group working be actively by detecting discussion stagnation and so on.

6. Conclusion

This study proposed a new learning environment concept as a substitute to using smart devices called FLE. In this concept, by using the projection area of a projector as the movable area of a student, a UI is projected dynamically on the student's desk or hands according to their behavior. This study discussed the design of a prototype system based on the FLE concept, which used one projector and one RGB-D sensor and showed how UIs can be projected on each student's desks according to the desk arrangement. Although many issues remain to be solved to practically implement FLE, the study showed that FLE has many advantages in educational settings.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP18K11580.

References

- Alvarez, C., Brown, C., & Nussbaum, M. (2011). Comparative study of netbooks and tablet PCs for fostering face-to-face collaborative learning, *Computers in Human Behavior*, 27(2), 834-844.
- Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk (2014). Augmented reality trends in education: A systematic review of research and applications, *Educational Technology & Society*, 17, 133-149.
- Ditzler, C., Hong, E., & Strudler, N. (2016). How tablets are utilized in the classroom, *Journal of Research on Technology in Education*, 48(3), 181-193.
- Henderson, S. & Yeow, J. (2012). iPad in Education: A Case Study of iPad Adoption and Use in a Primary School, *45th Hawaii International Conference on System Sciences*, 78-87.
- Karsenti, T., & Fievez, A. (2013). The iPad in education: Uses, benefits, and challenges—A survey of 6,057 students and 302 teachers in Quebec (Canada). Montreal, QC: CRIFPE.
- Roblin, N. P., Tondeur, J., Voogt, J., Bruggeman, B., Mathieu, G., & Braak, V. J. (2018). Practical considerations informing teachers' technology integration decisions: The case of tablet-PCs, *Technology Pedagogy and Education*, 27(2), 165-181.
- Valiente, O. (2010). 1-1 in Education: Current practice, international comparative research evidence and policy implications, *OECD Education Working Papers 44*, OECD Publishing.
- Yamaguchi, T., Mizutani, K., & Arai, M. (2015). A study of followable user interface to hand behavior, *International Journal of Knowledge Engineering*, 1(3), 240-243.

A Semantic Tag-based enhanced Learning Recommendation approach for enhancing student learning experiences

Tien-Yu Hsu^a, Jun-Ming Chen^{a*}, Hsin-Yi Liang^a

^a*Department of Operation, Visitor Service, Collection and Information Management,
National Museum of Natural Science, Taichung, Taiwan*

*jackychen.academic@gmail.com

Abstract: With the rapid development of semantic web and ubiquitous learning applications, constructing prior knowledge has also become increasingly important for on-line reading learning. Traditional studies on prior knowledge generation during reading activities have focused on extracting sentences from reading materials that are manually generated by website administrators and educators. This is time-consuming and strenuous, and hence personalized prior knowledge recommendation is difficult to perform. To cope with this problem, we combine the concept of prior knowledge with semantic tagging methods to assist the reading comprehension of students studying science subject, which then identifies suitable supplementary materials for quickly constructing a student's prior knowledge reservoir. Moreover, we incorporate tags into a tag based learning approach to automatically monitor running activities and student status. The experimental results clearly show that the proposed approach not only significantly improves the efficiency of knowledge learning but also helps teachers assist students in improving their reading ability.

Keywords: ubiquitous learning, intelligent tutoring systems, interactive learning environments, museum learning

1. Introduction

In the past, learning was considered a process of accumulating information or experience (Hwang & Chang, 2016). Building prior knowledge through reading is becoming increasingly important for students, as it helps students learn quickly and effectively. Thus, obtaining this skill early during the learning process is crucial. Research has shown that a student's prior knowledge often confounds an educator's best efforts to deliver ideas accurately (Sung & Hwang, 2018). Thus, to help students make the most of new experience educators need to understand how prior knowledge affects learning (Falk, J. H., & Dierking, 2018). However, despite the value of prior information, Taiwanese schools have largely focused on skill development rather than expanding a student's knowledge of the world, such that reading comprehension and prior knowledge instruction are still a challenge in English as a foreign language (EFL) classes in Taiwan's education (McNAMARA, Ozuru, & Floyd, 2017).

Additionally, studies have found that students who find it hard to link new knowledge in a social learning environment, resulting in students having problems understanding, recalling, and accessing the new knowledge later (Song, Kalet, & Plass, 2016). This leads to poor levels of reading comprehension among students, such that even above average students are unable to read and fully understand material. In short, there is a gap between the rich information received and the knowledge generated.

To cope with the problem, this paper proposes a Semantic Tag-based Enhanced Learning Recommendation approach (STLR) by employing semantic tag analysis and visualization monitoring mechanism, which provides prior knowledge learning environments that help promote critical thinking through article construction can activate a student's existing schema and help them identify new information from articles more easily (Falk, J. H., & Dierking, 2018; Peng, Zhang, & Ho, 2019). Such background information may even help students find clues for identifying the meanings of new vocabulary or sentence patterns (Cottrell, 2017). The most important value of this social tagging system,

however, is the promise it shows for dramatically improving student reading comprehension. Meanwhile, designing a monitoring interface that enables teachers to navigate to potentially relevant information, and assists teachers in both analyzing interactions from student tag behaviors and evaluating student learning performance.

2. STLR: a Semantic Tag-based enhanced Learning Recommendation approach

Reading involves a complex cognitive process for students who take the initiative in learning. In order to effectively assist the students in enriching prior knowledge and raise their reading comprehension, and is also helpful in assisting teachers in evaluating student reading comprehension by tag-cloud visualization, in this study, a Semantic Tag-based enhanced Learning Recommendation (STLR) system was developed by employing semantic web and its related techniques for conducting visiting learning activities in museum. The proposed system consists of a mobile reading learning module, a prior knowledge recommendation module, learning monitoring module, and a back-end databases module s for computer-aided reading practices, as shown in Figure 1.

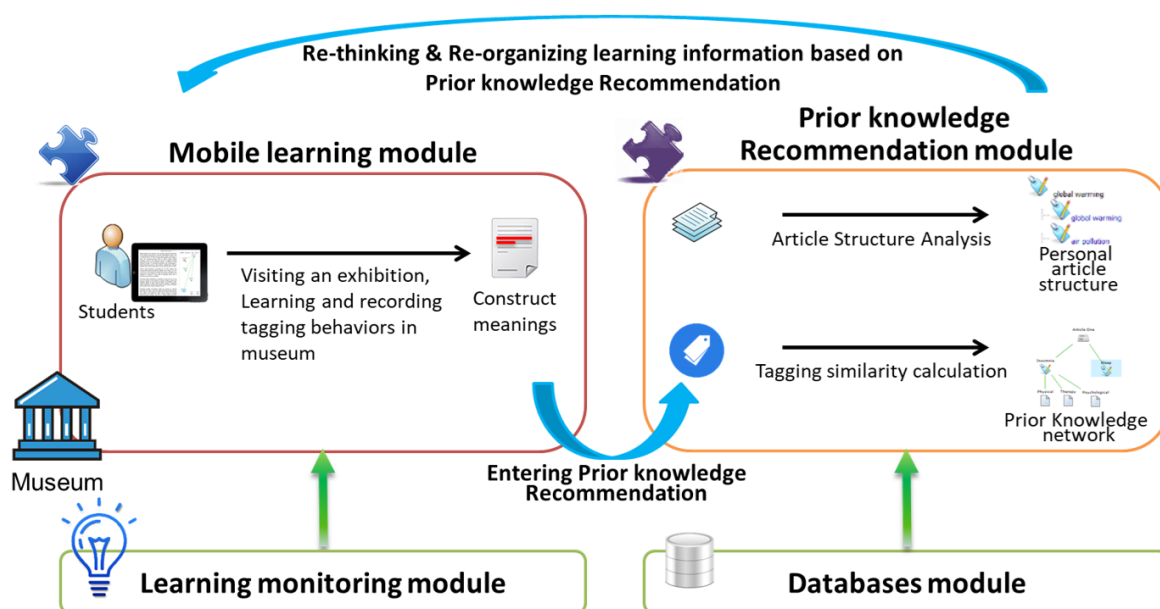


Figure 1. Structure of the proposed STLR system.

- (1) **Mobile Learning module:** students can use their tablets to learning via a wireless network, observe exhibitions, read on-line learning materials and record their learning behaviors. Students can annotate various online resources (materials) with freely chosen tags. Tagging certain activities can help students summarize new ideas and quickly grasp the structure and concepts of English articles (Chen et al., 2010; Klačnjaja-Milićević et al., 2018). Moreover, these tags are also designed to enhance critical thinking skills by directing students to evaluate and then support or oppose different viewpoints on their readings. The functions of this module contain viewing exhibition information and learning materials and tags recording.
- (2) **Prior Knowledge Recommendation module:** the functions of this module contain to construct personalize article structure and prior knowledge network, and then recommends appropriate supplementary knowledge to students and helps them gain new information from articles more easily. The major functions are presented as follows:
 - **Article structure analysis:** we combine the results of paragraph analysis and tag classification to decide which paragraphs are most relevant and which paragraphs represent subtopics of the article. The presenting information can help students hold each bit of information or connect it to background knowledge, so that enables students in understanding articles and related elements of exhibitions.
 - **Prior knowledge network:** in order to find suitable prior knowledge for a student, we use the key sentence of the paragraph as a query to search for similar articles. Meanwhile, to ensure that these prior knowledge articles are suitably useful for students, we search for topic sentences

from each paragraph by article structure analysis that the student is unfamiliar with or does not understand by students' tagging similarity calculation (Chen et al., 2014). We then recommends appropriate supplementary knowledge to students and helps them gain new information from articles more easily.

- (3) **Learning monitoring module:** teachers can observe the current status of a student's learning performance and tagging behavior. When students who have a learning disability are diagnosed by analyzing the tag cloud, this interface can assist teachers in providing focused feedback and questions to students as soon as possible. This information is helpful for assisting teachers in evaluating the reading comprehension status of students, and thus provides teachers with valuable input when they attempt to adjust teaching strategies or diagnose a student's learning obstacles.
- (4) **Back-End Databases module:** The module contains databases of the learner's personal information, learning materials, annotated tags, learning portfolios and semantic tag weighting sheets.

Figure 2 illustrates the user interface of the STL system. Students first select articles via a drop-down menu, after which the context of the selected article is displayed. The personal article structure consists of a root node (e.g. global warming), which presents the main topic of the article, and sub-nodes (e.g. global warming, air pollution, extreme weather, and Kyoto protocol), which present the topic of each paragraph. If the paragraph is not important, no sub-nodes are included. In addition, article content is highlighted whenever students click on the sub-node of the "personal article structure." This highlighted content can include entire paragraphs, any key sentence represented by the sub-node, and the key sentence from the key sentence computation. Overall, this highlighted information provides each student with a quick and useful personal snapshot of the reading material.

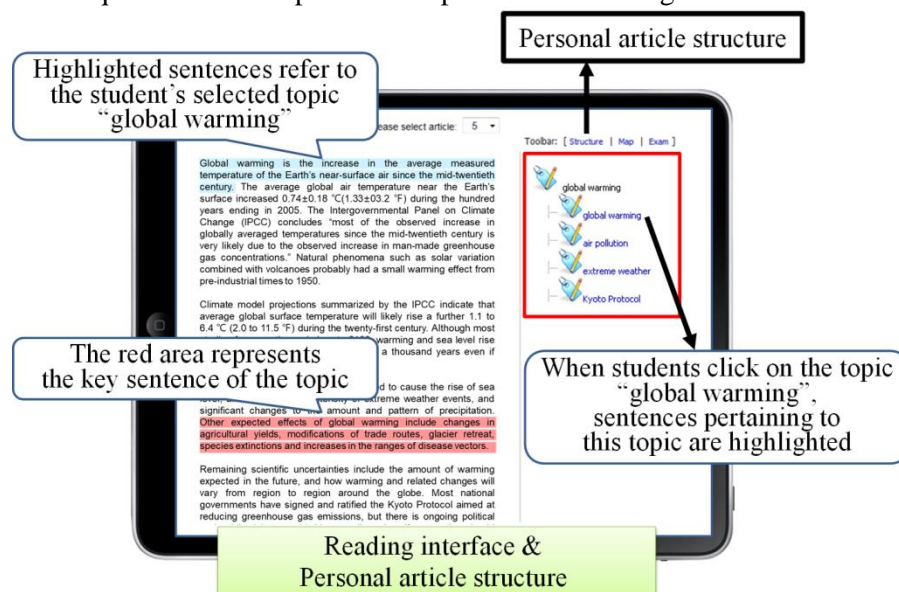


Figure 2. System Interface for Reading learning and Personal article structure.

Figure 3(a) illustrates the structure of the "prior knowledge network" which consists of a root node and sub-nodes, and the locations of nodes that correspond to the order of the paragraphs. This structure varies depending on the student. If a node of the structure represents the recommended topic for a student, such as "global warming," the node is highlighted. The most important function of the network of prior knowledge is its provision of adaptive and necessary prior knowledge. Students click any node in the prior knowledge network and a new window appears, displaying the prior knowledge article to the student. Here, the provided articles are in Chinese, because the goal is to supplement their language knowledge deficiency with prior knowledge, which is more easily consumed in the student's native language. Figure 3(b) shows the teacher interface presenting the current status of a student's learning performance and tagging behavior. When students who have a learning disability are diagnosed by analyzing the tag cloud, this interface can assist teachers in providing focused feedback and questions to students as soon as possible. This information is clearly helpful for assisting teachers in

evaluating the reading comprehension status of students, and thus provides teachers with valuable input when they attempt to adjust teaching strategies or diagnose a student's learning obstacles.

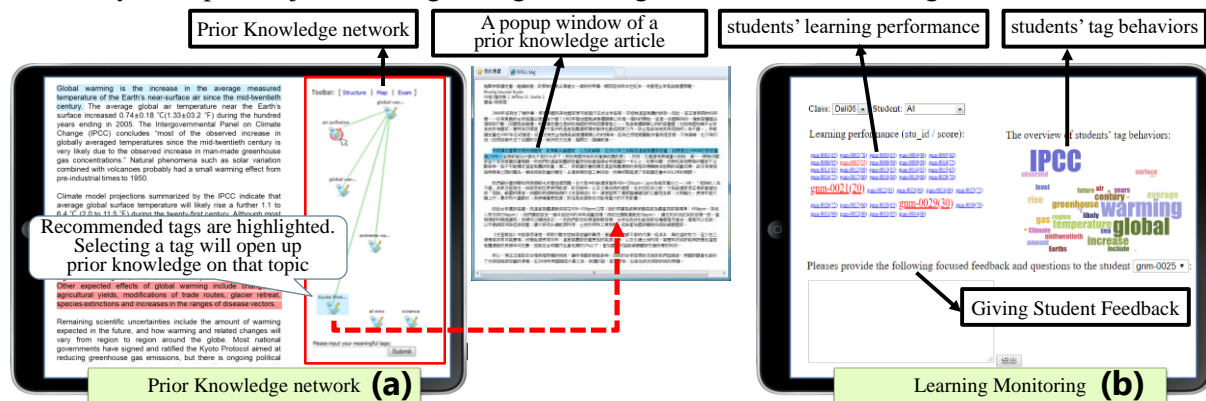


Figure 3. System Interface for Prior Knowledge Recommendation and Learning Monitoring.

3. Experimental design

To evaluate the effectiveness of the proposed approach, an experiment was conducted on visiting learning activity in a Taiwan's museum. The objective of the learning activity was to improve student comprehension, translating into overall better learning performance. The activity encouraged students to observe, read, and tried to help students get a better grasp of what is going on in the articles and associated elements of the exhibition they see. It was expected that the tag visualization and learning recommendation would not only be used by the students to get a quicker grasp of the meanings and the relationships between the learning targets, but also by teachers to assist them in tracing and analyzing the information search behaviors of students.

The participants were 68 students from one senior high school in Taiwan. Before the experiment, one class was assigned to be the control group and the other was assigned to be the experimental group (each with 34 students). The two groups of students did not have any interaction, so that they would not be affected by the other group during the learning activity. Next, each group was given a pre-test evaluation and then entering learning activity. Additionally, in order to evaluate the effectiveness of the proposed learning monitoring mechanism, five participating teachers from a senior high school in Taiwan were invited to experience the use of the teacher interface. The teachers were all experienced with over 5 years in the teaching profession.

4. Experimental Results

4.1 Learning achievement

In order to evaluate the feasibility and potential application of the STL system, we adopted a pre- and post-test experimental design that made use of before and after surveys to demonstrate the achievement of learning outcomes.

Pre-test: All students took a pre-test at the beginning of the reading activity. Table 1 shows the t-test values for the pre-test and post-test results. The mean score of the pre-test for the control group was 60.58, and 58.01 for the experimental group. The t-test result showed that these two groups did not differ significantly ($t = .437, p > .05$). In other words, before performing the experiment, the pre-test revealed that control and experimental group demonstrated a similar understanding of the learning topics at an alpha level of 0.05.

Post-test: After participating in the learning activity, the two groups of students took a post-test. The t-test results of the post-test indicated that the experimental group had a higher mean score than the control group. Furthermore, the results show that the learning achievement of the experimental group was significantly better than that of the control group ($t = -.2318, p < .05$). That is, the experimental

results indicate that the proposed STLR approach is an effective approach for enhancing student reading comprehension.

Table 1: Paired t-test of the pre-test and post-test results

Test	Group	N	Mean	Std. Deviation	Std. Error Mean	t-test
Pretest	Control Group	34	60.58	23.8609	4.0921	t = .437
	Experimental Group	34	58.01	24.4024	4.1850	p = .665
Posttest	Control Group	34	60.59	13.8041	2.3638	t= -2.318
	Experimental Group	34	67.65	14.4723	2.4819	p= .027*

*p < .05

4.2 Analysis of Perceived Ease of Use and Usefulness

To better understand the students' perceptions of the use of the STLR learning system, this study collected the students' feedback on the "perceived usefulness" and "perceived ease of use" of the system. The results indicated that most students gave positive feedback concerning the two dimensions of the STLR learning system. The average ratings for "perceived usefulness" are 3.87 and 2.82 for the experimental group and the control group, respectively; moreover, their average ratings for "perceived ease of use" are 4.14 and 3.89. In comparison with the ratings given by the control group, it should be noted that the students in the experimental group gave higher ratings for "perceived usefulness" and "perceived ease of use", implying that the students who learned with the STLR learning system revealed higher degrees of technology acceptance than those who learned with the mobile learning system.

In terms of perceived usefulness, the t-test result ($t=3.68$, $p<0.001$) shows a significant difference between the experimental group and the control group. From the students' interview feedback, it was also found that most of the students agreed that the proposed A semantic tag-based enhanced learning recommendation learning system is capable of helping them easily comprehend and learn the benefits of learning activity, and can help them improve their learning efficiency.

4.3 A Survey on teacher consensus with STLR recommendation mechanism

To evaluate the validity of the personalized prior knowledge recommendation quality, an experiment has been conducted by arranging five experienced experts. The experts were asked to rate importance and feasibility for STLR recommended prior knowledge based on their knowledge of the literature using 10 point scales. The rating of 10 represents the highest representative.

From the experimental results, it was found that a consistent consensus value with the experts' experience and opinion was 8.6, meaning that the content of the recommendation fits the scoring process of the domain experts. That is, the results demonstrate that the recommendation mechanism of STLR system is valid, and thus the proposed STLR approach serves as a useful tool for assisting teachers with student literacy assessment.

4.4 A Survey on teacher satisfaction with STLR and learning monitoring mechanism

To evaluate the effectiveness of the learning monitoring for a semantic tag-based enhanced learning, teachers' comments from the exit survey were also analyzed. Five participating teachers from a senior high school in Taiwan were invited to experience the use of STLR teaching interface. A questionnaire and interviews were conducted for collecting teacher feedback. From this data, there were two main findings:

- (1) Most of the teachers believed that annotated tag can enhance the ability of the students for knowledge acquisition. Moreover, those teachers considered that another reason for the effectiveness of the STLR systems is the ability to aggregate tags added by different students, resulting in a list of weighted tags. This means that each tag is not only present or not for each article, but it may have a weight of relevance, indicating how relevant the tag is for the article. As

such, the proposed learning assistance and teaching mechanisms can provide a different dimension to accumulate more learning experiences which can then make students not only aware that the text has structure but also familiar with the cues that exist within the text.

- (2) All teachers agreed that the learning monitoring mechanism is helpful to acquire more knowledge about the learning status of students and effective to decrease the teaching load. This teaching interface can assist teachers in evaluating the reading ability of the students efficiently, enabling constructive suggestions to be given to students, and allowing for the improvement of tutoring strategies. Furthermore, STLR provides a new way to assist teachers in exploiting tag information in the students' tag to capture student attention and keep the attention focused.

5. Conclusions and future work

The study is to propose a Semantic Tag-based enhanced Learning Recommendation approach (STLR) to provide insight on learner comprehension and activity as well as learning monitoring, which is useful for both students and teachers. Moreover, tag extension and application involves using semantic web and its related technologies to generate tag clouds of semantic concepts, which also provides a new way to extract meaningful learning results and to analyze learning behavior.

The experimental results showed that experiences with applying tagging and integrating tag navigation helped students improve their reading comprehension and quickly grasp the structure and concepts of English articles. Moreover, the proposed STLR approach is also helpful in assisting teachers evaluate student learning achievements by tag-cloud visualization. These finding and results can thus provide instructors and administrators with suggestions and references for the design of efficient mobile learning in the future. Further, these indications further warrant future experiments in social tagging applications in collaborative learning environments to improve reading and recommendations incrementally.

Acknowledgements

This work was supported by the National Science Council of Taiwan, R.O.C., under contract numbers MOST 106-2511-S-178-001.

References

- Chen, J. M., Chen, M. C., & Sun, Y. S. (2010). A novel approach for enhancing student reading comprehension and assisting teacher assessment of literacy. *Computers & Education*, 55(3), 1367-1382.
- Chen, J. M., Chen, M. C., & Sun, Y. S. (2014). A tag based learning approach to knowledge acquisition for constructing prior knowledge and enhancing student reading comprehension. *Computers & Education*, 70, 256-268.
- Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection*. Macmillan International Higher Education.
- Falk, J. H., & Dierking, L. D. (2018). *Learning from museums*. Rowman & Littlefield.
- Klašnja-Milićević, A., Vesin, B., & Ivanović, M. (2018). Social tagging strategy for enhancing e-learning experience. *Computers & Education*, 118, 166-181.
- Hwang, G. J., & Chang, S. C. (2016). Effects of a peer competition-based mobile learning approach on students' affective domain exhibition in social studies courses. *British Journal of Educational Technology*, 47(6), 1217-1231.
- McNAMARA, D. S., Ozuru, Y., & Floyd, R. G. (2017). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International electronic journal of elementary education*, 4(1), 229-257.
- Peng, Y. P. M., Zhang, Z., & Ho, S. S. H. (2019). A Study on the Relationship among Knowledge Acquisition Sources at the Teacher-and College-Level, Student Absorptive Capacity, and Learning Outcomes: Using Student Prior Knowledge as a Moderator. *Educational Sciences: Theory & Practice*, 19(2), 22-39.
- Sung, H. Y., & Hwang, G. J. (2018). Facilitating effective digital game-based learning behaviors and learning performances of students based on a collaborative knowledge construction strategy. *Interactive Learning Environments*, 26(1), 118-134.

Designing Pre-test Questions as Phone Notifications: Studying the Effects of a Mobile Learning Intervention.

Ingrid Yvonne HERRAS^{a*}, Don Romielito N. ABANES^b, Nico B. DEL ROSARIO^c & Jonathan DL. CASANO^d

^{abc}*Undergraduate, Ateneo de Naga University, Philippines*

^d*Instructor, Ateneo de Naga University, Philippines*

*herras.ingrid@gmail.com, abanes.donr@gmail.com, nicodlrsr@gmail.com,
jonathancasano@gmail.com

Abstract: Mobile devices are increasingly becoming more pervasive and emerging as part of our daily life, particularly with university students. From these devices developed in tandem with face-to-face class interaction it has opened new possibilities for ubiquitous learning. We present our work on designing a smart-phone Mobile Learning application that streamlines pre-test questions into a “set it then forget it” input system where students can answer quiz items as slide-down notifications within the day prior to a scheduled lecture. Teachers are afforded a web application to create pre-tests in advance and review class scores. The study was conducted to first-year Computer Science and Information Technology students of a university in the Philippines. Data collection techniques used in the study used experience questionnaires, usability tests, interviews, and tests of student learning outcomes. Usability testing showed consistent satisfactory scores across three iterations. Results from the Learning Experience questionnaire maps to the general answers from the focused group discussion presenting indicative of a positive learning experience. Evaluation of the pre and post test scores signified that using the mobile application can be an effective substitute to class administered tests.

Keywords: Mobile Learning, Android Notifications, Usability

1. Introduction

The act of teaching students requires several resources, such as manpower and especially time (Baker, Fabrega, Galindo, & Mishook, 2004). In our educational setting, extending class time is not a new idea. Governments spend millions of out-of-school programs such as tutorials, homework assistance and modifying class schedules such as block scheduling to expand learning opportunities for students (Saliva, 2007).

Institutions recognized that the learning environment extends beyond the classroom thus, an area of interest is finding varieties of technological support options that would allow students to engage with learning outside class hours (Vogel, Kennedy, Kuan, Kwok, & Lai, 2007).

As in every traditional classroom setting, pre-tests are common to schools and universities as it allows teachers to preview of what the students know and where to focus next. Given their prior knowledge with test scores, teachers are afforded to make changes to accommodate weak and strong points. Although one limiting point of classroom-based pre-tests is that pre-tests are not time sensitive which take time away from class instruction (Kelly, 2019).

We turn to Mobile Learning or m-learning as a pedagogy that addresses the need for giving more time for classroom instruction. Mobile Learning is learning made flexible with the use of mobile devices to access educational resources and can be shared with others, both inside and outside classrooms (Squire & Dikkers, 2012).

To make efficient use of the limited time available to learn in a classroom-based setting, we describe our work in attempts to redesign the process of conducting class administered pre-tests into utilizing android notifications to deliver pre-test questions day before the actual class lecture while looking into its impacts on the student learning experience.

1.1 Research Questions

This study is aimed to perform qualitative and quantitative approach on designing and developing a mobile learning pre-test tool which answers the following research questions.

- Is using Android notifications as a pre-test tool, usable in conducting pre-test quizzes?
- How will sending Android quiz notifications impact the student learning experience in the context of class-administered pre-tests?

2. Related Work

Current research approaches have adopted to use mobile phones as part of the new learning tools in the curricula. According to Duggan (2015) it was reported that among smartphone users in ages 18 to 29 years old used messaging apps in their daily lives as with the increased use of mobile devices this generation preferred collaborating and gaming than doing serious tasks and enjoyed the “customized, collaborative and interactive learning”. Common systems made using m-learning are implementations of an SMS quiz-based systems where students are sent quiz questions within a given time period (Vogel et al., 2007). This type of systems explored the usability and acceptance of the system outside university grounds and overcome the challenges of a classroom-based setting. Results of the study showed that it is possible to extend learning opportunities even outside the classroom. Empirically, results from these studies suggests that students are more confident and enthusiastic to use mobile applications to study and review for their classes combined with face-to-face traditional learning methods (Mehrotra, Pejovic, Vermeulen, Hendley, & Musolesi, 2016). Although these mobile applications may have some downsides in terms of user experience such as small screen size, problems with navigation, size of messages and speed of processing and connectivity limiting the user’s full engagement with the use of mobile applications (Kumar & Vasimalairaja, 2019).

In this work, we are to set-up a similar environment of mobile learning where students answer pre-test questions directly using a slide-down notification feature that allows less throughput and decrease interaction time as to reduce distractions by opening unnecessary applications thus an increase of engagement. We hypothesize that this simpler interaction would elicit sustained engagement among students. In the next subsections, we will discuss each feature on by one.

2.1 How do users engage with notifications?

Mobile notifications are extremely beneficial to the users, however, at the same time, they are a cause of potential disruptions, since it often requires users’ attention at inopportune moments (Mehrotra et al., 2016). Thus, there are several ways a user can engage with a notification. In this section we discuss on how we can understand the use of notifications for designing a better and effective notification feature in our mobile learning application.

There are five ways a user may interaction with notifications (Pradhan, Qiu, Parate, & Kim, 2017). First is read, this is when the user spends time reading the notification but does not take any action. Second, it can be sometimes be partnered with Read and Dismiss where the user spends some time reading the notification the but does not take an action. Third is Take action, is it another interaction wherein notifications provide users with embedded set of buttons to take quick actions. Fourth is Launch an app, as in this form is the highest level of engagement where a user launches the app corresponding to the notification and lastly, Ignore or dismiss is when the user may ignore the notification.

In summary, although there are numerous ways of how a user may interact with a notification, we take into consideration into using the five types of engagement in creating our application.

2.2 Mobile Learning Applications with Time Adaptability

Mobile learning systems that adapt to the learners learning time schedule is one of the many ways that a learning application can retain the user’s motivation to learn and use the application. A study that

implemented an SMS quiz system feature that sends quiz items for the user to answer. Initial test results showed that learners tend to lose interaction and feel burdened with too much information when learning systems spam quiz items by sending multiple SMS messages at random hours of the day (Li et al., 2010). Improvements had been made to adapt to the students learning schedules resulting to an increase of motivation and participation by synchronizing the users time schedule using time adaptation that studies habitual interaction between user and application (Li et al., 2010).

In our study we see this as one important aspect in our application to adapt and synchronize the sending of quiz items and made our own adaptation algorithm discussed in section 3.2.

3. Methodology

3.1 Scope and limitations

We implemented our system in a university where 18 students are taking up an Introductory Programming Course aging from 18 to 20 years old owning Android phones and run operating systems 6.0 Marshmallow and above as the custom notification features cannot be catered to lower Android operating systems.

3.2 Time Adaptation Algorithm

The application implements a time adaptation algorithm that computes a simple frequency check every time the students answers an item on a specific time of the day. For every user, an array of Preferred Time (PT) is assigned to record and receive quiz notifications. Each PT corresponds to the item during the day. Quiz items are arbitrarily appointed each time. However, an item might not be answered within the PT. In this note, we developed range periods.

Range periods are a set of interval block which varies per preferred time. The idea is to adapt to the student’s habitual time to answer quizzes. As aforementioned, range periods vary in relation to PT and the Quiz deadline which is at midnight. These sets of time will also be a cue to notify or re-notify students about dismissed items.

3.3 Implementation of Web and Mobile Application

The quiz system is packaged into a downloadable mobile application and a web application deployed on the server. Test group A composed of 10 Computer Science Students, upon testing did not have the time adaptation algorithm and solely relied in the static preferred time stated in the demographic’s questionnaire. Test group B of 8 Information Technology students had already been using the mobile application with the Time Adaptation Algorithm.

Test groups had been exposed to pen and paper and notification pre-tests. The intervention of the class instructor remained the same throughout the study as it can be seen in Figure 1.

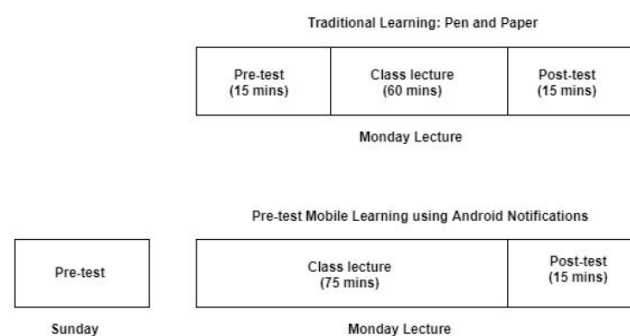


Figure 1. Classroom Design Set-up for Traditional and Mobile Learning.

We used the five concepts of a notification stated in the literature and tweaked the Ignore Notification feature into a Set Reminder when the user dismisses and ignores the quiz item. The Set Reminder feature is designed to remind the student of any unanswered items before the deadline. We opted to use these interactions to provide a byte size chunked of content as a notification for the user and set them to be personalized to their study habit with the Time Adaption Algorithm.



Figure 2. A set reminder notification for dismissed or missed quiz items.

Considering the design of the application, the main activities and functions for the application lets the users log in into the system, receive, and answer the quiz item or be dismissed through notifications and review the quiz result summary.

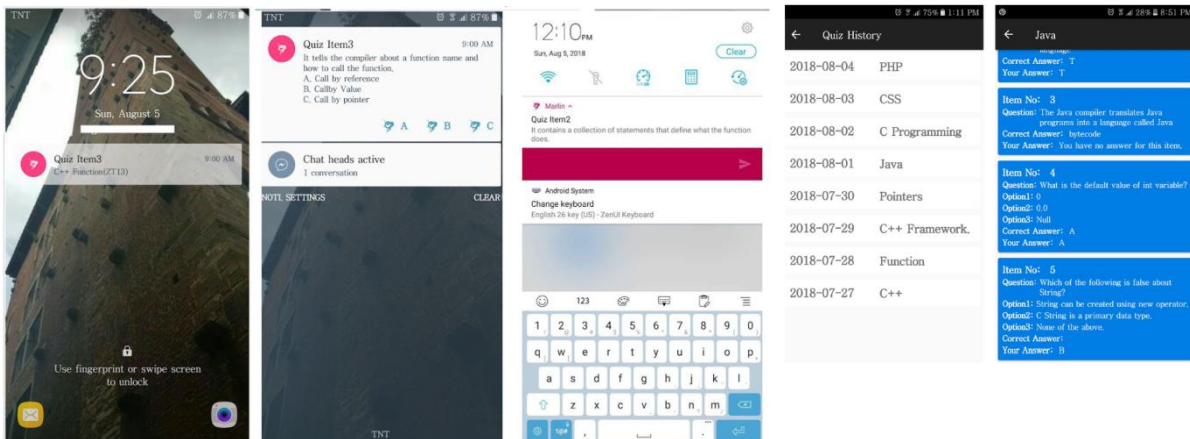


Figure 3. Pre-test items as Android Notifications (Left) and Quiz History Panel for reviewing pre-test items (Right).

4. Results and Discussion

4.1 System Usability Scale Results

To measure our mobile application systems usability, we used Brooke's (1986) System Usability Scale (SUS) as our evaluation matrix to measure usability.

4.1.1 Beta Testing

Preliminary testing revealed that the mobile application fell under the Acceptable-Good with a score of 73.85. Notable comments from the 13 testers were that the application lacked user feedback in terms of loading pages. Respondents were observed dragging down the screen multiple times to refresh pages. Thus, in the next iteration we implemented loading screens or splash screens to improve user experience.

4.1.2 Test Group A and B SUS Results

Results from the SUS questionnaire showed that the mobile application for both test groups fell under the Acceptable-Good category with the scores 71.75 for test group A and 77.18 for test group B. We note that during the implementation the Time Adaptation Algorithm was not yet implemented in test group A while test group B had their mobile learning application integrated with the Time Adaptation Algorithm.

4.2 Learning Experience Results

4.2.1 LEQ Results for Test Group A

For this group, majority or 80% of the participants strongly agreed that answering quizzes through the use of mobile notifications helped in their study habits by reminding them to review via Set Reminder Notification. Although they have been exposed to pen and paper pre-tests, 60% agreed that taking pre-tests on the same day has the same effect in their learning experience as taking quizzes with the mobile application.

4.2.2 LEQ Results for Test Group B

For test group B, 87.5% agreed that they were able to learn more by accessing and revisiting their answered quizzes through the application. The group in general made use of the app as a preparation for their next lecture while the other 12% said that mobile application made no difference in influencing their study routine. Mixed reactions from the group were observed when asked for the difference both methods to their learning routines as it shows that the majority answered Strongly Disagree/Disagree and Neither.

4.3 Focused Group Discussion

In general, both groups commented that they like how mobile pre-tests were sent to them in intervals because they felt that they were not bombarded with too much information at once. Groups A and B felt that they were more confident and prepared during class discussions because they have a sense of the lecture with the learning content sent to them is short and bite-sized content in comparison to condensed learning concepts delivered to them in a traditional classroom setting. However, improvements on the Time Adaption Algorithm can be improved. Most of the interviewees from test group B strongly agreed that delivery of quiz items as slide-down notifications is "... just enough for me to not be overwhelmed of what to study", and "I like how I don't feel stressed with learning all things at once" Both groups generally preferred the dismissal feature because it made them feel in control with the pace they want to learn. A particular instance was told by groups A and B that they formed peer groups during the study. High achieving members group A expressed enjoyment in competing along with the members as to who gets the most answers correct while Group B formed groups to tutor each other after receiving and answering items.

Overall the participants from both test groups strongly agreed that learning through the notification feature is one good way for them to learn as they were more confident for lectures having received their pre-test day before actual class lectures as their mobile phones were easily accessible to them than their laptops.

4.4 Test validation

At this point of the study, our main goal was to test the effectivity of taking pre-tests day ahead of the lecture using android notifications by comparing it to class-administered pre-test. In this evaluation the elements that were not as controlled as we should have wanted were block sections, difference in teaching methods, and topics per lecture session this is because of the lack of teacher distribution between classes and the availability of mobile devices used.

In order to determine if the two differ significantly for written and mobile learning pre-test in both test groups, we computed using paired t-tests. There was a significant difference in the scores for class administered test ($M=2.11$, $SD=1.81$) and mobile learning ($M=3.33$, $SD=1.1882$) conditions; $t(17)=-2.61$, $p = 0.02$. These results show that the use of the mobile learning application is an effective tool for administering pre-tests.

5. Conclusion and Future work

Our work primarily focused on designing and testing a mobile application wherein Android Notifications were utilized to streamline pre-test items and blast them to student's mobile devices before an actual class lecture. Evaluations across three iterations show that using notifications as pre-tests can be applicable and that the inclusion of the Time Adaptation Algorithm increased the SUS score. However, it is not conclusive but only indicative of its effects. Based on the interviews and experience questionnaire, overall students report that they were more confident and prepared during class lectures as they like to take learning at their own pace delivered in bite-sized content. While this study was not explicitly designed to be a collaborative learning tool, reports from the FGD affirms that the application can be an effective collaborative pre-test tool. We can infer that the mobile learning application does indeed enforce a positive learning experience.

For future work, we suggest confirming whether the pre-test questions in text-mode could be another cause of motivation as learners in the "App Generation" seem to prefer the use of video and games.

Acknowledgements

For their encouragement, support, and guidance we would like to thank our mentors, Ms. Michelle Santos, Mr. Glenn Fabia, Mr. Rey Herman Vidallo, Ms. Jenilyn Agapito, and Mr. Frederick Olaño. To the Ateneo Innovations Center and Ateneo Laboratory for the Learning Sciences especially to Dr. Ma. Mercedes T. Rodrigo, Jaclyn Ocumpaugh, and Dr. Joseph Beck for their feedback.

References

- Baker, D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional Time and National Achievement: Cross-National Evidence. *Prospects*, 34(3), 311-334. doi:10.1007/s11125-004-5310-1
- Brooke, J (1986). SUS – A quick and dirty usability scale. Retrieved from <https://hell.meiert.org/core/pdf/sus.pdf>
- Duggan, M. (2016, November 17). Mobile Messaging and Social Media 2015. Retrieved from <http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/>
- Kelly, M. (2019, January 16). The Importance and Use of Pretests. Retrieved from <https://www.thoughtco.com/importance-and-uses-of-pretests-7674>
- Kumar, G. P., & Vasimalairaja, M. (2019). *Digital Education*. New Delhi, India: APH Publishing Corporation
- Li, M., Ogata, H., Hou, B., Hashimoto, S., Liu, Y., Uosaki, N., & Yano, Y. (2010). Development of Adaptive Kanji Learning System for Mobile Phone. *International Journal of Distance Education Technologies*, 8(4), 29-41. doi:10.4018/jdet.2010100103
- Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., & Musolesi, M. (2016). My Phone and Me. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI 16*. doi:10.1145/2858036.2858566
- Pradhan, S., Qiu, L., Parate, A., & Kim, K. (2017). Understanding and managing notifications. *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. doi:10.1109/infocom.2017.8057231
- Silva, E. (2007). *On the clock: Rethinking the way schools use time*. Retrieved from <https://www.air.org/sites/default/files/publications/OntheClock.pdf>
- Squire, K., & Dikkers, S. (2012). Amplifications of learning. *Convergence: The International Journal of Research into New Media Technologies*, 18(4), 445-464. doi:10.1177/1354856511429646
- Vogel, D., Kennedy, D., Kuan, K., Kwok, R., & Lai, J. (2007). Do Mobile Device Applications Affect Learning? *2007 40th Annual Hawaii International Conference on System Sciences (HICSS07)*. doi:10.1109/hicss.2007.18

MOOD: A Mobile Phone-enabled Educational Data Collection Platform

Pei-Yan YUAN ^{a,b,*}, A-Mei DU ^{a,b}, Yong-Bo LI ^{a,b}, Xiao-Yan ZHAO ^{a,b}, Chuan WANG ^{a,b} & Xiao-Dong WANG ^{a,b,*}

^aSchool of Computer and Information Engineering, Henan Normal University, Xinxiang, Henan 453007, China

^bBig Data Engineering Laboratory for Teaching Resources & Assessment of Education Quality, Henan Province, Xinxiang 453007, China

*peiyang, wxd@htu.cn

Abstract: Formative evaluation is an indispensable way to evaluate students and teachers. However, it is difficult to collect classroom behaviors of students and teachers. To deal with this issue, this paper focuses on the design of the mobile teaching platform integrating into the WeChat. We develop the MOOD, a mobile phone-enabled educational data collection APP. MOOD mainly includes digital check-in, homework assignments, online testing, topic discussion, shared library, classroom performance and other modules, so it can collect data and calculate the daily scores of students automatically and scientifically, which enlarges its application area. It is not only beneficial to evaluate students and teachers formatively, but also provides a decision-making reference for the educational administrators.

Keywords: Formative evaluation, mobile teaching platform, crowd sensing

1. Introduction

Formative evaluation is a very important method in education and teaching (Bloom, 1971), which aims to improve the teaching and learning by providing teachers and learners with immediate and effective feedback (Dixson & Worrell, 2016). It is a developmental evaluation based on the continuous observation, recording and reflection of the whole learning process, which makes the students change from passive acceptance evaluation to the initiative participants of evaluation. Besides, formative evaluation helps students break the cognition that the score is the only criterion for learning. The final exam is no longer the only standard to evaluate students' achievement (Yao, 2015). This enables the teacher to understand the student's learning situation fully and discover the problems timely in the teaching process, so that the teacher can obtain continuous feedback in the teaching process (Curry, Mwavita, Holter, & Harris, 2016). Therefore, more and more people pay more attention to formative evaluation and actively carry out formative evaluation in practice.

One key issue in formative evaluation is how to collect the educational data especially the behavioral data of students and teachers. A general way is the questionnaire. However, it is high cost and low efficiency. Moreover, it is difficult to collect procedural data. Crowd sensing or mobile sensing

(Ma, H., Zhao, D., & Yuan, P., 2014) is a new paradigm in pervasive computing, where the owner of a portable device (such as a mobile phone) can be regarded as a sensing source. Mobile teaching platform integrating crowd sensing is an effective tool to implement formative evaluation in colleges. It fits well with the school-running philosophy that everyone can learn anytime and anywhere, and it adapts to new changes of continuous innovation and reform in demand for education. For example, teachers can ask questions and students race to be the first to answer a question. After that, teachers mark stars for the student. In the above process, both teachers and students are viewed as sensors.

Although there are some mobile teaching platforms, it is not used in the campus widely. The existing mobile teaching platforms also have some limitations. Most of them focus on the pushing or sharing of teaching resources and fail to collect classroom behaviors of students and teachers. Therefore, it is difficult to acquire the learning situation of students accurately. Students' learning results cannot be feedback timely. Since process evaluation is an important part of education and teaching, it is necessary to study and design a mobile teaching platform, which not only facilitates the implementation of formative evaluation, but also provides decision-making reference for the educational administrators.

In this paper, we conduct learning and teaching management on the mobile teaching platform. The proposed platform is based on the mobile terminal and is generated by the WeChat public account. Different from traditional platforms, it can collect students' classroom behaviors for formative evaluation and accumulate the data generated in education and teaching to meet the various needs of learning and interaction by being integrated into WeChat. What's more, our proposed mobile teaching platform has been put into use, which makes the theory into practice.

2. Related Work

In 1994, researchers have studied mobile education, mainly in economically developed countries such as Europe and the United States (Churchill, D., & Churchill, N., 2008). Blackboard Mobile (Kinash, & Mathew, 2012) is a smart terminal-based mobile learning platform developed by Blackboard Company of the United States. Relying on the powerful features of Blackboard Learn, Blackboard Mobile can teach and learn anytime, anywhere, and supports smart terminals currently such as iOS, Android, webOS and other operating systems. Jorge Villalon designed a collaborative marking platform in (Villalon, 2019) to support summative and formative feedback in higher education, which includes modules for printing management, scanning support, on-screen-marking, markers training and peer reviews by students. The study of mobile learning in China began with a report by Dr. Desmond Keegan, a famous international distance educator, at the 40th anniversary of Shanghai TV University (Keegan, 2003). Subsequently, China has carried out a lot of research on mobile learning and its application. In 2013, MOOC entered China as a large-scale online platform. The diversification of curriculum resources and the autonomy of curriculum participation provided a good learning environment for students. The effective teaching evaluation cannot be done through MOOC platform, and the subjective content of students' emotions or thinking ability is difficult to reflect (Zhou, P., 2016). For example, in (Yu, J., 2018), the authors proposed a SPOC (Small Private Online Course) platform based on the MOOC concept. This platform can analyze the frequency and residence time of students' online learning, but cannot be applied to the teaching and management in the classroom, since the classroom behaviors of students and teachers cannot be collected. In (Li, J., & Qiu, H., 2016), the mobile teaching platform is designed based on the mobile phone APP. Compared with the mobile teaching platform integrated into WeChat, the cost is higher, students need to install the APP, it is not

conducive to mass promotion, and the platform cannot collect student learning behavior data, so it cannot conduct formative evaluation.

Through research and comparison, some main problems existing in the development of mobile learning platforms at home and abroad. For example, the interaction design is simple relatively, and it is difficult to meet the various needs of interaction for learners in the mobile learning process. It can't be combined the characteristics at the same time such as low-cost, ultra-lightweight, and formative evaluation. Therefore, we need further research on the mobile teaching platform to improve its management functions, especially in the collection of students' behaviors for formative evaluation.

3. System Design

3.1 Framework Design of Mobile Teaching Platform

MOOD is a mobile teaching software based on WeChat development, covering all three situations before, during and after the class. It adopts the micro service architecture design concept (Zheng, Z., Cheng, J., & Peng, J., 2015) to meet the functional requirements of the mobile teaching platform. The core idea of micro service architecture design is to disassemble complex application systems into specific services of multiple independent business units. Each service is implemented with the most suitable technology and runs independently, thus applications which are easy to expand and suitable for mobile terminals are generated. In addition, each service implements a complete application, which has an independent development, deployment and operation. These features assure a high scalability and low coupling. The architecture of the mobile teaching platform is shown in Figure 1. It consists the following layers.

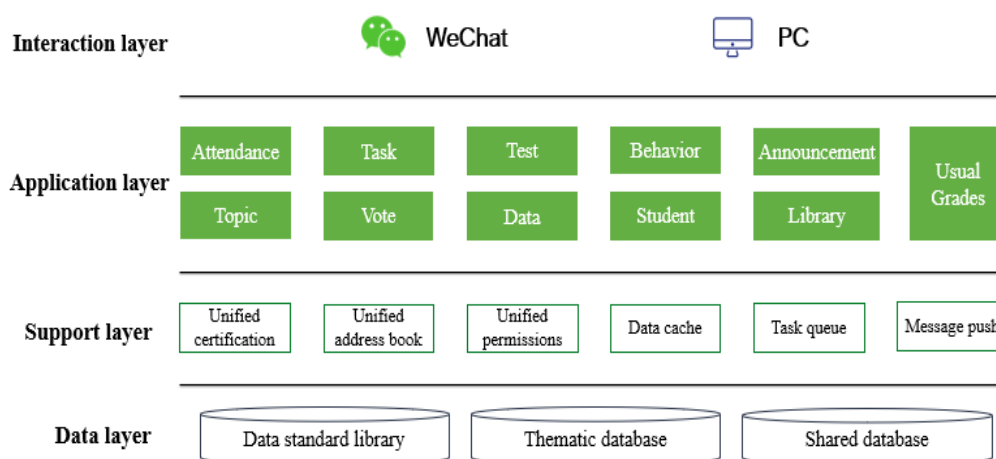


Figure 1. System Architecture

- Interaction layer: The mobile teaching platform supports the login by WeChat and PC, allowing students to use the mobile terminal to learn anytime and anywhere.
- Application layer: It consists of multiple independent services, and has multiple function modules such as check-in, operation, test, performance, etc.

- Support layer: As an integrated environment for information, it can aggregate distributed applications and information resources. It also realizes seamless access and integration of various application systems. In addition, it provides an integrated environment that supports access, delivery, and collaboration of information.
- Data layer: The data layer is divided into data standard library, special database and shared database. It not only has powerful test database resources, but also enables online testing at any time.

3.2 Function Design of Mobile Teaching Platform

The mobile teaching platform is divided into two parts: the student and the teacher. Students and teachers have the same functional modules such as check-in, homework, test, library, and grade. Through the student terminal, students can conduct related course learning, access to platform learning resources, sign on the platform, submit assignments and participate in topic discussions. Teachers use the platform to achieve course teaching, share learning materials, have an attendance checking and a classroom testing and arrange some assignments. The main interface of the mobile teaching platform is shown in Figure 2. The specific functions are as follows:

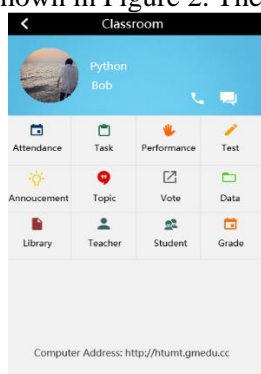


Figure 2. The Main Interface of the MOOD

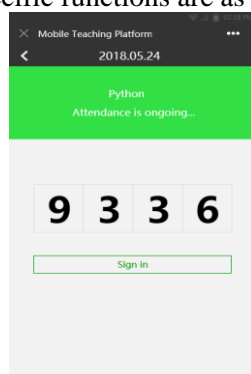


Figure 3. Digital Attendance

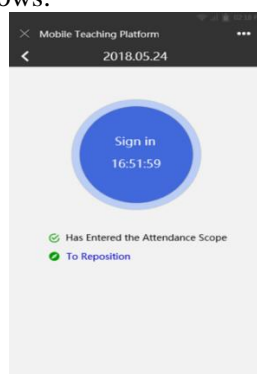


Figure 4. GPS Attendance



Figure 5. QR Code Attendance

3.3 Student Terminal

- Attendance: Students can sign on the platform according to the teacher's attendance method, after they log into the mobile teaching platform on the mobile phone. The attendance mode includes: (a) Traditional Attendance; (b) Digital Attendance as shown in Figure 3; (c) GPS Attendance as shown in Figure 4; (d) QR Code Attendance as shown in Figure 5.
- Job Submission: The mobile teaching platform supports students to submit pdf, word, excel, ppt and image format jobs. They can also update their assignments and resubmit them before the deadline. After the teacher corrects the assignment, the student can select the desired assignment to view.
- Classroom Performance: It is mainly a record of answering questions in the classroom. After the teacher releases the answer, the students enter the classroom and click on the performance to answer. The operation interface is shown in Figure 6.
- Participate in the Test: After the student enters the class list on the mobile phone, select the class, click the test, enter the test interface, select the test to be tested, or check the score after the test.
- Classroom Members: After the students enter the class, they can click on the teacher or classmate to view the member information in the class, as shown in Figure 7.
- Data: After the student enters the classroom on the mobile phone, click on the data to view the information published by the teacher in the classroom, as shown in Figure 8.

3.4 Teacher Terminal

- Class Attendance: After entering the class list, select the class, click attendance, enter the attendance interface, enter the attendance name, select the attendance mode, and then the teacher can check if the student is present. After the attendance is completed, the teacher can modify the attendance status manually, as shown in Figure 9. The same as student terminal, the class attendance mode includes: (a) Traditional Attendance; (b) Digital Attendance; (c) GPS Attendance as shown in Figure 10; (d) QR Code Attendance.

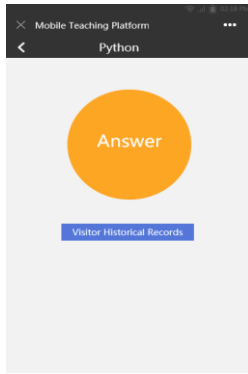


Figure 6. Classroom Performance

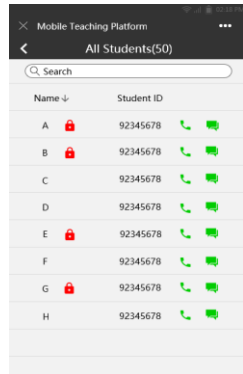


Figure 7. Classroom Members

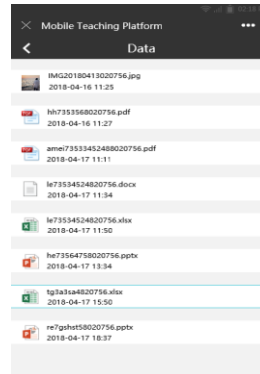


Figure 8. Data

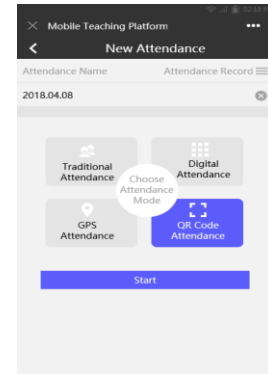


Figure 9. Class Attendance

- Classroom Work: the platform allows teachers to release tasks, which include: (a) Create A Job (b) Correcting Assignments (c) Job Settings: Select to hide or display grades for students; to remind students who have not submitted assignments, or to extend the deadline of assignments; teachers can click on the marker button to mark student assignments as needed; (d) Job Review: In the Job discussion interface, teachers and students can add comments to the assignment.
- Classroom Performance: After entering the class list, select the class, click on the performance to enter the performance interface, select the question mode to ask questions, and it divides into three questioning methods: random question, race to be the first to answer a question, and name-calling question, as shown in Figure 11. In the performance interface, the teacher can view previous questions and class summary. As shown in Figure 12, the performance interface also includes the interface of waiting for answer.
- Voting: At present, only the mobile phone supports the voting function. After the teacher enters the classroom, he/she chooses vote button. Click the plus sign on the upper right to add a vote, edit the voting content, and begin to vote after setting.

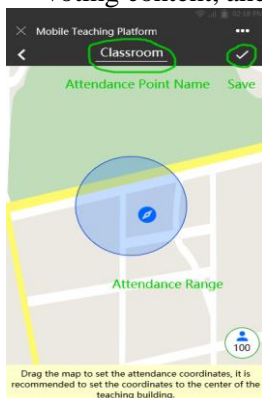


Figure 10. GPS Attendance

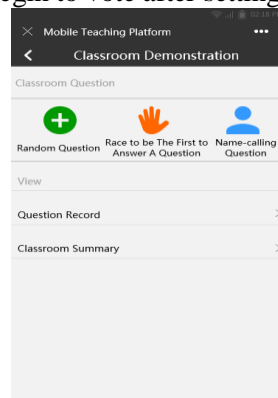


Figure 11. Three Methods

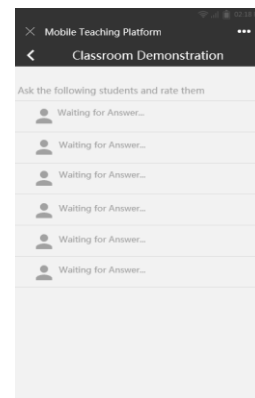


Figure 12. Answer

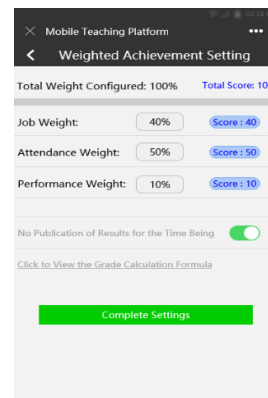


Figure 13. Grade Weighting

4. Implementation and Evaluation Results

We designed a mobile teaching platform software using the convenience of the WeChat public platform, which can collect data on students' daily grades. In this section, we show the relationship between the usual grades (daily performance) and the final exam grades, as well as the impact of the homework achievements, attendance results and classroom performances of the daily grades on the final exam grades. When calculating the usual grades, the teacher can click the score weighting setting on the total score interface to assign weights to each grade component, as shown in Figure 13. After the teacher performs the score weighting setting, click to get the latest data, and the system will calculate the usual score automatically.

In order to verify the positive impact of the usual grades on the final exam grades and the auxiliary role of the mobile teaching platform in student learning and teacher teaching, this paper uses

the mobile teaching platform to collect the students' usual learning data and final exam data. This data comes from 89 undergraduate students. We recorded a semester learning data for the PHP programming course through the mobile teaching platform, which made the learning evaluation method diversified, not only rely on the final exam scores. It can be seen from the change trend of the daily performance and the final exam grades, as shown in Figure 14. The daily grades have a certain influence on the final exam grades. Through analysis, 49 students in the final exam grades are above the median score. There are 54 students in the usual grades above the median score. It can be seen that the students with good daily grades have relatively good final exam grades. Through the analysis of Figure 15, it can be obtained that the homework grades are more correlated with the final exam scores. Among the homework results, 64 students are above the median scores. For the attendance, it is impossible to clearly see its impact on the final exam grades. This is mainly because most students have a higher attendance ratio. In addition, the classroom performances have not been analyzed because of the page limitation. In summary, paying attention to students' daily grades is very helpful for improving students' performance and teachers' guidance for students' learning, and the homework grades can better reflect students' learning.

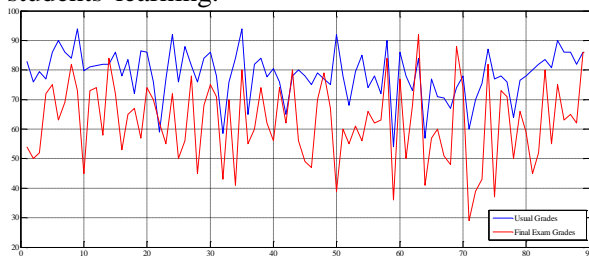


Figure 14. Daily Grades and Final Exam Grades

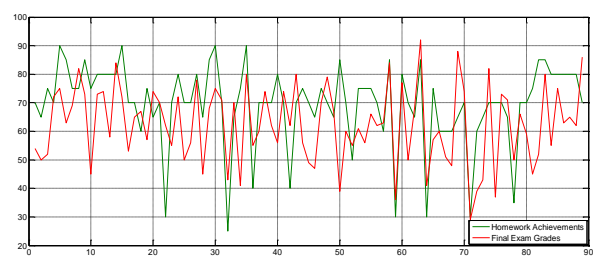


Figure 15. Homework Achievements and Final Exam Grades

5. The Conclusion

The mobile teaching system implemented in this paper is a third-party mobile learning platform based on the open interface function of Tencent WeChat public platform. The structure function is clear and the operation is simple. The system function includes various requirements of education and teaching. Teachers use this system for supplementary teaching, which is convenient and simple. Students only need to pay attention to the course WeChat public number loaded with this system, and they can use mobile intelligent terminal to study, discuss and evaluate.

The mobile teaching platform can monitor and record the student's learning progress and test situation, and provide a decision-making reference for the teacher when they monitor students' learning process. It is beneficial to evaluate students and teachers formatively. Due to the real-time interaction of the functions, it enhances students' ability to collaborate and explore when they discuss in a group and have a test in the class. And it improves students' initiative in learning while supervising students' classroom behavior. The model is not only low in development cost but also easy to implement, and it is ultra-lightweight and highly stable. And the mobile teaching platform has a good cross-platform nature. It can be used on both the WeChat and PC sides. Due to its technical simplicity and convenience, students are easy to use and it is conducive to mass promotion.

The mobile teaching model based on the WeChat public platform has practical application value. For related teachers, it is necessary to further develop the superiority of the WeChat public platform in the future work. We still have a long way to go especially in the use of mobile teaching platform to monitor student learning behavior. And we should find more new paths that will help to carry out mobile teaching. Finally, we can reach the goal of further improving the quality of teaching.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant U1804164 and U1404602 and Science and Technology Development Center of Ministry of Education under Grant 2018A03008, and in part by the Science and Technology Foundation of Henan Educational Committee under Grant 19A510015.

References

- Bloom, B. S. (1971). Handbook on formative and summative evaluation of student learning.
- Churchill, D., & Churchill, N. (2008). Educational affordances of PDAs: A study of a teacher's exploration of this technology. *Computers & Education*, 50(4), 1439-1450.
- Curry, K. A., Mwavita, M., Holter, A., & Harris, E. (2016). Getting assessment right at the classroom level: Using formative assessment for decision making. *Educational Assessment, Evaluation and Accountability*, 28(1), 89-104.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159.
- Ma, H., Zhao, D., & Yuan, P. (2014). Opportunities in mobile crowd sensing. *IEEE Communications Magazine*, 52(8), 29-35.
- Keegan, D. (2003). *Distance training: Taking stock at a time of change*. Routledge.
- Kinash, S., Brand, J., & Mathew, T. (2012). Challenging mobile learning discourse through research: student perceptions of Blackboard Mobile Learn and iPad. *Australasian journal of educational technology*, 28(4), 639-655.
- Li J, Qiu H. (2016). Design and implementation of auxiliary teaching APP. *Fujian computer* 32(9), 111-112.
- Villalon, J. (2019). Emarking: A collaborative platform to support feedback in higher education assessment. In *Foundations and Trends in Smart Learning* (pp. 113-118). Springer, Singapore.
- Yu, J. (2018). Research on Key Development Technologies for SPOC Platform. In *MATEC Web of Conferences* (Vol. 232, p. 01025). EDP Sciences.
- Yao, Y. (2015). Teacher Perceptions of Classroom Assessment: A Focus Group Interview. *SRATE Journal*, 24(2), 51-58.
- Zhou, P. (2016). The Exploration and Inspiration on SPOC Localization. *Journal of Chendu University: Educational Sciences Edition*. Issue 5, 108-112.
- Zheng, Z., Cheng, J., & Peng, J. (2015). Design and implementation of teaching system for mobile cross-platform. *International Journal of Multimedia and Ubiquitous Engineering*, 10(2), 287-296.

Analysis of Current Situation of Classroom Interactive Teaching Based on Mobile Devices: A Case Study of Middle School Mathematics Classroom Teaching

Jing-Wen PAN*, Di-Fei WANG, Mian LIU

Department of Education Information Technology, East China Normal University, China

*pan_jingwen@126.com

Abstract: With the development of mobile technology and the popularization of smart devices, mobile learning has penetrated into people's life. In this paper, by referring to the interactive analysis system of different experts and scholars, and combining the actual situation of this study, the content analysis method of video is adopted to carry out the classroom interactive analysis of a junior middle school mathematics class using mobile devices for teaching. The study found that mobile devices to some extent effectively support classroom interaction. The specific performance is as follows :(1) the Android tablets mainly supports the learning behavior of students' independent inquiry; (2) the class supported by the Android tablets can provide instant feedback on students' answers; (3) Android tablets can help students' explicit thinking and support their thinking development.

Keywords: Classroom interaction, Mobile devices, Middle school mathematics

1. Introduction

The 43rd statistical report on the development of Internet in China (February 2019) pointed out that as of December 2018, 829 million Internet users were in China, and 29.8% of them used tablet computers to surf the Internet. In terms of place of use, the proportion of students accessing the Internet in school rose to 22.1 percent in 2018 from 19.1 percent last year. From the perspective of users' occupation structure, students account for the largest proportion of Internet users, as high as 25.4%. In terms of the usage rate of individual Internet applications, a total of 202.13 million netizens conducted online education, with a usage rate of 24.3 percent, an increase of 29.7 percent over last year. With the development and upgrading of mobile technology and the popularization of smart devices, it has successfully laid a foundation for mobile learning and provided more opportunities for people to learn lightweight, fragmented and structured knowledge through mobile terminals. Therefore, mobile learning has become one of the important trends in the application of new technology in education. Mobile learning is an educational interaction through mobile technology that students can access from any place (Traxler, 2013), which is also considered as the use of mobile devices and Internet connection for educational purposes (Kinash, Brand, & Mathew, 2012). Many studies have shown that in a learning environment, mobile learning can build and maintain the capabilities of creativity (Ranieri & Bruni, 2013), collaboration (Niels & Carreira, 2018), interactivity (Shen, Wang, & Pan, 2010). In traditional classroom teaching, the classroom interaction between teachers and students is only through simple interpersonal interaction and interaction based on simple mediation (such as blackboard). The form of interaction is monotonous, insufficient in depth and unidirectional. In the current classroom teaching, other technical equipment has been provided to improve classroom teaching, such as tablet computers. Because the tablet supports instant feedback, it provides more support for the real-time and efficient classroom interaction. Based on this, this study selects tablet computer as a third-party intermediary media for teacher-student interaction, and studies its influence on classroom interaction from multiple dimensions such as interpersonal interaction, human-computer interaction and mediation-based

interaction. The research will focus on two research issues: first, the supportive role of tablet computers for classroom interaction; second, the depth of classroom interaction.

2. Methods

Classroom interaction is the basic form of classroom teaching (Qu Bo & Sun Li-li,2019). The nature of interaction requires participants to understand the meaning of interaction with each other. Therefore, classroom interaction is related to the effectiveness of classroom teaching. The study will select “classroom interaction” as the entry point for video image analysis, and analyze the “classroom conversation” by using the content analysis method of classroom video. This research method encodes interactive behaviors and languages based on specific types of functions and analyzes classroom interactions. Based on this, this study takes a public course “Image and Nature of a Function” taught by a junior high school teacher in Beijing as a research case. There are 24 students in the class, study the classroom video of this lesson, and then use the classroom interactive analysis system. The classroom recorded video is sliced and coded to provide an in-depth analysis of the classroom interaction.

3. Interactive analysis and coding system based on mobile devices

In the 1960s, American scholar Ned Flanders proposed a system for classroom interaction analysis (Flanders Interaction Analysis System, FIAS). FIAS consists of three parts: a coding system that describes the interactive behavior of the classroom, a prescribed standard for observation and record coding, and a matrix table for displaying and analyzing the data. Although FIAS can present the classroom structure, behavioral patterns and teaching styles in a quantitative way, there are also many limitations, such as neglecting the interaction between teachers and students and the classroom environment; interaction analysis is limited to the interaction of the established classification, and Ignore the true intentions of teachers and students. Based on this, domestic scholars have made improvements to the system. For example, Gu Xiaoqing (2004) added technical classes based on FIAS and further refined these four categories into 18 coding rules. Feng Wisdom (2016) improved and optimized FIAS to form a Student Centered Interaction Analysis System (SIAS), which increases student behavior and technology, and is subdivided into 18 Article encoding rules. The author will refer to the research results of various scholars and combine the actual situation to organize a classroom interactive analysis system that is more suitable for this research. The classroom interaction analysis system used in this paper is shown in Table 1.

Table 1. *Classroom interactive analysis system*

Classification		code	representation	content
Teacher language	indirect influence	1	Accepting emotions	Accept or clarify student attitudes and emotions in a non-threatening manner
		2	Praise encouragement	Encourage and praise students' behavior through language or software
		3	Adoption of opinions	Adopt and repeat the student's speech perspective
		4	submit questions	Ask students questions and expect students to answer
	direct impact	5	Content teaching	Use the blackboard, technology, etc. to present facts or insights, express the concept of the teacher, and present the teacher's own interpretation or other authoritative scholars' views.
		6	Behavioral instruction	Command the student to do something
		7	Criticize and correct	Teachers criticize or correct students' classroom behavior with authority or self-referential

Student language	8	Passive response	The teacher named the student to answer the question.
	9	Active response	Students take the initiative to answer questions from teachers
	10	Active questioning	Students take the initiative to ask questions to teachers
	11	Brainstorming	Students and peers discuss and exchange
Quiet	12	Silent confusion	Temporary quietness or confusion that does not help teaching
	13	Thinking problem	Students thinking about problems
	14	Classroom practice	Students use paper or pencil or school tools for classroom exercises
Technology	15	Teacher operation technique	Teachers use technology to clarify ideas, teach demonstrations, evaluate works, view student data, etc.
	16	Student operation technique	Students use technology to do experiments, practice, express opinions, view materials, etc.

4. Analysis results and discussion

Through the method of case study, the author selects a junior high school mathematics class "Image and Nature of Once Function" as a case study of this study. The course uses the combination of Rui Xuatang teaching software and GeoGebra drawing software. The study will use a combination of qualitative and quantitative methods to analyze the classroom interaction of mobile devices in this class. The following is an analysis of the classroom interaction in this lesson:

4.1 Time sequence analysis of teaching process

The time sequence analysis of the teaching process of "images and properties of primary functions" is shown in table 2. Through analysis, it is found that teachers mainly guide students to learn by independent inquiry and independent learning, and can carry out situational teaching in combination with real life. In the whole class, the teacher guides the students, enlightens their thinking and enables them to think independently to deepen their impression of what they have learned.

Table 2. *The Teaching Process of "Image and Nature of a Function"*

Time	teaching activities	teaching description		Android tablets support
		teacher	students	
00:31-01:27	Situational import	The picture shows the situation and leads to the teaching content of this lesson — the first function	According to the teacher's guidance, think about the quantitative relationship between x and y	
01:27-08:50	The independent inquiry	Inquiry question: the position relation of the solution of the bivariate quadratic equation in the plane rectangular coordinate system?	Under the teacher's inspiration, students can understand the solution of binary quadratic equation in a straight line through software self-exploration	software : GeoGebra
08:50-11:30	With practice Course teaching	The concept of first order function is derived from the practical problems in life	Learn the definition of a function through the teacher's guidance	
11:30-	The	Q: what does the graph of a	Independent thinking, the use	software :

13:30	independent inquiry	function look like?	of geometric sketchpad software to explore the first function of the image	GeoGebra
13:30-34:50	Autonomous learning Course teaching	Assign tasks, guide learning, ask questions, summarize and comment	Using the geometry drawing board to learn independently, take pictures and upload the function image, and summarize the image properties of the function	software : GeoGebra 、Sharp school
34:50-41:42	Independent practice	Assign tasks, direct, summarize and comment	Draw a graph of the function on the graph paper and upload it. Summarize the drawing method of function image	software : Sharp school
41:42-43:48	Contents summary	Summary of content Classroom reflection	Share the knowledge and methods gained in this lesson	

4.2 The analysis matrix is formed according to the interactive coding system

In the observed case course, excluding the useless parts at the beginning and end of the class, the total length of the class is 44 minutes. According to the coding rules of Ned Flanders, a coding number is determined every 3 seconds, and 879 pairs of coding sequence pairs are finally formed. According to this arrangement, a 16*16 order Ned Flanders transfer matrix is formed (table 3).

Table 3. Interactive Analysis Matrix of "Image and Nature of Once Function" Video

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	2
2	0	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	17
3	0	0	19	1	0	0	0	0	0	0	0	0	0	0	0	0	20
4	0	0	0	73	1	0	0	0	0	0	0	0	0	0	0	0	74
5	0	0	0	0	179	1	0	0	0	0	0	0	0	0	4	2	186
6	0	0	0	0	0	38	1	0	0	0	0	0	0	0	0	0	39
7	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	3
8	0	0	0	0	0	0	0	90	1	0	0	0	0	0	0	0	91
9	3	0	0	0	0	0	0	0	37	0	0	1	0	0	0	0	41
10	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
11	1	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	9
12	0	0	0	0	0	0	0	0	0	0	0	20	1	0	0	0	21
13	0	0	0	0	0	0	0	0	0	0	0	0	12	1	0	0	13
14	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1	0	100
15	0	0	0	4	0	0	0	0	0	0	0	0	0	0	45	1	50
16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	210	211
total	4	17	20	78	180	39	3	92	38	1	9	22	13	100	50	213	879

According to the interactive analysis matrix of video in "the image and nature of the first function", the statistical table of the ratio of interactive behaviors in classroom teaching can be obtained (table 4), through which the classroom can be analyzed in depth from different perspectives. From the perspective of classroom structure, the teacher's speech ratio is 38.79%, the student's speech ratio is 16.27%, the silence ratio is 5.24%, and the technology ratio is 29.69%. It can be seen that the teacher occupies a dominant position in the classroom, and the use of technology is reasonable, and the student's speech rate is also high. Among them, the question asking rate of teachers is 2.28%, and the thinking rate of quiet middle school students is 84.33%, indicating that teachers can guide students to think through questions and have deep classroom interaction with students. Students use technology ratio was 80.84%, the ratio of teachers' use of technology was 19.16%, more can show students the use of more technology issue to explore, practice and browse other students answer, teachers use technology demonstration for operation, answers and comments on student responses from the student, to improve the students' participation and self-efficacy, in order to achieve the depth of the subject and object interaction.

In addition, it can be seen from the diagonals of the matrix in the above table that the three largest data are (16,16), (5,5) and (14,14) successively, which also indicates that students make full use of technology to conduct classroom exercises and knowledge exploration under the guidance of teachers. From the perspective of the emotional atmosphere of teachers and students, teachers and indirect effects directly affect the rate of 49.55%, in line 1-3 and 1-3 columns of the area of the positive grid frequency is 36 times, on line 7-8 and 6-7 column defect of the frequency of 2 times, that teachers pay attention to the interaction with the students in the classroom, active communication with students, be good at to adopt the opinions of the students, the classroom atmosphere is very harmonious whole, student's participation is also very high, and the teacher's direct impact is mainly composed of lectures and guide students to think about, to help students better understand knowledge.

Table 4. Classroom teaching interaction behavior ratio statistics

Analysis content	calculation formula	proportion
Teacher speech ratio	1-7 total number of rows/total number of rows	38.79%
Student speech ratio	8-11 total number of rows/total number of rows	16.27%
Silence ratio	Total number of rows from 12 to 14 / total number of rows	5.24%
Android tablets ratio	Total number of rows of 15-16 / total number of rows	29.69%
The rate of students thinking about problems	The total number of rows from 13 to 14 over the total number of rows from 12 to 14	84.33%
Teacher question ratio	The total number of rows divided by the total number of rows	2.28%
The percentage of students using tablets	The total number of rows divided by the total number of rows between 15 and 16	80.84%
The percentage of teachers using tablets	Total number of rows of 15 / total number of rows of 15-16	19.16%

As a whole, this case video lesson in the technology and the classroom teaching can fully mix, and conducive to the students in the classroom interaction and teacher, student and equipment, teachers and equipment, equipment - students such as deep multi-dimensional interaction, and classroom atmosphere is very harmonious, friendly, is conducive to teachers inspire students to think about, to mobilize students' learning enthusiasm and initiative.

4.3 Mobile device support, interactive depth, interactive feedback and interactive engagement

After the whole process of observing classroom interaction, this study also interviewed several teachers who observed this class on the spot. Teachers basically believe that mobile devices are better supported, can better support interaction, significantly improve students' participation and efficacy, but have not helped students achieve high-level learning goals; for the depth of interaction, teachers It is believed that high-effect interaction can be achieved, mainly in the multi-dimensional and overall interaction; for

interactive feedback, teachers believe that mobile devices can promptly feedback right and wrong, encourage and promote attitudes and emotions; and for interactive participation, teachers believe that interactive participation The degree is good, showing that most students can deeply participate in the interaction.

5. Conclusion

Based on the above analysis, this study believes that mobile devices effectively support classroom interaction to some extent in this class. The specific performance is as follows :(1) the tablet computer mainly supports the learning behavior of students' independent inquiry. In class, students make full use of the learning resources in the tablet computer, and complete the tasks assigned by teachers through resource viewing, inquiry and simulation experiments, so as to help students think actively and promote their self-directed learning. (2) the class supported by tablet computer can provide instant feedback on students' answers. Real-time feedback technology can be used to accurately reflect students' learning data to teachers, so as to make data-based teaching strategy adjustment. For example, through students' answers, students' knowledge confusion points or knowledge blind spots can be systematically explained. (3) tablet computers can help students' explicit thinking and support their thinking development. As a learning tool, tablet computers can help students make their thinking process explicit step by step. Learning scientific research shows that when students express the knowledge they are forming externally, the learning effect will be better, thus promoting students' independent thinking and developing thinking.

The real classroom interaction is to help students to construct knowledge and develop their thinking ability. Although the cases examined in this study are not necessarily representative, teachers' use of mobile devices in class does improve classroom interaction. Therefore, this case can help other teachers to reflect on teaching and take practical actions to some extent. This kind of case study will be more conducive to the practice and development of mobile learning in classroom teaching.

References

- Flanders, & Ned, A. . (1968). Interaction analysis and Inservice training. *The Journal of Experimental Education*, 37(1), 126-133.
- FENG Zhihui. WU Wenchen. & HU Xiaoyong. (2016). Research on the Coding of Classroom Interaction Analysis in the Flipped Classroom. *Journal of Distance Education*, 34(4), 59-64.
- GU Xiaoqing. & WANG Wei. (2004). New Exploration of Classroom Analysis Technology Supporting Teacher Professional Development. *China Educational Technology* (7), 18-21.
- Kinash, S. , Brand, J. , & Mathew, T. . (2012). Challenging mobile learning discourse through research: student perceptions of blackboard mobile learn and ipads. *Australasian Journal of Educational Technology*, 28(4), 639-655.
- QU Bo. & SUN Li-li. (2019). Classroom Interaction Based on Image Ethnology, with a Discussion on the Ontological Value of Video Analysis. *Journal of Capital Normal University(Social Sciences Edition)* (1).
- Ranieri, M. , & Bruni, I. . (2013). Empowering creativity in young people through mobile learning: an investigation of creative practices of mobile media uses in and out of school. *International Journal of Mobile and Blended Learning (IJMBL)*,5.
- Shen, R. , Wang, M. , & Pan, X. . (2010). Increasing interactivity in blended classrooms through a cutting-edge mobile learning system. *British Journal of Educational Technology*, 39(6), 1073-1086.
- Sievertsen, N. , & Carreira, E. M. . (2018). Apoc social: a mobile interactive and social learning platform for collaborative solving of advanced problems in organic chemistry. *CHIMIA International Journal for Chemistry*.
- Traxler, J. . (2013). Mobile learning: starting in the right place, going in the right direction?. *Innovations in Mobile Educational Technologies & Applications*.

Multimodal Recording System for Collecting Facial and Postural Data in a Group Meeting

Yusuke SONEDA^{a*}, Yuki MATSUDA^a, Yutaka ARAKAWA^{bc} & Keiichi YASUMOTO^a

^a*Nara Institute of Science and Technology, Japan*

^b*Kyushu University, Japan*

^c*JST PRESTO, Japan*

*soneda.yusuke.su2@is.naist.jp

Abstract: By the spread of active learning and group work, the ability to collaborate and discuss among the participants becomes more important than before. Although several studies have reported on that micro facial expressions and body movements give psychological effects to others during conversation, most of them are lacking in quantitative evaluation and there are few datasets about group discussion. In this research, we proposed a highly reproducible system that helps to make datasets of group discussions with multiple devices such as an omnidirectional camera (360-degree camera), an eye tracker and a motion sensor. Our system operates those devices in one-stop to realizing synchronized recording. To confirm the feasibility, we built the proposed system with an omnidirectional camera, 4 eye trackers, and 4 motion sensors. Finally, we succeeded to make a dataset by recording 8 times group meeting by using our developed system easily.

Keywords: Group discussion, active learning, multimodal communication dataset

1. Introduction

In the 21st century, the thinking and cooperating abilities gained through discussion have regarded significant skills (Arpan, 2017). A model called elaboration likelihood model (ELM) exists as a hypothesis as a communication model to persuade the listeners, and it is a content that has been actively studied for a long while (Richard, 1986). Small-group discussions are more effective on the retention of knowledge than classes in the large classroom. Hence, the classes utilizing small-group discussions should become mainstream in the future (Philip, Kerstin, Bruce, & Wilson, 2011). To improve the ability of discussion, it is important to assess the group discussion quantitatively. However, there are no related work that has realized easy and quantitative evaluation with multimodal devices.

We can divide communication with human beings into verbal communication and nonverbal one. Nonverbal communication has a large impression on the conversation partner and is a very important ability to build social relationships (Peter, 1987; Michael, 1988). Micro-expressions on the face has a large impression too, and it is important information for understanding the deep psychology of the human (Ekman & Friesen, 1975). On the other hand, analysis of the content of the conversation may include sensitive private information, so it is difficult to introduce a system that analyzes the content of the conversation in an actual educational scene. From the above, it is important to develop a system that evaluates the quality of communication from micro behavior and facial expression, and assists the students participating in the discussion.

Many comprehensive face image datasets for facial expression recognition are famous and some papers report the high accuracy results in emotion recognition (Kanade, 2000; Yan, 2014). But there are few datasets that have been labeled for multiple discussions. In this paper, we propose a system for making datasets about group discussion and describe the consideration of some datasets measured.

This article is organized as follows: Section 2 provides a description of related work, and Section 3 describes a proposed system and the experimental condition. Then, the results of the experiments are shown in Section 4, and Section 5 concludes this paper with future perspectives.

2. Related Work

Hori et al. (2012) developed the system that analyzes face direction and the contents of the statement real time with combining an omnidirectional camera and several directional microphones. But the audio analysis performed by this system is highly influenced by the form of the meeting room, so it works only in pre-optimized rooms. In addition, very expensive and high-spec equipment is required to analyze, it is difficult to introduce in general schools and companies.

Ohnishi et al. (2019) have developed the system that recognizes nodding, speaking and looking motions from acceleration and angular velocity data from a head motion with a 9-axis acceleration sensor. By amelioration of the algorithm, it is expected that the accuracy of recognition will be further improved, this paper reports head motion acceleration and angular velocity is very useful data in recognizing head behavior such as nodding. However, this system needs to synchronize the recorded video with the timestamp of the accelerometer and it is not automated. It is very laborious to match the labeling data of the recorded video with the timestamp of the acceleration sensor by human hands. If there is a time stamp gap between the video and the sensor data, there is a risk that the accuracy will decrease when performing activity recognition.

Thus, to make datasets for group discussions in various educational scenes, the reproducibility of the recording system and the mechanism synchronize each device data are important. We propose a system that has two major features. The first feature is that it is a highly reproducible system which can easily make group discussion datasets and can be used in any room. The second feature is that the proposed system uses multiple types of devices to acquire more data and each timestamp does not slide from the exact time by operating in one-stop.

3. Proposed Methods

3.1 Assumed environment

Four people hold a five minutes discussion two times per experiment. The topic in this discussion does not depend on the knowledge possessed by the participants, each participant can talk equally. We prepared the themes which the answer is narrowed to two ways. For example, “If you have a time machine, which do you want to travel to the past or the future?”, “Which do you like cat or dog?” and so on. From the presented agendas, we asked the participants to select two themes which all of them don’t have the same opinion, and each participant discusses their views. After each discussion is over, we make surveys of discussion for them.

3.2 Devices

We developed a measurement system using three types of devices to make data sets for multi-person communication. The devices are THETA V, Pupil Mobile Eye Tracking Headset, (hereinafter called “Pupil”) and LPMS-B2.

THETA V (<https://theta360.com/en/about/theta/v.html> Last access 15 Aug 2019) is the omnidirectional camera developed by RICOH, it can record video with 29.97 fps frame rate and 4K pixels. In addition, since it can behave like a Wi-Fi base station, it is possible to get shooting commands from a PC or mobile terminal. It records the faces and bodies of every participant, we prepared only one.

Pupil (<https://pupil-labs.com/pupil/> Last access 20 May 2019) is a glasses-type eye tracking system developed by Pupil Labs. From the eye direction and movement, it is possible to measure which direction the participant was looking at. In the case of this study, it is used to detect the person who is watching during the discussion. Pupil connects to a laptop to perform analysis and recording. In this experiment, we used Surface provided by Microsoft. When Pupil measurement application is activated, it can receive commands for recording from the background, so it is possible to send commands for shooting the video remotely using wired LAN or Wi-Fi. Each participant attaches Pupil, we prepared four Pupils and four Surfaces.

LPMS-B2 (<https://lp-research.com/lpms-b2/> Last access 15 Aug 2019) is a lightweight 9-axis wireless motion sensor developed by LP-Research. The measure of this device is $39 \times 39 \times 8$ mm and

weight is 12g. It can transmit acceleration information to a PC or mobile terminal via Bluetooth. OpenMAT, GUI open source application, control and collect data sent from LPMS-B2. LPMS-B2's timestamp recording starts when a measurement is started by OpenMAT. Moreover, since it is possible to simultaneously acquire data from several LPMS-B2s, acceleration data obtained from them don't have a difference of timestamp. In this experiment, the sampling frequency was set to 100Hz, and data on the 9-axis acceleration of the head was acquired by attaching to a Pupil. Each Pupil has LPMS-B2, so we prepared four LPMS-B2s.

3.3 Proposed System

In this research, we define data measured by these three types of devices and data annotated to omnidirectional video as a dataset in the group discussion. As we described in Section 2 (Related Works), for accurate analysis, it is very important that each timestamp of the data acquired by several devices is not out of alignment. We developed a system which sends commands for the THETA V to record the video via Wi-Fi, for Surfaces to record the pupil video via wire connection and for OpenMAT to start acquiring data from LPMS-B2. A diagram of the proposed system is shown in Figure 1.

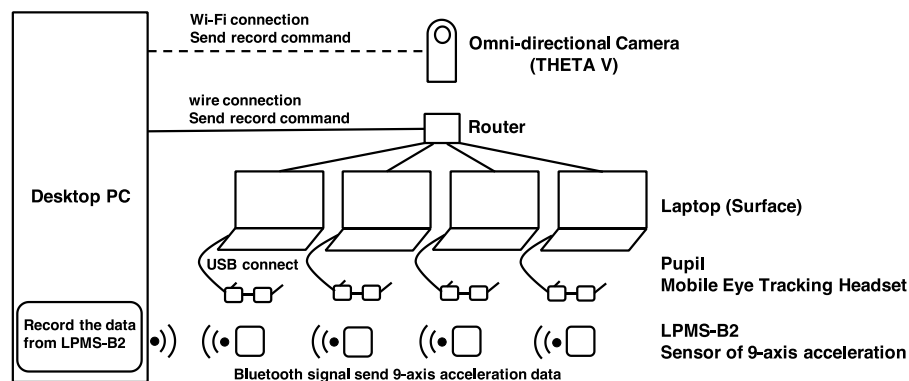


Figure 1. Proposed System Diagram.

Originally it is necessary to manually activate all devices and record a discussion, our proposed system possible to automatically acquire all data simultaneously. Hence, the timestamps of acquired data match with each other, we can make easily high-quality multi-sensor datasets about group discussion.

Figure 2 shows the image which a participant wearing Pupil with LPMS-B2. Figure 3 shows the image of an actual experiment.

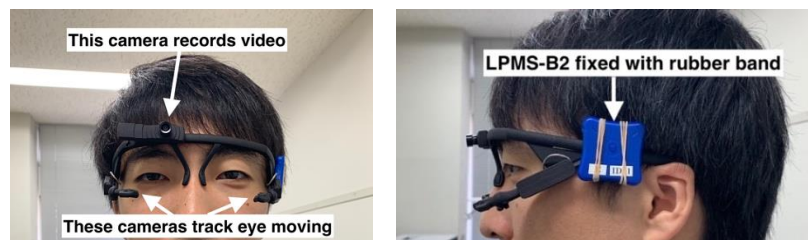


Figure 2. Pupil with LPMS-B2.



Figure 3. Experiment image of group discussion.

3.4 Annotation

Each participant annotates videos of THETA V about their behavior by using ELAN, a software that can perform time-series annotation on video. Figure 4 shows the image of annotating with ELAN. THETA’s video is expanded in a rectangle. When annotating, we ask them to annotate with paying attention to the reason for behavior. Table 1 shows a list of labels of annotation. These labels are decided with reference to the book of Peter (1987).



Figure 4. Annotation with ELAN.

Table 1

The contents of labeling

Label name	Category	Detail
smile	response	does not have special meaning generated by the listener
	agree	smile means consent
	interesting	smile that occurs when the discussion is interesting
	sympathy	seeking empathy
nodding	response	does not have special meaning generated by the listener
	agree	nodding means consent
talk	description	statement when explaining your opinion
	objection	statement when denying the opinion of the speaker
	agree	statement when you agree with the speaker’s opinion
	say	giving the right to speak to a third person
other	-	other behavior

4. Result

4.1 Data Collection

16 students from our laboratory cooperated, we made a dataset of eight discussions. These participants have their native language Japanese, 14 males, 2 females, and ages 22 to 24. The total number of hours is about 70 hours in total to make this dataset. Most of the experiment time was spent mainly on annotation.

After each discussion, we conducted a survey in Table 2. A1 and A2, if the participant’s opinion is the former, the answer is point 1. On the other hand, if it is the latter, the answer is point 5. Therefore, point 3 means a neutral opinion. B1 to B8, the participants answer each item between 1-5 point, closer 1 means “No” and closer 5 means “Yes”.

Table 2

Discussion survey questions

Survey ID	The content of survey
A1	Before the discussion, is your opinion the former or the latter?
A2	After the discussion, is your opinion the former or the latter?
B1	I was satisfied with the discussion.
B2	I could talk my own opinion.
B3	I heard what people with the same opinion say.
B4	I heard what people with the opposite opinion say.
B5	The discussion was enjoyable.
B6	The group often cut off my speech.
B7	I would like to have discussion with this group in future.
B8	I have an attention for the omnidirectional camera.

4.2 Discussion of findings

The results of the survey are shown in Figure 5. Comparing (a) and (b) shows that the participant’s opinion approaches neutral opinion through the discussion. This is because the survey ID B4: “I heard what people with the opposite opinion say.” is 4.125 point on average, so it seems that there was a change in opinion through hearing the opposite opinions.

Survey ID B8: “I have an attention for the omnidirectional camera.” has a score of 1.43 point on average, most people did not care about the omnidirectional camera. Only one person answered 4 point and we interviewed directly after the end of the experiment. The participant said, “I was concerned about omnidirectional camera recording the video”. However, most participants stated, “I was not particularly my mind on the omnidirectional camera”. Placing the omnidirectional camera in the center of the table does not have a psychological impact.

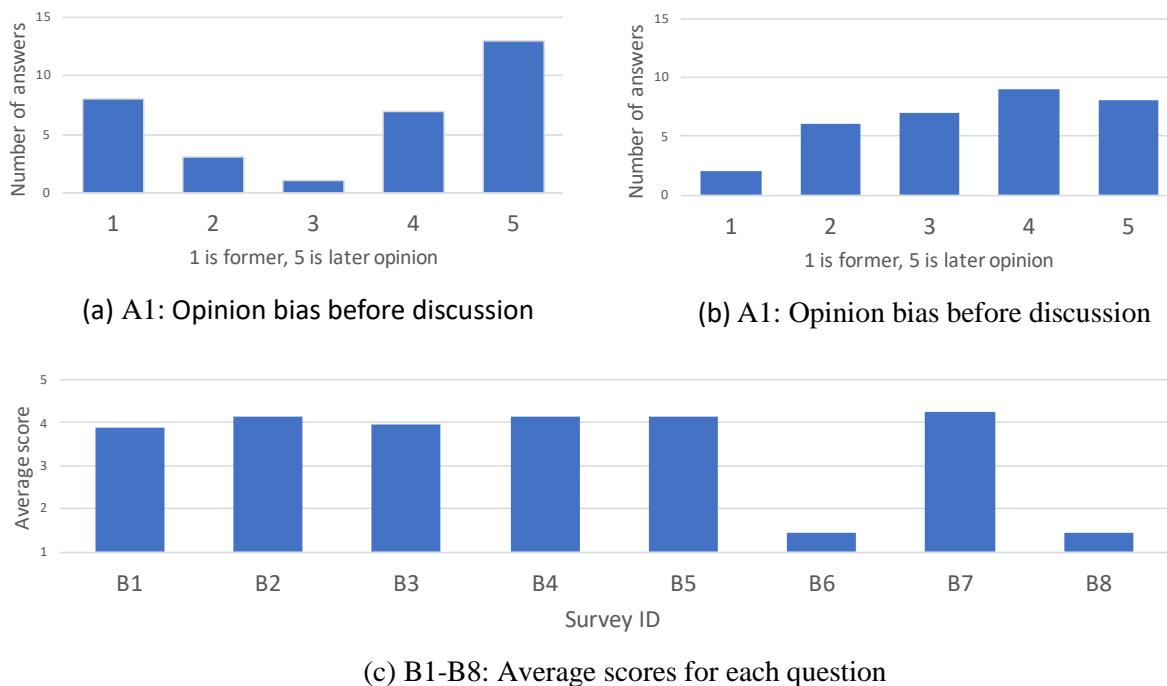


Figure 5. Results for the survey listed in Table 2.

5. Conclusions and Future Work

The ability to discuss in a group discussion is expected to become even more important in the future. We proposed a system for making datasets on group discussions in one-stop, and we have made up a dataset of eight discussions. We confirmed that high-quality datasets could be made by using three types of devices and constructing a system which records consistent datasets. Also, the participants do not attract attention to the omnidirectional camera placed at the center, so our proposed method is considered to be a system that is useful to introduce to the various education scenes. We plan to increase the number of experiments and acquire a dataset.

At the current stage, videos recorded by THETA V directly deal with face images of the participants in the discussion, it is difficult to release them as open data from the viewpoint of privacy. Applying OpenFace: it converts face image to point cloud data, and OpenPose: it converts body image to point cloud data, to the THETA V's video, we will release as open data in the future in a state where it is easy to analysis and protecting private information. There are converted images of the discussion into point cloud data by OpenFace and OpenPose in Figure 6.

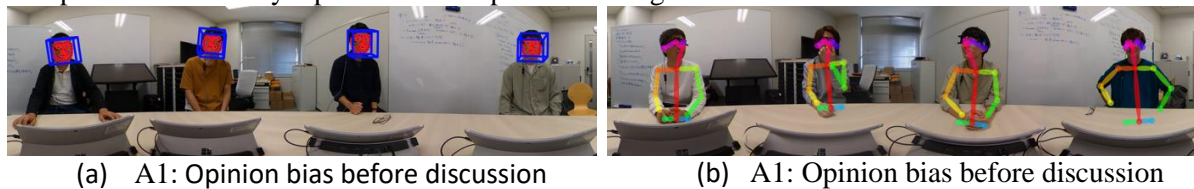


Figure 6. The point cloud data of face and body.

Acknowledgements

This research is partially supported by JST PRESTO and Innovation Platform for Society 5.0.

References

- A. Ohnishi, K. Murao, & T. Terada (2019). A method for structuring meeting logs using wearable sensors. *Internet of Things*, 5, 140-152. doi:10.1016/j.iot.2019.01.005.
- Argyle M (1988). *Bodily Communication*. London: Routledge.
- Arpan. C (2017). *Active Learning through Discussion*. Bridge 1-4.
- Ekman. P, & Friesen. W, V (1975). *Unmasking the face: a guide to recognising emotions from facial clues*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Philip, H. Pollock, Kerstin. H., & Bruce. M. Wilson (2011). Comparing the Benefits of Small-Group and Large-Class Discussions. *Journal of Political Science Education*. 48-64. doi:10.1080/15512169.2011.539913
- Peter. E, Bull (1987). *Posture and Gesture*. Pergamon Press.
- Richard. E, Petty, & John T Cacioppo (1986). The Elaboration Likelihood Model of Persuasion. *Springer Series in Social Psychology. Communication and Persuasion*, 1-24.
- T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, A. Ogawa, K. Otsuka, D. Mikami, K. Kinoshita, T. Nakatani, A. Nakamura & J. Yamato (2012). Low-latency Real-Time Meeting Recognition and Understanding Using Distant Microphones and Omni-directional Camera. *IEEE Transaction on Audio, Speech, and Language Processing* 20(2), 499-513. doi:10.1109/TASL.2011.2164527.
- T. Kanade, Jeffrey, F. Cohn., & Y. Tian (2000). Comprehensive Database for Facial Expression Analysis. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.
- W, J. Yan., X, Li., S, J Wang., G Zhao, Y, J. Liu., Y, H. Chen., & X Fu(2014). CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation. *PLoS ONE*. doi:10.1371/journal.pone.0086041

Study of Augmented Reality Interaction Mediums (AIMs) towards Collaboratively Solving Open-Ended Problems

Pratiti SARKAR^{a*}†, Prabodh SAKHARDANDE^{a†}, Utsav OZA^b & Jayesh PILLAI^a

^aIDC School of Design, Indian Institute of Technology Bombay, India

^bAhmedabad University, India

*pratiti@iitb.ac.in

†These authors contributed equally

Abstract: Open-ended problem solving involves multiple approaches in solving a problem. This can help students to think divergently to relate and apply their classroom learnings in real-life examples. At the same time, through active collaboration, students get to exchange and enhance their knowledge, thus increasing productivity beyond that of an individual. The aim of our study was to develop a collaborative open-ended learning environment using Augmented Reality Interaction Mediums (AIMs) as scaffolds. We conducted a study in a classroom with 12 students of 7th grade who collaboratively used different AIMs to solve certain open-ended problems based on their Mathematics syllabus. We observed their interactions and performance with AIMs as compared to the controlled treatment for each task. Further, we evaluated the creativity through divergent thinking scores using the parameters of fluency, flexibility and originality, where the experimental groups using different AIMs had better creativity score (M=86.3) as compared to the control group (M=79). Thus, a collaborative open-ended approach using AIMs as scaffold can be explored further in improving creative problem solving.

Keywords: Divergent Thinking, Creativity, Augmented Reality, Collaborative Classroom Environments

1. Introduction

The 21st century K-12 education involves a deep understanding of complex concepts, to further creatively generate new knowledge and enhance the Science, Technology, Engineering, Arts, and Mathematics (STEAM) skills. A critical part of STEAM education involves experiential learning, where the learners learn from their experiences and reflect on those with minimal help of the adults. One such approach involves using the learning theory of Constructivism, where students tend to construct their own knowledge (Mughal, & Zafar, 2011). The students can be made to solve a real-life problem by constructing their knowledge on top of prior experiences. This learning in classrooms can be enhanced while exploring the multiple solution approaches in the Open-Ended learning environments with few resources and tools as the scaffold (Biswas, Segedy, & Bunchongchit, 2016). Collaboration among students in this process can further help in exchanging knowledge and developing social skills, critical thinking and creative problem solving ability (Laal, & Ghodsi, 2012).

With the advent of technology, the learnings are now being imparted with one of the emerging technology called Augmented Reality (AR), which helps in superimposing virtual objects in the real world in real time (Azuma, 1997). These virtual graphics can be in the form of images, 3D models, textual information, audio, video, animation, etc. Thus, when it comes to classroom education, AR can be useful as a scaffold in providing affordances that are not readily available in classroom environments. We have attempted to provide such an experience in the school curriculum where students collaboratively explore different ways of using AR in creatively solving open-ended problems.

In this paper, we have discussed an experiential learning study which involved collaboratively solving open-ended problems using different AR Interaction Mediums (AIMs) on a tablet. The broad goal of the study was to understand the interaction of the students with the AIMs and the effect on their creativity while solving open-ended tasks.

2. Background Work

Experiential learning emphasizes learning to be a process of gaining and constructing knowledge through reflection on prior experiences. As per Kolb's Theory of Experiential Learning (Kolb, 1984), the knowledge is constructed in a cyclical manner involving the transformation of experience in each stage. With the benefit of active participation of students, classroom-based experiential learning is thus being highly adopted and implemented (Huang, 2019). Kolb (1984), Piaget (1966) and Dewey (1938) have explored experiential learning through constructivism. Among the pioneers of constructivism, Jean Piaget in his Constructivism Theory states that people generate knowledge and form meanings based upon their experiences (Ackermann, 2001). The theory also states that by the age of 10-14 years, middle school students reach the stage of formal operation with the ability to think logically and conceptualize the things not seen in the actual surroundings (Ojose, 2008). Thus at this age, the students can be guided towards building up their creative imaginary skills.

Vygotsky hypothesized creativity as any human act that produces something new (Vygotsky, 2004), and calling imagination as the basis of all creative activities. Describing it as a complex process of dissociation and association of various elements in new ways towards creation of a new entity, imagination builds upon material supplied by reality (Vygotsky, 2004). There are tests (Guilford (1967), Wallach & Kogan (1965), Torrance (1962)) which evaluate creativity as a measure of divergent thinking. Also, in a study with 7th grade students, it was found that open-ended problems in Mathematics led to an increase in creativity through divergent thinking (Kwon, Park, & Park, 2006). Similarly, our study involves the students to participate collaboratively to help in developing problem solving ability, creativity, critical thinking and social skills (Laal, & Ghodsi, 2012).

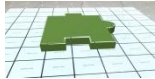
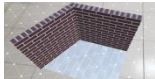
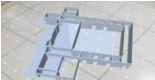

This creativity is brought forth in our study using Augmented Reality (AR). Experiential learning theory in AR suggests that gaining personal experience from AR activities, can enhance the learning achievement of the students (Hung, Chen, & Huang, 2017). Thus, in the classroom, AR as a scaffold can provide an interactive, engaging experience by helping students visualize the concepts which are otherwise difficult to imagine. Our previous work explored the use of AR medium in middle school classroom to collaboratively solve closed problem, showing enhancement in spatial visualization skills (Sarkar, Pillai, & Gupta, 2018). However, along with visualization skills, this study focuses on enhancing the imagination and creativity skills by collaboratively solving certain open-ended problems using different AR Interaction Mediums in the classroom.

3. Design and Implementation

In the study, open-ended tasks were designed (Table 1) as per the 7th grade Mathematics syllabus.

Table 1

Designing the Tasks: The four open-ended tasks given to students to solve

	Topic	3D Models	Task Description	Learning Goals
Task 1	Area	 Field on a 9x9 grid	To think of ways in which its area could be calculated.	To understand what a square unit is. To provoke discussion on the ways in which the area for irregular shapes could be calculated.
Task 2	Lines & Angles	 Walls with 120° angle	To think of methods to find the internal angle between the two walls.	To understand the basis of formation of Lines & Angles and their measurement. To leverage concepts and laws of geometry like parallel lines and adjacent angles.
Task 3	Symmetry & Congruence	 Floor plan	To think of ways to fill this structure (leaving no space) with objects of any shape and size.	To evaluate the ways in which different shapes of different sizes fit together and their fitting gets affected by the scale.
Task 4	Visualizing 3D Solids	 Mountain	To find ways of climbing the mountain in the fastest way possible.	To be able to relate the model to an actual mountain and develop a thorough understanding of the shape as this was a major factor in path and method planning.

In our study, the experimental groups used one of the three different AIMs (Table 2) in a task to solve the given open-ended problems. Each group was also once the control group where the task had to be done seeing a 2D isometric image of the 3D objects, shown to other groups as 3D models in AR.

Table 2

Defining the Augmented Reality Interaction Mediums (AIMs)

Imagine	Draw	Cube
It emphasized creative use of imagination as a method of problem solving. The students could move around this model, viewing it through different perspectives and level of detail by zooming in and out. A three finger tap cleared all elements on the screen.	One could draw in the complete 3D space by moving the phone around and drawing anywhere on the screen with one finger. A three-finger tap completely cleared the screen and a two-finger tap was used to place the task object on a detected plane.	Two tangible marker cubes were provided which overlaid different 3D models, based on the task being performed: Task 1: a scale and a protractor Task 2: a protractor and a ladder Task 3: a round table and a cupboard Task 4: a flagpole (post) and a rope.

The applications were designed in Unity and exported as Android packages on Samsung Galaxy S4 Tablets. Google AR Core SDK was used in the draw and imagine AIMs to enable use of AR. Plane detection and raycasting for placing objects on the plane were the primary AR Core facilities used. In the cube AIM, Vuforia was used to enable tracking. Multi image tracking (cuboidal) was used for the cubes and single image target (QR Code) was used for placing the 3D objects on the given cubes.

4. Method

Our study addresses the following Research Questions:

1. How do students collaboratively interact with AIMs to solve open-ended problems?
2. What is the effect of AIMs on students’ creativity as compared to a traditional medium?

The participants were 7th grade students of a sub-urban Indian school. Through convenience sampling, study was conducted with 12 students (5 boys and 7 girls) of age group 12-14. They were randomly divided into 4 groups of 3 students each. The study was conducted a few days after their end-semester examinations, to ensure they all are familiar with the concepts covered in the AR tasks. Each group was assisted by a researcher to guide them about the tasks and observe their actions. The task and its corresponding AIM for a group was selected using the balanced Latin square design (Figure 1).

	Area	Lines & Angles	Symmetry & Congruence	Visualizing 3D Solids
Group 1	Draw	Imagine	Cube	Control
Group 2	Imagine	Control	Draw	Cube
Group 3	Control	Cube	Imagine	Draw
Group 4	Cube	Draw	Control	Imagine

Figure 1. The balanced Latin square design for task and AIM distribution to a group

The students were encouraged to collaboratively think of multiple answers for a given problem. The interactions were captured using video recording. Observation Logs were used to note group behaviour, involvement and interaction with each other and the AR interface. Further, the students wrote their answers on a sheet, in forms of writing sentences, sketches, diagrams etc. At the end of each task, the students were interviewed about their approach to solve the problem.

The answers of each task written by the groups, were digitized. To answer RQ1, their answers were evaluated, video data was observed, the recorded interview were transcribed and the observation log was assessed to determine the behaviour and approach of the groups in solving the open-ended tasks in each of the different AIMs provided. To answer RQ2, the answers that students gave were evaluated for creativity score by taking inspiration from Guilford’s test of divergent thinking (Guilford, 1967).

5. Results and Discussions

5.1 Collaborative Interaction with AIMS to Solve Open-Ended Problems

5.1.1 Approaches in Solving Tasks

AR has previously helped learners to collaboratively play and visualize mathematics with everyday objects (Khan, Trujano, & Maes, 2018). While solving the Area problem, students suggested materials and objects from their surroundings as the measurement tools. The square grid placed around the field helped them calculate the area in terms of units. All AIMS using groups used this grid in some form of measurement. The Imagine and Draw AIMS groups, split the given field shape into its individual rectangular components to calculate their respective areas as per the learned formulae. However, all students neglected the semicircular shape of the field, for not being able to understand the way of calculating it using the learned formulae.

In the Lines & Angles task, the Imagine and Draw AIMS using groups, initially suggested multiple angle degrees between the walls as per their perspective of looking at the walls using tablets. On viewing bit longer, they realized that the angle was constant and the angles looked different because of different perspective views. The Draw group placed the augmented model of the wall against the actual wall of the classroom. They then hypothesized that the augmented wall model had an obtuse angle. Even in this task, students used objects from their immediate surroundings to compare and measure the angles between the walls.

All the groups related the floor plan to that of their homes in the Symmetry & Congruence task. They thought of ways to place at least 20 household items while considering their shape and space available. The Cube AIM group partitioned the entire floor plan based on size and classified the objects to be placed based on a size proportional to these partitions. This group also suggested organic materials like salt and powdered spices to fill in minute empty spaces. The control group proposed filling up the remaining space by increasing the number of smaller sized objects. The group using Imagine AIM proposed scaling up objects to fill space.

The mountain climbing problem in Visualizing 3D Solids task required students to apply problem solving skills learnt in their syllabus towards the real life scenario of a macro problem. As seen in previous studies (Schneider, Weinmann, Roth, Knop, & Vorderer, 2016), influence of entertainment media was observed where the Control group took an aerial approach to suggest jumping off a helicopter as the fastest way as seen in television. The Imagine AIM group categorized their answers based on risk and speed involved to climb.

5.1.2 Use of Augmented Reality Interaction Mediums (AIMs)

Using the Draw AIM, the group with the Area task communicated their ideas instead of drawing the solutions. In Visualizing 3D Solids, the group precisely drew a ladder and a rope, realizing the need of an anchor with it. In the Lines & Angles task a triangle was drawn on top of the augmented wall. It was then replicated on a paper to calculate the angle. Thus, students identified their own effective and unique ways of using the draw AIM.

With the Imagine AIM, the group of Lines & Angles task on collaboratively discussing and using hand gestures to while viewing the wall from different sides, deduced that it was indeed the same angle which was obtuse. In the Area task, the group calculated area by using objects which were around them, e.g. the field was approximately 10 lunch boxes or 15 pencil boxes large. Thus, they were able to associate and dissociate the meaning of square units.

The cubes and the respective overlaid objects were used as a stimulus for students to answer. The use of cubes was less effective for Visualizing 3D Solids task as they used the rope but not the flagpole in the solutions. In Lines & Angles and Area tasks, students attempted to calculate the exact values using the objects of cubes, considering them to be realistic measuring tools. For the cubes to be within camera view to get tracked one student held the tablet and the others held the cubes. One group, however, placed both cubes on the floor and changed their perspectives by moving the tablet around.

The control group in Visualizing 3D Solids task sketched to communicate complex ideas. They numbered the multiple drawn paths on the hill to discuss the ways of using these paths. In the Area task,

they drew a top view of the field from the given isometric image and used a protractor to measure the drawn semi-circular area. In the Symmetry & Congruence task, the group noted the approximate number of items to cover the floor, e.g. 40 cellphones, 14 newspapers etc. However, this group restricted themselves to their household scenarios and did not think further. Overall, the control group had limited responses due to the inability to visualize the hypothetical scenario on seeing a 2D image.

5.2 Effect of AIMS on Students' Creativity as Compared to a Traditional Medium

The students' answers were evaluated for creativity inspired from Guilford's test of divergent thinking (Guilford, 1967). The answers were evaluated based on fluency, originality and flexibility. The fluency score was the total number of answers given by a group. The flexibility score was the number of categories or different ways of thinking answer. The originality score was calculated as a percentage of uniqueness of the answers. If the answers were rarer than 10%, 20 % or more than 20% of all the answers for a particular task, it was given a score of 2, 1 or 0 respectively. The inter rater agreement of the two raters on the scores of flexibility and originality was 93.15%. The scores were categorized into groups, tasks and AIMS. A creativity score was calculated by adding the scores for fluency, flexibility and originality for a particular category. As shown in Figure 2, Group 2 had the highest creativity score of 155, the task of Symmetry & Congruence had the highest creativity score of 150 and the creativity score of 99 was the highest for AIM Imagine.

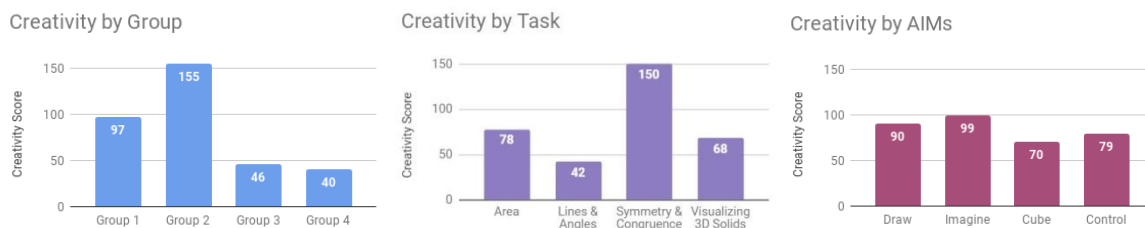


Figure 2. Creativity evaluated across groups, tasks and AIMS

The mean creativity score of the experimental groups (AIMs) was higher ($M=86.3$) than the control group ($M=79$) across all tasks. In terms of creativity, comparatively the Cube AIM lacked. This AIM by its design, provided students with a prompt of two tools to stimulate their thinking ability to find solutions. However, it was observed that their thoughts were limited to the two tools. Thus, even though students liked the Cube AIM's interactive environment, such AIM might be more beneficial for closed problems or convergent thinking tasks. For example, an AR Mathematical education game was developed using three tangible marker cubes to teach certain defined operations (Lee, and Lee, 2008).

In our study, the AIMS provided to experimental groups, helped in visualizing the problem and generating a higher number of creative ideas, as similarly seen in the study by Huang (2019) to enhance students' creativity using AR. The control group had a high flexibility and originality score but a lower fluency score. Group interactions and dynamics were essential in shaping the ideas of students. The discussions were overall positive and helped in the formation of finished solutions, as was seen in our earlier study as well (Sarkar, Pillai, & Gupta, 2018). AIMS provided a stimulus to discussion. All the groups had a positive response to the AIMS, one of the groups, while being the control group, did not want to do the task on paper but wanted to use one of the AIMS. Like observed by Huang (2019), it was seen in our study that prior knowledge and experiences played a major role in the generation of ideas by students. Most of the solutions were directly inspired from either school or household scenarios along with media like TV shows and online videos. Thus, we could claim that AIMS have the potential to provide experiences that are otherwise not possible. The design of AIMS ensures that students use AR not only as a visual tool but also as an immersive experience to think beyond the screen.

6. Conclusion and Limitations

We explored students' approach towards solving open-ended problems in a collaborative environment through Augmented Reality (AR) mediums as scaffolds. We developed three AR Interaction Mediums (AIMs) on tablets providing immersive experiences. These were applied through four different

open-ended tasks based on 7th grade Mathematics syllabus. The results of RQ1 report qualitative differences on the ways in which participants approached various tasks as well as their positive experience using AIMS. The results of RQ2 evaluated creativity based on the multiple solutions that students gave for the problems. We found that the groups which solved the problems using the AIMS had higher creativity score (M=86.3) than the control group (M=79) using traditional pen and paper.

While these results are promising towards a direction in the way AIMS can be used in classrooms, there are certain limitations to our study. The study was conducted with a small sample size (N=12) over a day. Thus, a larger sample size with the study conducted over a longer time would give more in-depth insights. Another limitation pertaining to plane detection and occlusion using Cube AIMS requires the improvement of the technical aspect of the application.

Acknowledgements

The project was funded by Tata Centre for Technology and Design, IIT Bombay. We thank the school authorities, teachers, students and their parents for their consent and valuable time to conduct the study. Thanks to the volunteers at IIT Bombay – Amarnath Murugan, Jonathon Mathew and Amal Dev.

References

- Ackerman, E. (2017). Piaget's constructivism, Papert's constructionism: what's the difference?, 2001. URL http://learning.media.mit.edu/content/publications/EA_Piaget_Papert.pdf.-(URL geprüft: 05/2009).
- Azuma, R. T. (2018). A survey of augmented reality. 1997. Disponível in: <http://www.cs.unc.edu/~azuma/ARpresence.pdf>.
- Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education*, 26(1), 350-364.
- Dewey, J. (1938). Experience and Education. New York: Collier Books
- Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1), 3-14.
- Huang, T. C. (2019). Seeing creativity in an augmented experiential learning environment. *Universal Access in the Information Society*, 1-13.
- Hung, Y. H., Chen, C. H., & Huang, S. W. (2017). Applying augmented reality to enhance learning: a study of different teaching materials. *Journal of Computer Assisted Learning*, 33(3), 252-266.
- Khan, M., Trujano, F., & Maes, P. (2018, June). Mathland: Constructionist Mathematical Learning in the Real World Using Immersive Mixed Reality. In *International Conference on Immersive Learning* (pp. 133-147). Springer, Cham.
- Kolb, D. A. (1984). *Experiential Learning: Experiences as a source of learning and development*, Englewood Cliffs, NJ: Prentice-Hall
- Kwon, O. N., Park, J. H., & Park, J. S. (2006). Cultivating divergent thinking in mathematics through an open-ended approach. *Asia Pacific Education Review*, 7(1), 51-61.
- Laal, M. & Ghodsi, S. M. (2012). Benefits of collaborative learning. In *Procedia-social and behavioral sciences*, 31, 486-490.
- Lee, H. S., & Lee, J. W. (2008, June). Mathematical education game based on augmented reality. In *International Conference on Technologies for E-Learning and Digital Entertainment* (pp. 442-450). Springer, Berlin, Heidelberg.
- Mughal, F. & Zafar, A. (2011). Experiential Learning from a Constructivist Perspective: Reconceptualizing the Kolbian Cycle. *International Journal of Learning and Development*. 1. 10.5296/ijld.v1i2.1179.
- Ojose, B. (2008). Applying Piaget's theory of cognitive development to mathematics instruction. *The Mathematics Educator*, 18(1).
- Piaget, J. (2005). *The psychology of intelligence*. Routledge.
- Sarkar, P., Pillai, J. S., & Gupta, A. (2018, December). ScholAR: a collaborative learning experience for rural schools using Augmented Reality application. In *2018 IEEE Tenth International Conference on Technology for Education (T4E)*(pp. 8-15). IEEE.
- Schneider, F. M., Weinmann, C., Roth, F. S., Knop, K., & Vorderer, P. (2016). Learning from entertaining online video clips? Enjoyment and appreciation and their differential relationships with knowledge and behavioral intentions. *Computers in Human Behavior*, 54, 475-482.
- Torrance, E. P. (1962). Guiding creative talent.
- Vygotsky, L. S. (2004). Imagination and creativity in childhood. *Journal of Russian & East European Psychology*, 42(1), 7-97.
- Wallach, M. A., & Kogan, N. (1965). Modes of thinking in young children.

Supporting Job-hunting Students to Learn Job-hunting Related Terms with SCROLL eBook and InCircle

Noriko UOSAKI^{a*}, Kousuke MOURI^b, Takahiro YONEKAWA^c, Chengjiu Yin^d, & Hiroaki Ogata^e

^a*Osaka University, Japan*

^b*Tokyo University of Agriculture and Technology, Japan*

^c*Brain Signal, Inc., Japan*

^d*Kobe University, Japan*

^e*Kyoto University, Japan*

*n.uosaki@gmail.com

Abstract: In this paper, we describe the support system for job-hunting students to learn job-hunting related terms using an eBook and a chat system. Job-hunting process is very unique and complicated in Japan. Job-hunting students face difficulties in many phases. Some job-hunting related terms are not used in daily conversation and very new to them. Therefore, it is necessary to support them. In fact, many universities in Japan have started providing their students with career education. The objective of this study is to examine whether or not the use of our chat system was effective in learning job-hunting related terms. The result of the evaluation showed there was no statistically significant difference. However the highest score was given when they were asked it's helpfulness.

Keywords: Career education, career support, chat tool, digital textbook, eBook, InCircle, job-hunting, SCROLL

1. Introduction

Job-hunting process is complicated in Japan. It imposes a heavy workload on their academic lives. Students start job-hunting more than one year before graduation. They start with writing CVs (curriculum vitae) and entry sheets, taking exams, written or web-based, such as general knowledge tests, aptitude tests, and personality tests, participating group discussion observed by recruiters, and receiving group interviews and individual interviews at the final stage until they finally obtain official job offers. There are many terms used in job-hunting processes, which are rarely used in daily conversation. The objective in this study is to propose an effective job-hunting related term learning system to facilitate their job hunting in Japan. The system seamlessly supports individual learning and collaborative learning with SCROLL eBook and a chat system called InCircle. The results of our previous study (Uosaki et al, 2019) showed some superiority of SCROLL eBook-based learning to the Blogger-based learning though there was no statistically significant difference. However the effectiveness of InCircle was not confirmed. Our research question in this study is: (1) whether or not the use of our chat system was effective in learning job-hunting related terms. The rest of this paper is constructed as follows. Section 2 describes related researches to clearly identifying the difference between related works and our research. Section 3 describes the design of SCROLL eBook and InCircle. Section 4 describes the evaluation and Section 5, discussion and our conclusions.

2. Related Researches

2.1 Technology enhanced career education

The emergence of IT technologies such as multimedia technologies, Internet technology, ubiquitous and mobile technologies provoked new learning concepts such as WBL (web-based learning), CSCL (computer supported collaborative learning), and MAL (mobile assisted learning) (Ogata & Uosaki, 2012). Besides, various kinds of learning supports have been made into reality by accessing resources of web sites, or by linking learners and numbers of learning objects (Inoue et al. 2014). However career education using information technologies is still in the stage of emergence. There are some reports on ICT implementation to career education such as portal sites for students' career support (Calitz et al., 2015), the use of ePortfolio in career education (Arame et al., 2013), e-Learning in career development for university students (Teshima et al., 200), and e-mentoring for career development (Headlam-Wells et al., 2005). But no such learning system using seamless mobile learning technologies to enhance career education has been developed yet.

3. System Designs

3.1 SCROLL eBook

SCROLL stands for System for Capturing and Reusing Of Learning Log, which has been developed since 2010 (Ogata et al. 2014). SCROLL supports learners to record what they have learned in both informal and formal settings as a log using a web browser and a mobile device and to share them with other learners anytime and anywhere beyond the limits of time and space. This on-going project is still in progress with new functions being added to the system one after the other. SCROLL eBook is one of the functions of SCROLL developed based on EPUB format.

Teachers create e-book contents using PowerPoint or Keynote prior to class and use them in their courses. The uploaded e-book contents are converted to EPUB format and it is supported to access the contents by using smartphones and PCs. Figure 1 (left) shows digital textbooks uploaded by the teachers.

Figure 2 shows the eBook viewer interface and its functions. When a learner clicks the highlight button, he/she can highlight the word. he/she can find the page number corresponding to the target word in the e-book by clicking the search button. When a learner clicks the memo button on the digital textbook viewer system, he/she can write a description concerning the target words.



Figure 1. SCROLL eBook contents. Figure 2. SCROLL eBook viewer interface

3.2 InCircle

InCircle is a product developed by AOS Mobile Inc., Tokyo, Japan with our third author joining this project as a chief software architect. It is a client-server application. The server side runs on Linux OS and Windows Server. The client side is working on iOS, Android, and PC Web browser. Chat messages are transmitted and received through the network (Figure 3).

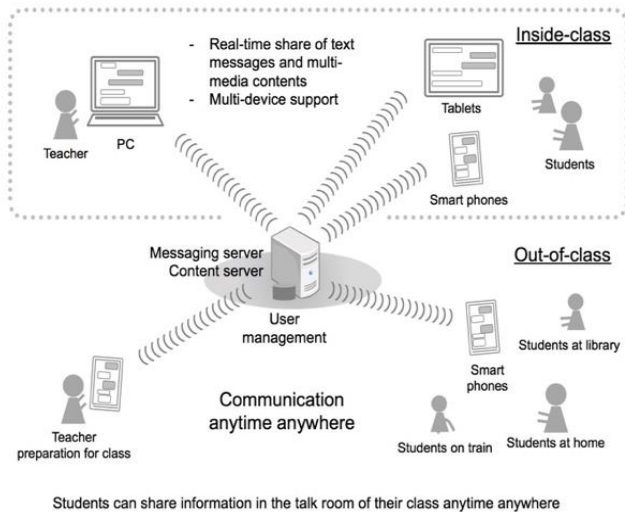


Figure 3. InCircle system configuration

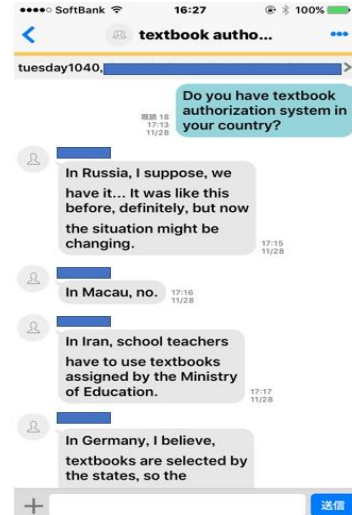


Figure 4. InCircle chat room interface on mobile

The system allows users to create groups. Group members are able to send and receive messages and multimedia files in their chat room with an easy operation. Chat messages are synchronized in real-time to realize smooth communication. Figure 4 shows a chat room interface when the instructor posted an interrogative sentence: "Do you have textbook authorization system in your country?" since interrogative sentences trigger active interaction among learners which leads to mutual cross-cultural understanding.

3.3 SCROLL eBook & InCircle combined learning

In this study, the learning scenario was designed to combine self-learning with group learning. Figure 5 shows SCROLL eBook & InCircle combined learning. Students learn career-related terms with SCROLL eBook alone and interact with other classmates and the teacher at the same time. In order to encourage students to collaborate during the task, the teacher posted a topic which would be helpful to learn career-related terms via InCircle. Students were encouraged to interact via InCircle by telling them the number of InCircle posts will affect their grades.

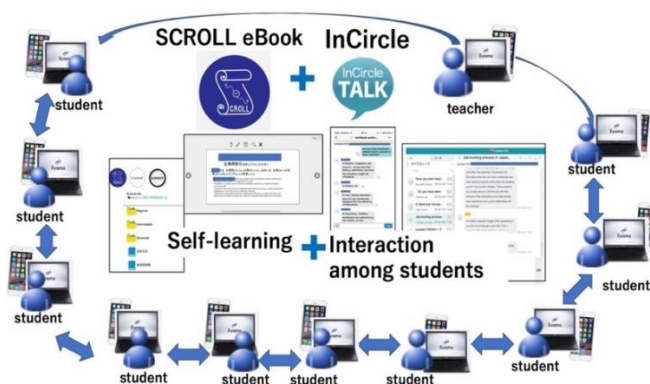


Figure 5. InCircle chat room interface on mobile

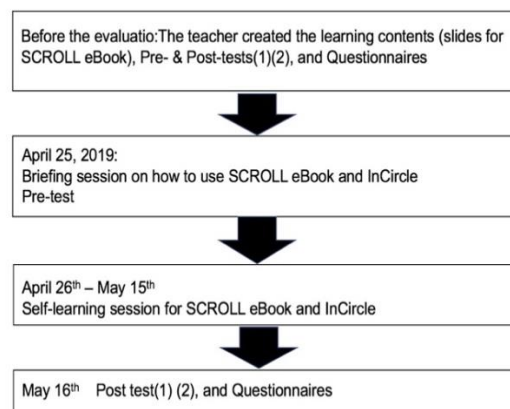


Figure 6. Evaluation Procedures

4. Evaluation

4.1 The target class

Nine graduate students (6 Japanese, 2 Chinese, 1 Brazilian), who were taking class called Career Design and Business Communication at the University in the western part of Japan participated in the evaluation experiments. The target class was held on a once-a-week-basis during the spring semester, 2019. Each student had a laptop PC in class. The objective of the target class was to develop one's self-concept during preparation for job hunting, and to develop the skills of problem finding and solving, and cross-cultural communication in a diversity of workplace environments. The evaluation was conducted during April 25th to May 16th as an out-of-class activity.

4.2 Procedures

Figure 6 shows the learning scenario. The teacher created 20 contents for SCROLL eBook of career related terms such as "分離礼(bunrirei)"(bow after words), "PREP法"(PREP method). The objective of the contents was to learn useful terms in terms of job-hunting. At the beginning of the session (Day 1), the participants received a briefing session to learn how to use SCROLL eBook and InCircle and took the pre-test to examine whether they know the meanings of 20 target terms. They were assigned to learn the target terms on a self-learning basis using SCROLL eBook. Then the teacher delivered 10 learning contents via InCircle. In order to examine the effectiveness of InCircle, the comparison was made between with and without InCircle delivery. In order to give an equal opportunity of education using the cutting-edge technology, there was no control group created. Therefore the whole class experienced both with and without InCircle delivery. During the evaluation session, students were free to use InCircle on PC to communicate with other classmates and the teacher. After the evaluation, Post-tests (1) & (2) were taken by the participants and the questionnaire was conducted.

4.3 Results

Table 1 shows the result of the Pre- and Post-test (1) and (2). Pre- and Post-test (1) were identical to ask them the meaning of 10 Japanese career-related terms to be taught via eBook with InCircle delivery. Pre- and Post-test (2) were also identical to ask them the meaning of 10 career-related terms to be taught via eBook without InCircle delivery. The full mark was 10 points for Pre- and Post-test (1) and Pre- and Post-test (2). The mean scores of the Pre-test (1) and (2) were 3.66 and 3.33 with the standard deviation (SD) of 1.12 and 1.22. The result of Post-test (1) was 9.55 with the standard deviation of 1.33, while that of Post-test (2) was 9.33 with the standard deviation of 2.00. T value shows that there is no statistically significant difference between them. As Figure 9 shows, there was no significant difference in the mean score increase in both medias.

Table 1

An The result of Pre- and Post-tests

	Pre-test (1) (full mark 10)	Post-test (1) after eBook and InCircle learning (full mark 10)	t-value of Pre-&Post test difference of with/without InCircle delivery
Mean	3.66	9.55	0.68 (p < 0.05)
SD	1.12	1.33	
	Pre-test (2) (full mark 10)	Post-test(1) after eBook learning without InCircle (full mark 10)	
Mean	3.33	9.33	
SD	1.22	2.00	

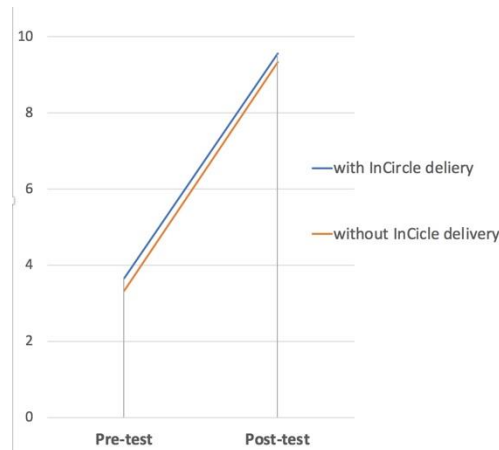


Figure 9. Comparison between eBook learning with InCircle and without InCircle in terms of the means of Pre- and Post-tests

5. Discussion and Conclusion

5.1 Discussion

At the end of the evaluation, they were asked to answer the five-point-scale-questionnaire as shown in Table 2. Q1 and Q6 were created to examine the fun factor of SCROLL eBook and InCircle. Q2 and Q5 were created based on the technology acceptance model proposed by Davis (1989). Q3 and Q7 were created to examine the user acceptance of its interface. The highest score, 4.88 was given when they were asked about the helpfulness of InCircle (Q.4). Even though there was no statistically significant difference between with and without InCircle delivery, they regarded it as helpful. It implies that InCircle contributed to effectiveness of learning job-hunting related terms. However the lowest score, 4.22 was given when they were asked about rating the interface of InCircle (Q.7). Since its interface is just like other chat tools, they seem to regard it as nothing new.

Table 2

The results of the 5-point-scale questionnaire

	Questions	Mean	SD
Q.1	Was it fun to learn career related terms with SCROLL eBook contents	4.56	0.53
Q.2	Was it easy for you to handle SCROLL eBook?	4.44	0.53
Q.3	Please rate its interface of SCROLL eBook	4.22	0.83
Q.4	You also learned career related terms via InCircle. Was it helpful?	4.88	0.35
Q.5	Was it easy for you to handle InCircle?	4.56	0.53
Q.6	Was it fun to learn career related terms via InCircle ?	4.75	0.46
Q.7	Please rate its interface of InCircle.	4.22	0.97

Table 3

The students' impressions of the eBook and InCircle combined learning

	Comments
#1	It was useful to learn career-related terms
#2	The contents were well-organized and easy to learn
#3	There were many abbreviated terms which I did not know
#4	Difficult
#5	Interface looks old. I prefer other free apps.
#6	It was difficult to adjust the slide size

Table 3 shows the participants' free comments on the eBook and InCircle combined learning. Comments #1, 2, 3 are positive ones, but they referred to the contents not mentioning the system itself.

As for Comment #4, it is not clear what was difficult in what way. Comments #5 and #6, which turned out that they referred to SCROLL, were rather negative. SCROLL was renewed right after this evaluation. The feedback will be expected better in our next evaluation. As the comment #6 pointed out, some students had some difficulty to enlarge the slide on smartphone, but it was solved after it renewal.

5.2 Conclusions

In this study, we describe facilitating the learning of job-hunting related terms using SROLLeBook and InCircle. Our research question was (1) whether or not the use of our chat system was effective in learning job-hunting related terms. Our hypotheses that InCircle content delivery system contributes to their learning job-hunting related terms could not be proved since there was no statistically significant difference between with and without InCircle delivery in terms of Pre- and Post-test improvement. It is among our future works to improve learning scenarios to prove the effectiveness of our system.

Acknowledgements

Part of this research work was supported by the Grant-in-Aid for Scientific Research No.18K02820, No.17K12947, No.16H06304 and 16H03078 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References

- AOS Mobile Inc., Tokyo, Japan: <http://www.aosmobile.com/> (last access: January 29th, 2019).
- Arame M., Naganuma S., Kobayashi M., Komatsu M. & Tamaki K.: Consideration about how to use the e Portfolio in career education (Japanese) *IPSJ Transactions on Computers and Education (Kenkyu Houkoku:Computer to Kyouiku)* 2013-CE-120(3), 1-8 (2013).
- Calitz, A., Evert, C., & Cullen, M.: Promoting ICT careers using a South African ICT career portal. *The African Journal of Information Systems*, 7(2), Article 1 (2015).
- Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13 (3), 319–339 (1989).
- Headlam-Wells J, Gosland J, & Craig J.: “There's magic in the web”:e-mentoring for women's career development. *Career Development International*, 10 (6/7), 444 - 459 (2005).
- InCircle: <https://www.incircle.jp/> (last access: January 29th, 2019).
- Inoue, G., Uosaki, N., Ogata, H., & Mouri, K.: Enhancing Outside-class Learning Us-ing Online Tools: A Review Work, *Proceedings of LTLE2014*, 332-337 (2014).
- Japan Student Services Organization (JASSO) https://www.jasso.go.jp/sp/about/statistics/intl_student_e/2018/index.html (last access: April 30th, 2019).
- Japan Revitalization Strategy 2016 https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/hombun1_160602_en.pdf (last access: April 30th, 2019).
- Ogata, H., & Uosaki, N.: A new trend of mobile and ubiquitous learning research: to-towards enhancing ubiquitous learning experiences. *International Journal of Mobile and Learning Organization*, 6(1), 64-78 (2012).
- Ogata, H., Hou, B., Li, M., Uosaki, N. Mouri, K. & Liu, S.: Ubiquitous learning project using life-logging technology in Japan, *Educational Technology and Society Journal*, 17(2), 85-100 (2014).
- SCROLL: <https://scroll.let.media.kyoto-u.ac.jp/learninglog/signin>(last access: May 14th, 2019).
- Teshima H., Kawasaki C., and Komatsu Y., Integrated Program of the Academic Skills and the Career Development for University Freshers : A Report on the Course "Skills for Self Establishing" at Osaka Jogakuin College (Japanese) *Bulletin of Osaka Jogakuin College* 5, 119-144, (2008).
- Uosaki, N., Mouri, K., Yin C., & Ogata, H. How We Can Support International Students' Job Hunting in Japan Seamlessly, *Proceedings of International Conference on Computer in Education, November 26-30, Philippines*, 527-529 (2018). <http://icce2018.ateneo.edu/wp-content/uploads/2018/12/C4-26.pdf>

The adoption of Facebook mobile application for managing learning

Edwin Theng Fung SIEW^a, & Cheng Ean (Catherine) LEE^{a*}

^a*Department of Communication, School of Arts, Sunway University, Malaysia*

*catherinelee@sunway.edu.my

Abstract: Mobile technologies such as mobile device applications play a significant role within formal education and the use of social networking site such as Facebook through mobile devices has influence students' management of learning in higher education. Researchers have identified a range of uses of Facebook mobile application in higher education. Despite the increasing use of Facebook mobile application by students and lecturers in higher education for educational purposes, very little empirical studies that concern the adoption of Facebook mobile application by university students in managing learning in the Malaysian higher education context. In this paper, we describe the likelihood of adoption of Facebook mobile application for managing learning by students of a private university in Malaysia, as well as how the perceived ease of use, perceived usefulness, and motivations of usage affect the behavioural intention of the students in adopting Facebook mobile application for managing learning. Through an online questionnaire survey, data were analysed using descriptive statistics, correlations and regressions. Results from the survey questionnaires (N=101) indicated a significant positive association between motivations of usage towards adoption of Facebook mobile application for managing learning as well as between perceived ease of use and perceived usefulness towards students' adoption of Facebook mobile application for managing learning. In conclusion, students' perceived ease of use, perceived usefulness, education purposes, information-seeking motive and interactivity were significant predictors of behavioural intention in adopting Facebook mobile application for managing learning.

Keywords: Facebook mobile application, managing learning, university students, Malaysia

1. Introduction

Mobile technologies are increasingly popular as useful educational tools to enrich teaching and learning process, to engage and retain students, and to improve educational outcomes (Arokiasamy, 2017). The benefits of mobile technologies for teaching and learning in higher education has caused a new paradigm of learning to emerge known as mobile learning. This new way of learning allows university students to utilise mobile technologies such as smartphone, tablets and laptops to learn and to obtain learning materials anywhere, anytime. The features of mobile technologies - mobility, the ease of software access and integration of media and applications - enable students 'to work more continuously across home and school settings, allows activities to be initiated outside the school, or practice on exercises to be undertaken when or where desired' (Passey, 2010, p. 69). Therefore, evidence from prior studies indicates the potential of mobile technologies in supporting learning (Passey, 2010).

Research on Facebook use in Malaysian educational contexts has been growing at a rapid rate. However, only ten out 35 existing studies are implemented in formal classroom curriculum. Despite the increasing use of Facebook mobile application by students and lecturers of higher education institutions for educational purposes due to its affordances and benefits illustrated in prior literature; very little empirical studies that concern the adoption of Facebook mobile application by university students for managing learning in the Malaysian higher education context. This study aims to find out the likelihood of adoption of Facebook mobile application for managing learning by students of a private university in Malaysia, as well as how the perceived ease of use, perceived usefulness and three motivations of usage (education purposes, information seeking and interactivity) affect the behavioural intention of the students in adopting Facebook mobile application for managing learning.

2. Methodology

The questionnaire consists of two sections, namely Section A for demographic profile and Section B for testing the independent variables (i.e., construct of usage motivations) and dependent variables (i.e., behavioural intention of adoption). During data analysis, a total of 101 responses were usable and analysed using IBM SPSS 25.0 for descriptive statistics, correlation and regression analyses. The reliability score of 32 scale-items using Cronbach's Alpha are found to be highly reliable, achieved 0.958 which sufficiently exceeded the minimum requirement of 0.70 (Nunnally, 1978). The respondents of the survey comprise of 101 students of a private university in the Klang Valley, Malaysia. This private university is known to have a focus on innovative uses of technology in teaching and learning. The demographic profile of the respondents and the descriptive statistics show that most of the respondents are female students (67.3%) and 32.7% of the respondents are male students. A total of 54 respondents are in the age range of 20-22 years old (53.5%); 23.8% of the respondents are within the age range of 17-19 years old, followed by 22.8% of them in the age range of 23-25. Majority of the respondents are Chinese (65.3%), followed by Malay and Indian students (12.9%) as well as other races (8.9%).

3. Findings

3.1 *The Adoption of Facebook Mobile Application for Managing Learning*

This section reveals the results answering the research question on the likelihood of adoption of Facebook mobile application for managing learning from students' level of agreement on 15 statements of three motivations of usage - education, information seeking and interactivity - towards the adoption of Facebook mobile application for managing learning in Malaysian higher education. Findings show that students slightly agreed that Facebook mobile application could be used for providing others with information ($M=3.90$), allows the exchange of information among peers ($M=3.83$), for knowledge-sharing in the education system ($M=3.79$), supports informative discussion related to education between friends ($M=3.71$), allows the exchange of information with educator ($M=3.55$), and gives the opportunity for discussion to take place ($M=3.53$).

Next, from the bivariate correlation findings, respondents' adoption of Facebook mobile application indicated a strong correlation with the opportunity for discussion to take place ($r=.563$, $p<.01$) and for obtaining information and course related materials ($r=.516$, $p<.01$); a moderate correlation with interaction among peers ($r=.492$, $p<.01$), for educational purposes ($r=.478$, $p<.01$), for interaction with the educator ($r=.452$, $p<.01$) and exchange of information among peers ($r=.442$, $p<.01$); but shows a weak association with providing others with information ($r=.279$, $p<.01$) and for knowledge-sharing in the education system ($r=.273$, $p<.01$). Based on the results, because $p<.01$, the relationship between education purposes, information-seeking motive and interactivity with Facebook mobile application adoption are statistically significant and they are positively correlated.

3.2 *The Relationship Between Perceived Ease of Use, Perceived Usefulness and Motivations of Usage towards Facebook Mobile Application Adoption for Managing Learning*

This section describes the relationship between perceived ease of use and perceived usefulness and behavioural intention of adopting Facebook mobile application for managing learning, as well as the relationship between education purposes, information-seeking motive and interactivity and behavioural intention of adoption Facebook mobile application for managing learning. A multiple linear regression was carried out to predict students' adoption of Facebook mobile application for managing learning based on perceived ease of use, perceived usefulness, and three motivations of usage (education purposes, information seeking and interactivity). The R values for the perceived ease of use and perceived usefulness as well as motivations of usage are 0.774 and 0.797 respectively, which indicate a high degree of correlation with the students' adoption of Facebook mobile application.

Besides, a significant regression equation was found ($F(15, 85)=8.445$, $p<.000$), with an R^2 of .598. This shows that the respondents' perceived ease of use and perceived usefulness of Facebook mobile application are accounted for 59.8% of variance towards the adoption of Facebook mobile application for managing learning. In another analysis, a significant regression equation was found

($F(15, 85)=9.884, p<.000$), with an R^2 of .636. This indicates that respondents' perception on the motivations of usage is accounted for 63.6% of variance towards the adoption of Facebook mobile application for managing learning. The findings show that all independent variables (perceived ease of use, perceived usefulness, education purposes, information-seeking motive and interactivity) were significant predictors of behavioural intention of the students in adopting Facebook mobile application for managing learning.

4. Discussion and Conclusion

This study investigated the likelihood of adoption of Facebook mobile application for managing learning by students of a private university in Malaysia. To answer the question on the likelihood of adoption, 30.7% of the respondents agreed and strongly agreed that they intend to continue using Facebook mobile application to manage learning. The motivations to adopt Facebook mobile application were due to knowledge-sharing, supporting informative discussions related to education between friends, providing others with information and the exchanging of information among peers. The results are consistent with the studies of Hamat, Embi and Hassan (2012), Al-Rahmi and Othman (2013), Chen (2015), Balakrishnan (2016), and Lee (2019) that the benefits of Facebook mobile are for sharing of information and knowledge, supporting educational communication between students and lecturers as well as discussing academic matters. This study concludes that students' perceived ease of use, perceived usefulness, education purposes, information-seeking motive and interactivity were statistically significant, positively correlated and significant predictors of behavioural intention of adopting Facebook mobile application for managing learning.

The study has practical implications for scholars, educators, students and policy-makers in Malaysia who are interested in the use of Facebook mobile application for higher education. The importance of this study is to offer Malaysian higher education a new perspective on the role of Facebook mobile for educational purposes, a new framework of practice within Malaysian higher education to support optimum use of Facebook mobile in formal classroom education, as well as to show evidence that Facebook mobile has a major role in supporting the management of learning activity. Nevertheless, the two limitations of this study are that its focus is mainly on students of one private university in Malaysia and the predictor of Facebook mobile adoption is limited to five variables - perceived ease of use, perceived usefulness, education purposes, information-seeking motive and interactivity. Similar studies could be conducted with students of other universities in Malaysia for generalisation to the Malaysian student population as well as to examine other predictors of behavioural intention for adopting Facebook mobile application in managing learning.

References

- Al-Rahmi, W.M., & Othman, M.S. (2013). The impact of social media use on academic performance among university students: A pilot study. *Journal of Information Systems Research and Innovation*, 1–10
- Arokiasamy, A.R.A. (2017). A systematic review approach of mobile technology adoption in higher education. *Economics, Management and Sustainability*, 2(2), 48-55. doi:10.14254/jems.2017.2-2.5
- Balakrishnan, V. (2016). Learning via a social media enabled tool – Do learning styles make a difference? *International Journal of Information and Education Technology*, 6(1), 35–38. doi:10.7763/IJET.2016.V6.654
- Chen, Y. (2015). Linking learning styles and learning on mobile Facebook. *International Review of Research in Open and Distributed Learning*, 16(2), 94-114
- Hamat, A., Embi, M.A., & Hassan, H.A. (2012). The use of social networking sites among Malaysian university students. *International Education Studies*, 5(3), 56–66. doi:10.5539/ies.v5n3p56
- Lee, C.E. (2019). Facebook use in Malaysian higher education classroom. In Tatnall A. (Eds) *Encyclopedia of Education and Information Technologies*. Springer, Cham. doi:10.1007/978-3-319-60013-0
- Nunnally, J.C. (1978). *Psychometric theory (2nd Ed.)*. New York: McGraw-Hill.
- Passey, D. (2010). Mobile learning in school contexts: Can teachers alone make it happen? *IEEE Transactions on Learning Technologies*, 3(1), 68-81

Characteristics Analysis for Technology Enhanced Learning Maturity: A Qualitative Approach

Amalia RAHMAH^{a*}, Harry B. SANTOSO^b & Zainal A. HASIBUAN^c

Faculty of Computer Science, Universitas Indonesia, Indonesia

amaliarahmah2@gmail.com

Abstract: IT advancement allowed us to experience technology enhanced learning (TEL). However, a learner should implement TEL gradually, following a maturity roadmap, by paying attention to TEL characteristics. TEL covers an extensive discussion embedded in hundreds of research articles. Thus, this research explores the TEL key terms from 12 research digital libraries using the Significant Words approach, postulated by Luhn. The 65 key terms were then analyzed to conceive TEL characteristics. There are nine characteristics for TEL maturity, discussing TEL coverage in facilitating formal-informal learning and various alternatives of learning methods, applying state-of-art technology along with social media and learning management system, and also improving the learning experience and the learning environment to enable self-directed learning. This result may become a foundation to formulate TEL maturity measures.

Keywords: Luhn significant words, maturity measurement, technology enhanced learning, TEL characteristics, TEL key terms

1. Introduction

The shifting of the learning paradigm follows the supporting technology advancement. However, digital discrepancy and learner's capability to utilize the technology are topics of concern. Thus, we need to comprehend ways of using state-of-art technology gradually, which is described as technology enhanced learning maturity (Rahmah, Santoso, & Hasibuan, 2017). Learners should understand the concerning characteristics in implementing TEL with its progression, to comprehend the maturity of TEL implementation. Maturity roadmap follows the shifting paradigm with ubiquitous and seamless learning as the highest level (Ogata & Uosaki, 2012) (Milrad et al., 2013),

This research tries to identify these characteristics by exploiting the key terms of TEL, using the significant words approach (Luhn, 1958) from a massive number of research articles on TEL. Introduced in text analysis research, this method explains that medium frequency words become significant words due to their discriminating power, meanwhile common and rare words are insignificant (Luhn, 1958). The significant words are the ones located in between the upper-cut and lower cut in the Zipf law distribution. The interesting poin is there is no defined principle on how to choose the lower and upper-cut. Van Rijsbergen (1979) stated that it requires a certain amount of arbitrariness. The resulting key terms are not yet in the form of characteristics. We determine the meaning and context using vocabulary control theory to formulate the TEL maturity characteristics.

2. Methodology

The methodology consists of data collection, pre-processing, processing, extracting significant words, and analyzing TEL characteristics. At first, we collect, filter, and choose full-accessed 548 research articles based on the search query "technology enhanced learning" found on the article's title, abstract, or keywords. The contributions are limited to proceedings, journals, and book chapters articles, published after 2010 in 12 accessible research digital libraries. Then, we run data cleaning by eliminating the article's metadata, punctuation, non-alphabetical words, one-letter words, and

stop-words. We also run the lemmatization procedure, parse the words to generate bigram and trigram, and continue with terms frequency of occurrence counting and descending sorting. Bigram and trigram are the chosen format because they can deliver meaning, without having to put in a context.

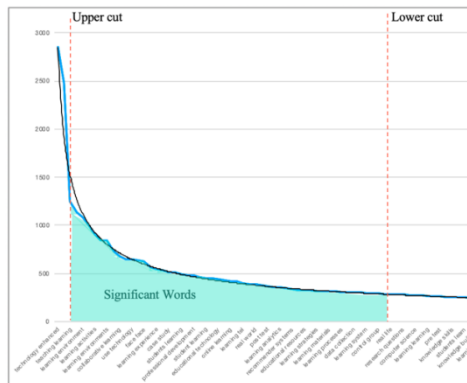


Figure 30. Luhn Distribution

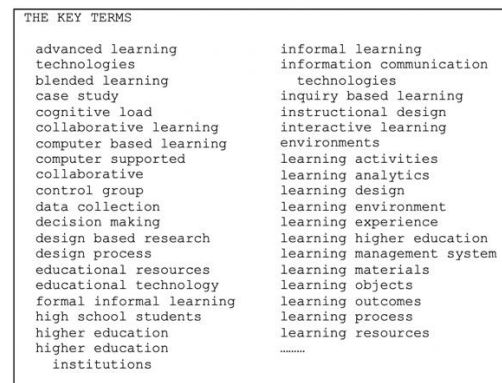


Figure 31. The Resulting Key Terms

In the following activity, we map the terms into Zipf distribution, delivered in Figure 30. The upper and lower cuts are chosen with prudence. The upper-cut lies on the slope with a significant jump, meanwhile the lower-cut lies before the slope become insignificant. We follow the principle that the number of to-be listed significant words are not too few or too many. It is too many if there are many generic terms with no discriminating power. It is too narrow, for example, less than 20 are not adequate to cover the scope of discussion. We then check for the resulting significant words for their eligibility, of which each term has meaning and relevance to the topic.

3. Result and Discussion

There are 65 important terms, which consist of 29 bigrams, and 36 trigrams (as shown in Figure 31). We formulate the characteristics using facet analysis and thesaurus to structure the filtered terms, comprehend their meaning, and acquire the underlying knowledge. Figure 6 delivers the sample data presentation¹. Then, we formulate the TEL characteristics for maturity measurement candidates.

FACETED KEY TERMS	THESAURUS FOR KEY TERMS
LEGEND (R) significant words (C) conceptualization result	LEGEND BT: broader term NT: narrower term RT: related term USE: use for
TECHNOLOGY (R) information communication technologies (ICT) (C) Technology trend (R) mobile devices (R) web based (R) social media	Mobile Learning BT: Online Learning RT: Educational Technology NT: Mobile Seamless Learning System USE: M-learning
LEARNING (C) Learning situation (R) informal learning (R) formal informal learning	

Figure 3. Example of Faceted Key Terms and Thesaurus

The first characteristic (C1) is about the learning situation, formal and informal. It is based on facet analysis on terms: formal, informal, and both. We define the second characteristic (C2) discussing to what extent technology facilitate online learning, starting from simple web-based learning to full-online learning (without face to face). It is based on the trait in supporting learning with online resources. The key terms are blended learning and web-based learning.

The third(C3) discusses how sophisticated technological trends are applied to support learning. The key terms discuss web technology for learning and mobile learning. Both are associated with the learning shifting paradigm. However, ubiquitous and seamless learning have not yet entered the

¹ Complete list can be accessed in <http://bit.ly/tel-keyterms>

significant words list. However, we can interpret it as a prospect to pursue. The other key terms are about social media, interpreted as to what extent social media can improve the learning process (C4).

Characteristics (C5) and (C6) take learning methods and drivers as a viewpoint. C5 covers the key terms consisting of collaborative learning, inquiry-based learning, problem-based learning, project-based learning, and self-directed learning. The last one delivers learner viewpoint. Thus we determine characteristic C6 on to what extent self-directed learning and its affecting traits could be improved TEL.

The seventh characteristic relates to what extent the learning environment configured to improve the learning experience (C7). It is also a concern how the technology could enhance learning environments, in terms of facilitating real-time, real-world context, virtual learning environment, and interactive learning environment (C9). Additionally, there are key terms expressed learning environment in higher education and learning systems, including learning management systems (LMS), open educational resources (OER), and massive open online learning (MOOC). Hence, to what extent the learning system is applied in high education becomes a consideration (C8).

4. Conclusion

This study has extracted key terms using Luhn's Significant Words and conceived the nine TEL characteristics' by analyzing these terms using facet analysis, and thesaurus concept. It then becomes TEL maturity measures candidates. This study attempted to explore characteristics of a concept using a cross-discipline technique, usually used in natural language research. The proposition is that key terms have discriminant power to explain a concept. Therefore, the result may not always suit our common sense. As an example, particular trending technology did not emerge as significant words.

Future research may study this finding in greater depth to formulate characteristics with the TEL maturity leveling systems. The study may continue using other natural language techniques and data interpretation to strengthen the discovery. We can also examine related works which deduce characteristics and affecting factors of TEL maturity. Further, the study may also continue by putting a context to the terms, such as sentiment, to explain how a concept affects TEL maturity.

Acknowledgements

This research was supported by Hibah Publikasi Internasional Terindeks 9 (PIT 9) 2019 at the Universitas Indonesia (NKB-0008/UN2.R3.1/HKP.05.00/2019) and Doctoral Program of Computer Science funded by Indonesia Endowment for Education.

References

- Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159–165. <https://doi.org/10.1147/rd.22.0159>
- Milrad, M., Wong, L.-H., Sharples, M., Hwang, G.-J., Looi, C.-K., & Ogata, H. (2013). Seamless Learning: An International Perspective on Next Generation Technology Enhanced Learning. In Z. L. Berge & L. Y. Muilenburg (Series Ed.), *Handbook of Mobile Learning* (pp. 95–108). New York: Routledge.
- Ogata, H., & Uosaki, N. (2012). A New Trend of Mobile and Ubiquitous Learning Research: Towards Enhancing Ubiquitous Learning Experiences. *International Journal of Mobile Learning and Organisation*, 6(1), 64–78. <https://doi.org/10.1504/IJMLO.2012.046883>
- Rahmah, A., Santoso, H. B., & Hasibuan, Z. A. (2017). E-Learning Process Maturity Level: A Conceptual Framework. *Journal of Physics: Conference Series*, 978, 012028. <https://doi.org/10.1088/1742-6596/978/1/012028>
- Van Rijsbergen, C. J. (1979). Automatic Text Analysis. In *Information Retrieval* (2nd ed.). Retrieved from <http://www.dcs.gla.ac.uk/Keith/Chapter.2/Ch.2.html>

Integration of cloud-based mobile learning to improve students' creative thinking in a visual arts course

Jing-Wei LI ^a, Iwen HUANG ^{a*} & Hui-Yu JIANG ^a

^a *Department of Information and Learning Technology, National University of Tainan, Taiwan*

*huangi@mail.nutn.edu.tw

Abstract: Recently visual arts education has been influenced by technology development and delighted research attention. This study focuses on cloud-base mobile learning in visual art teaching, to improve the ability of creativity for high school students. The findings showed that student performance on scales of fluency, and resistance to premature closure in Torrance Creative Thinking test have been improved.

Keywords: G Suite for Education · Creative Thinking · Mobile Learning · Visual Arts

1. Introduction

Arts courses have the potential for fostering creativity in schools. The process of creating arts needs to be simulated by a lot of visual materials such as images and videos (Xu, 2019). However, it is difficult to prepare teaching materials and present artworks during course time (Ulger, 2016). The integration of visual arts and digital technology is one of the trends in recent years (Ceng, & Chang, 2013) . By using digital resources for visual arts teaching can solve these problems and stimulate students' creative thinking by multimedia (Chang, Chen, Yu, Chu & Chien, 2017).

Integration of cloud-based teaching resources into the classroom, provide students a flexible way in mobile environment (Clmrompton, Burke, & Gregory, 2017). The G Suite for Education is a combination of online tools and available at all levels of schools. Teachers and students can share unlimited cloud space and Internet resources to make learning contents more valuable (Iftakhar, 2016). They can also use the Internet platform to interact with each other. The National Curriculum for England encourages teachers use Internet resource to improve art and design techniques to help students learn visual arts, and develop creativity and imagination.(UK National Curriculum Online, 2010). Through active participation in art appreciation, students could develop new ways to enhance their power of creative thinking and presentation skills (Pateman, 2016).

The purpose of this study was to integrate cloud-based mobile learning in high school visual arts course to promote students' creativity. The main research question was that did students who learn with cloud-based mobile learning have better creative thinking than students who learn with traditional learning?

2. Methodology

2.1 Research Design

The 12-week experimental course had conducted for grad 10 students in the south of Taiwan. There were 40 participants in the experimental group and 35 in the control group. The experimental group used cloud resources with mobile learning. The control group used paper handouts and the textbook. G Suite for Education was used in experiment group that was cloud-based enterprise communications tools. Teachers integrated learning resources on the Internet to provide arts gallery and multimedia. The students in the experimental group could connect to G Suite for Education at anytime and anywhere through the mobile device. The control group learned from textbook and handouts.

The whole teaching experiment process was divided into four stages, namely, “Creative design”, “Art skills accumulation”, “Creation thinking practice” and “Creative achievement”.

1. The creative design-" assemblage graphic": The element of art was modeling. The teaching objective was to familiarize students with graphic recombination techniques
2. Art skills accumulation-"perspective drawings": The elements of art were color, texture material, space. The teaching objective was to build the concept and experience of arts techniques.
3. Creative thinking practice-" association": The element of art was form. The teaching objective was to familiarize students’ creative techniques
4. Creative achievement-" Creating reality ": The elements of art were color, texture, and space. The teaching objective was applying creative techniques and complete a creative arts work.

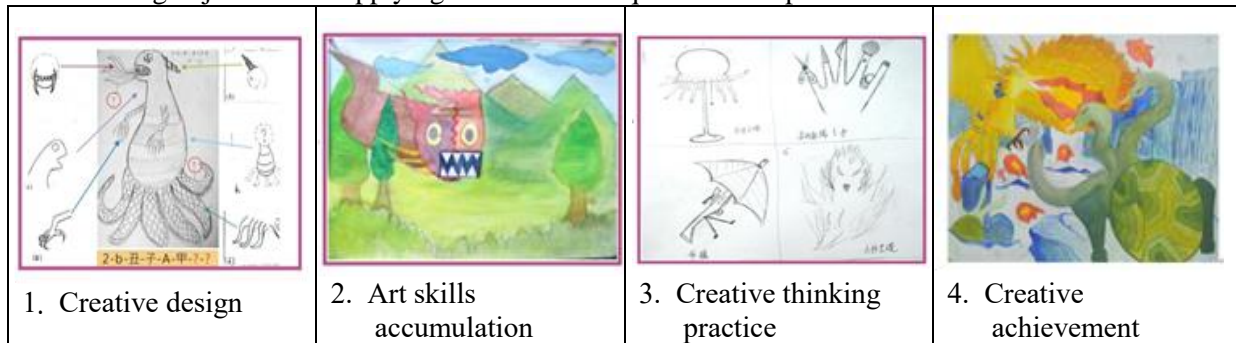


Figure 1. The arts work of four stages.

2.2 Research Instrument

The Torrance Tests of Creative Thinking (TTCT) is a test of creativity that reflected a person’s attitude to creativity, problem solving and decision making (Torrance, 1974). The figural version of the test was used in this study that consists of activities to drawing (Tan, 2007).

3. Research Results and Discussion

The pre-test of the TTCT, the average score of the experimental group was 84.17 points, and the control group was 82.12 points, the overall result of *t*-test showed no significant difference ($t= 0.78$, $p= .43 > .05$). The mean of the experimental group and the control group are 91.06 and 83.15, respectively; moreover, the post-test scores of the two groups reached a significant difference with $t=3.36$ ($p= .00 < .05$). There were significant differences in scale of fluency ($t= 4.46$, $p= .00 < .05$); and resistance to premature closure ($t= 3.47$, $p= .00 < .05$).

Table 1

Independent sample t – test of the post-test for Experimental group and control group

Scale	Group	Mean	SD	<i>t</i>	<i>p</i>
Fluency	Experimental	122.35	18.18	4.46**	.00
	Control	105.03	15.00		
Creative strengths	Experimental	116.00	23.97	1.74	.07
	Control	108.06	13.16		
Elaboration	Experimental	64.20	8.91	-.75	.44
	Control	65.49	5.19		
Abstractness of titles	Experimental	61.60	25.68	.83	.40
	Control	57.31	17.42		
Resistance to premature closure	Experimental	91.13	14.94	3.48**	.00
	Control	79.86	12.84		
Overall	Experimental	91.06	11.67	3.36**	.00
	Control	83.15	8.08		

The results of this study show that the results are similar to the research that picture books could improve the creativity and have significant difference in scales of "Fluency" (Yang, & Chang, 2013). By using digital resources, the students of the experiment group get obvious effects which is consistent with the previous study (Wang & Zhang, 2014).

4. Conclusions and Suggestions

The results show that integration of cloud-based mobile learning in arts course could improve creative thinking. Therefore, teachers can design mobile learning experiences and create a paperless classroom by integrating digital resources to enhance students' creative ability in the arts courses. The G Suite is a cloud-based productivity tool that helps teacher organize the digital resources and communicate with students. To evaluate the effects on creativity, 12 weeks might be too short to get significant results (Wakil, Omer, & Omer, 2017). Further research could extend research periods to get better results

References

- Ceng, B. W., & Chang, S. H. (2013) Research on the Inference Ability and Creativity of Intel K-12 Image Thinking Assistant into the Teaching Plan , *Special Education* , (14) , 1-19 .
- Chang, Y. S., Chen, S. Y., Yu, K. C., Chu, Y. H., & Chien, Y. H. (2017). Effects of cloud-based m-learning on student creative performance in engineering design. *British Journal of Educational Technology*, 48(1), 101-112.
- Clmrompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, 110, 51-63.
- Iftakhar, S. (2016). Google classroom: what works and how? *Journal of Education and Social Sciences*,3, 12-18.
- Pateman, T. (2016). *Key concepts: A guide to aesthetics, criticism and the arts in education*. Routledge.
- Tan, A. G. (2007). *Creativity: A handbook for teachers*. World Scientific.
- Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking (Norma-technical Manual)*. Bensenville, IL: Scholastic Testing Service
- UK National Curriculum Online. (2010). *Introducing the new primary curriculum*.
- Ulger, K. (2016). The creative training in the visual arts education. *Thinking Skills and Creativity*, 19, 73-87.
- Wakil, K., Omer, S., & Omer, B. (2017). Impact of computer games on students' GPA. *European Journal of Education Studies*.
- Wang, D. M. (2003) . The impact of Internet technology on the development of digital art. *Digital Art Education Journal* , (4) , 1-18 .
- Wang, Z. Q., & Chang, S. H. (2014) . Creative Thinking Skills Integrated into Game Design Program on Creativity for the Elementary Gifted Students , *Special Education* , (16) , 62-78 .
- Xu, Y. (2019). Digital Inheriting Application of Traditional Folk Art in the College Art Teaching. *Argos*, 36(74).
- Yang, Y. T. C., & Chang, C. H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education*, 68, 334-344.

MAIN CONFERENCE C5

EDUCATIONAL GAMIFICATION AND GAME-BASED LEARNING (EGG)

- FULL PAPER -

- EFFECTIVENESS OF USING AN IMMERSIVE AND INTERACTIVE VIRTUAL REALITY LEARNING ENVIRONMENT TO EMPOWER STUDENTS IN STRENGTHENING EMPATHY AND MASTERY LEARNING..... 495**
RHODORA ABADIA, JAMES CALVERT, RATNA DASIKA
- EFFECTS OF GAME-BASED LEARNING ON INFORMAL HISTORICAL LEARNING: A LEARNING ANALYTICS APPROACH 505**
XUANQI FENG & MASANORI YAMADA
- EXAMINING THE EFFECT OF GAMIFICATION IN INFORMATION SCIENCE, COMPUTER AND ENGINEERING EDUCATION: A META-ANALYSIS OF STUDENT LEARNING PERFORMANCE 515**
SHURUI BAI, KHE FOON HEW & BIYUN HUANG

- SHORT PAPER -

- DRUG DEFENSE: A MOBILE GAME FOR PREVENTION OF ALCOHOL ABUSE 525**
DIVINE-KIA T. TAN, RODRIGO EMMANUEL ROY, JOSE RAFAEL LASALA, ANTON GABRIEL YAP, MA. MERCEDES T. RODRIGO, WALFRIDO DAVID DIY & MA. REGINA M. HECHANOVA
- THE EFFECTS OF VIRTUAL REALITY SYSTEM APPLIED TO SHOOTING TRAINING COURSE FOR SENIOR HIGH SCHOOL STUDENTS..... 531**
YI-CHENG LIAO, HAO-CHIANG KOONG LIN, PO-CHUN KUO & KUAN-YU LIN
- ANALYSIS OF STUDENT AFFECT AND BEHAVIOR WHILE PLAYING A MOBILE GAME FOR ENGLISH COMPREHENSION 537**
MA. MERCEDES T. RODRIGO, JENILYN L. AGAPITO, DOMINIQUE MARIE ANTOINETTE B. MANAHAN
- CORE GAMIFICATION OF LEARNING ACTIVITIES THROUGH A METHOD BASED ON INFORMATION STRUCTURE MANIPULATION 543**
PEDRO GABRIEL FONTELES FURTADO, TSUKASA HIRASHIMA & YUSUKE HAYASHI
- A USABILITY STUDY ON SHAPE SHAPE HOORAY: AN ADAPTIVE EDUCATIONAL GAME ASSOCIATING 3D GEOMETRIC SHAPES TO DAILY OBJECTS..... 549**
WINNA MIA VICTORIA D. BUENVIAJE, MA. ANNIELA B. DELA CRUZ & INGRID MARIE THERESE P. FADRIQUELA
- INTEGRATING COMPUTATIONAL THINKING WITH DIGITAL STORYTELLING TO ENHANCING EXPRESSION ABILITY 555**
CHIH-KAI CHANG & XIN-YU DAI
- A MIND-MAPPING GUIDE BASED ON UNIFIED MODELING LANGUAGE FOR DEVELOPING EDUCATIONAL ROLE-PLAYING GAMES..... 561**
AHMED TLILI, MOUNA DENDEN, FATHI ESSELMY, MOHAMED JEMNI, KINSHUK, NIAN-SHING CHEN, MAIGA CHANG & RONGHUAI HUANG
- DANCING A TREASURE: A VIDEOGAME TO MOTIVATE YOUNG AUDIENCES TOWARD SPANISH DANCE..... 567**
ALEJANDRO ROMERO-HERNANDEZ, MANUEL GONZALEZ-RIOJO, MERIEM EL-YAMRI, BORJA MANERO
- PROTOTYPING OF AN EARTHQUAKE EVACUATION LEARNING GAME WITH VR REPRODUCING THE ENVIRONMENT THAT IS FAMILIAR TO LEARNERS..... 573**
RAIYA YAMAMOTO, ASAKA KITAOKA & KEI INOUE

RESEARCH ON APPLICATION OF STEAM TEACHING IN PRIMARY SCHOOL BASED ON GAMIFICATION TASK ORIENTATION-TAKING THE FIFTH GRADE CAMPUS SANDBOX AS AN EXAMPLE	579
XUE-FEI DING, DONG-MING QIAN & MENG-NI FENG	
INNERVATE AR: MOBILE AUGMENTED REALITY FOR STUDYING MOTOR NERVE DEFICITS IN ANATOMY EDUCATION	585
MARGARET COOK, AUSTIN PAYNE, JINSIL HWARYOUNG SEO, MICHELLE PINE & TIMOTHY MCLAUGHLIN	
ENHANCING COMPUTER ASSEMBLY SKILL USING VIRTUAL COMPUTER LABORATORY IN CONJUNCTION WITH TEAM GAME TOURNAMENT METHOD.....	591
ACHARAPORN KHANTHUSAENG & NADH DITCHAROEN	
- POSTER PAPER -	
INSTRUCTIONAL DESIGN OF STEAM EDUCATION BASED ON VIRTUAL REALITY TECHNOLOGY —— TAKING LEGO BRICKS AS AN EXAMPLE.....	597
FANGYUAN TONG, JIAHUI RUAN ,LINGLING MENG	
CODAR: AN AUGMENTED REALITY BASED GAME TO TEACH PROGRAMMING	600
VANDIT SHARMA, JEEVANKUR TALUKDAR & KAUSHAL KUMAR BHAGAT	
ENGLISH LANGUAGE LEARNERS’ PERCEPTIONS AND EXPERIENCES OF AN ENGLISH MOBILE URBAN GAME.....	603
YU-FENG (DIANA) YANG	

Effectiveness of using an Immersive and Interactive Virtual Reality Learning Environment to Empower Students in Strengthening Empathy and Mastery Learning

Rhodora ABADIA, James CALVERT, Ratna DASIKA

University of South Australia
Torrens University Australia
Torrens University Australia

*rhoda.abadia@unisa.edu.au

Abstract: In this paper we present the significance of immersive virtual reality learning environment by a conducted research which emphasizes experience in learning. The case study presents the analysis and comparison of the results in students' empathy and mastery learning between using a 360° video of a modern history topic – the World War II Kokoda campaign, against an immersive and interactive virtual world learning environment (Kokoda VR). Kokoda VR was developed to have an accurate recreation of places and events. By using photogrammetry of real locations and artefacts, in combination with animated characters, Kokoda VR places the students in the centre of the action. While each of the design techniques are not new to the development Virtual Reality Learning Environments (VRLEs), the combination and then application, created an immersive experience where students not only witness the event but also relive the experience. Participants in this study are high school and university students, who were divided into groups to use the 360° video and immersive virtual world application. Surveys and post-tests were conducted to identify how significant the enhancement is in the learners' empathy, mastery learning and their relationship. Results show that using immersive and interactive virtual reality learning environment have significant improvement in empathy and mastery learning. The study also shows that there is a positive association between empathy and mastery learning but the results does not show sufficient evidence of a strong relationship.

Keywords: immersive virtual reality learning environment, empathy, mastery learning, photogrammetry

1. Introduction

The capacity of Virtual Reality (VR) to provide meaningful enrichments for educational purposes is gaining significant interests and the possibilities for immersive learning using VR are endless and evolving. Although virtual reality has existed for decades now in different forms, the high cost of using this technology inhibits it from being used in education. With the emergence of affordable head mounted displays (Southgate & Smith, 2017) and the push for development of application for the classrooms by some of the major technology companies, such as Google (Google for Education, 2019) and Facebook's Oculus Education pilot program (Oculus VR, 2018), it is now making it possible to use immersive VRLEs in the classrooms. However, there is still a gap between claims for the usefulness of VR in academic learning and scientific research testing these claims.

The goal of the present study is twofold. The first part is to compare the instructional effectiveness of immersive VRLE in increasing empathy and mastery learning, compared to a 360° video. Empathy and mastery learning are two of the important salient features of VR applied in education (Abadia, Calvert & Tauseef, 2018). This research question has important practical implications concerning whether it is worthwhile to use immersive VRLEs in classrooms because

developing them is resource intensive compared to 360° videos. The second part is to analyze the relationship of increased empathy to student's mastery learning. For the purpose of this study, empathy is defined as "sensitivity to, and understanding of, the mental states of others" (Smith, 2006); and mastery learning, which was introduced by Bloom (Bloom, 1968), maintains that students must achieve a level of mastery (e.g., 80% score on a test) in pre-requisite knowledge before moving forward to learn subsequent information.

Several researches have shown that empathy assist students in having better academic achievement (Decety & Ickes, 2011; Bonner & Aspy, 1984; Mangione, Discepolo, Tore, Tore, Cozzarelli and Corona ,2013) and immersive VRs are effective tools in increasing empathy from its users (Shin, 2018; Shin & Biocca, 2017). The significance of this objective is to determine how the level of empathy is positively correlated to students' engagement in the learning environment that leads to mastery learning in the immersive VRLEs.

To serve this purpose, we created a VR experience for teaching history called Kokoda VR. Kokoda VR was designed and developed as a 40-minute educational experience to immerse Australian high school students in a significant moment in modern Australian history. By blending archival information such as photos, film and diary entries with virtual re-enactments, Kokoda VR provides the most complete portrait of a specific time and place. Using Kokoda VR and 360° videos, the research is conducted, and the results are analyzed and presented for the said above stated goal.

The paper is organized as follows: section 2 reviews previous work on VRLE's that afforded increase in empathy and learning. Section 3 discusses the development of Kokoda VR. Section 4 presents study design while section 5 discusses the results and analysis respectively. The paper is concluded in section 6 outlining the future work.

2. Related Literature

This section presents relevant studies of VRLEs that measures empathy and learning; and literatures that looks at the benefits of increased empathy to learning.

2.1 Empathy in VRLE

Empathy is an important instructional element that can increase students' interests (Rehan Dar, 2016). Preceding studies have demonstrated that VR can increase the empathic response felt by users, when compared to the same experience viewed in 2-dimensional presentation (Kandaurova and Lee, 2018; Schutte and Stilinović, 2017; Shin and Biocca,2017). However, these studies were not tested on a VRLE with an immersive narrative set in a realistic 3D environment. This paper presents the role of a VRLE with a linear narrative has on user empathy in an immersive learning environment.

2.2 Learning in VRLE

The following studies show that virtual reality has great potential in improving learning outcomes for education across different fields. Mastery learning were measured by studies in VR applications in the areas of space education (Bhargava et al, 2018), aviation (Chittaro and Buttussi, 2015), medical education (Bruno, Ongaro &Fraser, 2007), and construction (Lucas, 2018). These studies used pre and post-tests to measure learning. Butt, Kardong-Edgren, & Ellertson (2018), who also applied VR in clinical education, used a different approach to measure mastery learning by asking participants to answer two questions regarding the topic of the study. Kuronen-Stewart et al (2015) also applied the use of VR in clinical education. Their study used quantitative questionnaires to measure perceived effectiveness in learning and similarity to real-life setting. Experimental results by Alhalabi (2016) showed higher performance of the students in leaning the engineering education using immersive VR as a result of more engagement in the learning environment. The study (Akbulut, A., Catal, C., & Yildiz, B., 2018) particularly in computer engineering course using VR evidenced more successful student results compared to traditional teaching approaches.

2.3 Empathy and Learning

There is evidence from neuroscience, psychology and other educational research of the strong relationship of empathy in learning (Brennan, Daily, and Kaman, 2018; Cooper, 2011; Mangione, Discepolo, Tore, Tore, Cozzarelli, and Corona, 2013), academic performance and increase in both critical and creative thinking skills (Bonner and Aspy, 1984). There is no current study that looks at the relationship between empathy and learning when using VRLEs. In this study, the effect of immersive VRLEs in influencing higher empathy in the participants is compared if it results to better mastery learning.

3. The Development of Kokoda VR

Kokoda VR is a VRLE designed to immerse students in a significant moment in modern Australian history. Thousands of World War Two soldiers died on the Kokoda track in really hard conditions. The VR experience contextualizes and visualizes the Kokoda campaign, the intense conditions and the sacrifices made by many. Located in the jungle mountains of Papua New Guinea, it's hard to get to Kokoda and impossible to go back in time, but through VR, students can still attain a meaningful experience.

To fit in with the Australian curriculum and be a useful teaching resource, Kokoda VR needed to be historically accurate. However, the full experience needed to be condensed to be to a one class time of 40 minutes. Therefore, only key moments in the campaign could be described and linked together in a series of 12 scenarios, or chapters. Each chapter had to describe a significant historical moment, be engaging for the student and elicit a feeling of empathy for the soldiers. In addition, for Kokoda VR to act as a virtual field trip, high levels of realism and accuracy were required (see examples in Figure 1). This motivated the use of photogrammetry capture in the dense jungle and mountains of Papua New Guinea. Photogrammetry is the science of eliciting information about physical objects and environment by making measurements from photographic images. In this research, the focus was on the use of stereophotogrammetry which involves the estimation of 3D coordinates of an object by comparing multiple photographic images taken from different positions (Foster and Halbstein, 2014).

For each location, an area of approximately 10 meters in diameter was captured. A 360° photograph was taken from the central point of each location.



Figure 1. A sample screenshot from Kokoda VR (Calvert, Abadia & Tauseef, 2019)

In addition to the location, photogrammetry capture of museum artefacts from the Kokoda campaign was also performed. These included soldier uniforms, weapons and equipment from both sides in the conflict.

The models created from the photogrammetry capture were turned into a virtual set, approximately 10 meters in diameter, that the user can explore and where the key pieces of action take place. 360° photographs taken at each location were used to show the distant environment. To hide the seam between the photogrammetry stage and the 360-photograph, custom made 3D models and terrain

created using a variety of model building and photogrammetry techniques were added to the scene (see Figure 2).

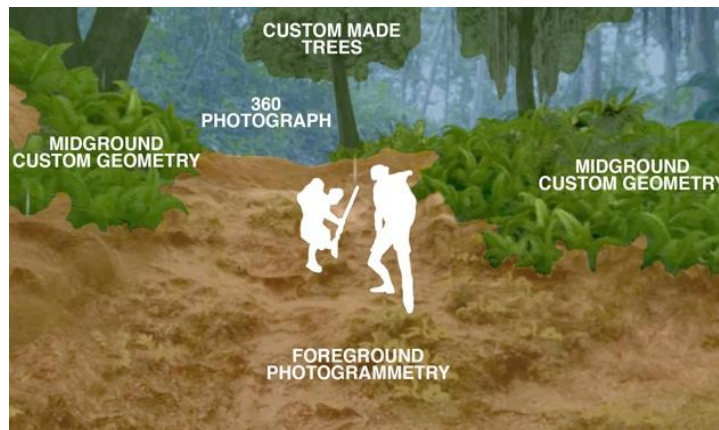


Figure 2. Screenshot from Kokoda VR with overlays showing how the scenes were constructed with a mixture of photogrammetry models and custom-made models (Calvert, Abadia & Tauseef, 2019)

To further fill out each scene, more than 40 real museum artefacts from World War Two captured from location or museums were placed in the scene. These photogrammetry locations and artefacts, motion graphics, voice over recording, location audio, musical score and character animation were combined to produce an immersive and interactive VR.

Integral to the story of Kokoda was the sense of hardship experienced by the Papuans, Australian and Japanese soldiers. Table 1. provides examples of how scenarios were designed in the Kokoda VR experience, with the goal of having the user feel empathy for the soldiers.

Table 1.
Example of Scenarios Designed to Elicit Empathy

Real scenario from the Kokoda campaign	Affordance based design manifestation	Desired empathic condition/response
It rained for most of the Kokoda campaign	Simulated rain falling around the user, with matched spatial audio	Environmental conditions were unpleasant for the soldiers
The Kokoda track is very steep	Place the user midway up a steep incline and search for a branch to use as a walking stick	The track was physically demanding
When vital supplies were dropped in from the air, they smashed on the ground	User sees planes approaching and releasing supplies, then see and hear (in 3D audio) the supplies smash on the ground	Soldiers already in harsh conditions weren't adequately supported
Soldiers often had to dig their own protective trenches, at night and without adequate equipment	User must shine a flashlight around a dark jungle environment to find an empty food tin to dig with	Soldiers had minimal equipment and often had to work in dark and harsh conditions

4. Study Design

The major participants in the study included high school and university students. The participants consist of high school students who will be studying the Kokoda history. For high school, two schools were chosen to participate, one public school and another private to cover different socio-economic background (analysis of this is not presented in this paper). For each school, one class was chosen by the

history teachers to use the Kokoda VR (experimental group) while the other group was chosen to use the 360° video (control group).

The study approach for high school students were replicated to university students. Two universities participated, one from Australia and another university from India. This is to identify if the empathy results are different from students who have no background knowledge of the historical content (students from India) compared to Australia university students with knowledge about Kokoda. This helps the study understand if the use VR increases the feeling of empathy regardless of their knowledge of the story presented.

All participants are self-selected. Students under 18 need permission from their parents. The parents and participants were informed about the study and they can voluntarily choose to participate or not. There was a total of 96 students who participated in the study (42 high school students and 54 university students).

4.1 Data Collection

All participants were asked to complete the student demographics survey before participating in the experiment. The experiment was conducted on the week before the Kokoda topic is introduced to high school students. All high school students who participated understand what the Kokoda campaign is about but have not yet have lessons about the details of this historical events. For university students, Australian participants have knowledge of the historical background while participants in India has none.

The students during the text were asked to have a 40-minute Kokoda VR or 360° video experience. Immediately after using Kokoda VR or 360° videos, participants completed a post-questionnaire which asked five questions relating to empathy (see Table 2 for an example survey question about empathy). The participants had to respond using a scale from 1 to 5. Two days after the students have completed the activity, the participants were asked to complete the online test to determine mastery learning.

Table 2

Sample Empathy Survey Question

Q1: I felt sad while seeing what the soldiers went through at that time
Q2: When the soldiers spoke about the conditions on the track, I could empathise with their struggle
Q3: I could understand the difficulties faced by both the Australian and Japanese soldiers.

Q4: I have a better understanding of the sacrifices made by soldiers on both sides of the conflict
Q5: Being a soldier in the Kokoda campaign would have been an exciting experience (scores to this question were reversed, to give an indication on student's willingness be a soldier in Kokoda)

Included in the data collection but not presented in this paper is the gathering of additional feedback form the students. Students were asked by their teachers to talk about their experience. They were also asked to take a quiz a month after the experiment to test and identify if there is an association between empathy and long-term recall; and if VRLEs are effective tools for long-term recall.

5. Results and Analysis

In all high schools in Australia, the Kokoda campaign is part of the history curriculum. In the data collection, high school students who participated have been given an overview of the Kokoda campaign before they participated in the experiment. The university participants, on the other hand, are a mixture of students who learned about the Kokoda campaign and those who have not heard of the Kokoda campaign. Before the experiment, the university students were only given a brief overview of what the Kokoda campaign is all about. The results and analysis of the experiment for empathy, mastery learning, and their correlation are presented below.

5.1 Empathy

Overall, there is a significant increase in empathy felt by participants in the VR version over 360°video (Figure 4). Participants in the Kokoda VR experienced 85% empathy (4.2 out 5 scale) while 360° participants experienced 76% empathy (3.8 out 5 scale). Figure 4 also shows that for both Kokoda VR and 360° groups of participants, university participants have higher empathy (mean =4.27) than those who are in high school (mean =3.81). The standard deviation is 0.2308 (calculated considering the mean empathy (360°video and Kokoda VR) of high school and university participants with reference to all participants mean empathy).

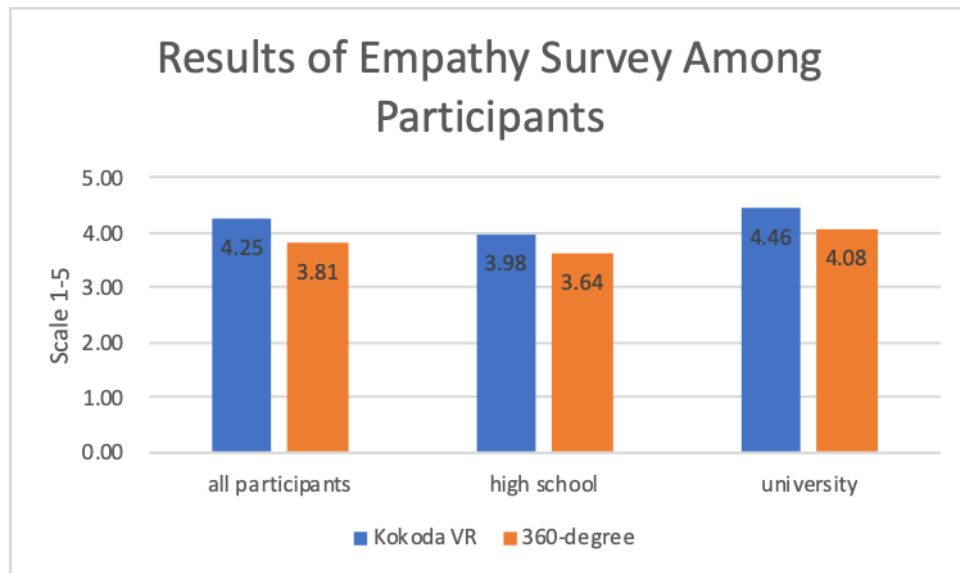


Figure 4. Graph showing the result of the empathy survey among participants.

However, our study shows that empathy is increased in VR over 360° video, but only when the specific scenario has been effective in the application of the technological affordances of 6-DOF head-tracking VR. Table 3 shows the responses to the empathy questions.

Table 3

Mean, Standard Deviation and t-Test scores for the empathy questions

Question	Kokoda VR		360° video		t-Test
	Mean	SD	Mean	SD	
Q1	4.43	0.75	4.05	0.93	0.1457
Q2	4.29	0.71	4.15	0.93	0.0080
Q3	4.55	0.59	4.26	0.90	0.0019
Q4	4.55	0.54	4.36	0.82	0.0004
Q5	3.16	1.46	2.87	1.06	0.1194

For questions 3 and 4, responses from the VR group are clustered more around ‘agree’ and ‘strongly agree’ compared to the 360° video group. The responses to questions 1 and 2 are evenly spread across both conditions. The responses to question 5 have a low empathy for both Kokoda VR and 360° groups.

When users hear soldiers speak of their discomfort in the jungle (Q2), or better understanding of the sacrifices made by soldiers on both sides of the conflict (Q3), there is no statistically significant increase in empathy in VR over 360° video. It is our view that the user can see the harsh conditions and arduous events the soldiers speak about in both conditions, 360° video and VR, but the user is not

experiencing the harsh conditions for themselves. Therefore, the full benefits of the immersive environment in VR are not being utilized.

The study also compared the empathy results of university participants who have learned the Kokoda campaign history in high school to those who have not heard of the Kokoda campaign until the introduction before the experiment. Results (see Table 4) show that participants with no background knowledge of Kokoda felt 7% more empathy using Kokoda VR compared to 360° video. This means these participants are more engaged in the immersive virtual learning environment that increased empathy. The same case with participants of Kokoda with background, having an empathy 5% higher than 360° video participants. In both cases, there is an increased empathy using Kokoda VR. Considering that if participants do not have background of Kokoda, they are not biased in their engagement. So, we can give weightage to the results of participants having no Kokoda background. With this we can conclude that using Kokoda VR has a higher level of empathic experience.

Table 4

Comparison of Level of Empathy Between University Participants – Background Knowledge vs No Background Knowledge

	Kokoda VR	360-degree
No Kokoda Background Knowledge	88%	81%
With Kokoda Background Knowledge	90%	85%

5.2 Mastery Learning

Table 5 summarizes the result of the tests given to the participants to measure mastery learning. Test scores show that VR participants in the study performed better than 360° video with an overall score of 85% compared to a 79% for 360° video participants. The high school participants performed better than university participants. The test results for high school participants have a 10% difference between VR and 360° video test results. VR participants' mean test score is 94% compared to 84% for 360° video participants.

Table 5

Test Results of All Participants (Score out of 10)

	All Participants (n=96)		High School (n=51)		University (n=45)	
	Mean (%)	SD	Mean (%)	SD	Mean (%)	SD
Kokoda VR	85	0.94	94	0.74	80	0.85
360° video	79	1.7	84	1.7	74	1.7

The study also looked at test results of university participants with background knowledge compared to those without. Results (see Figure 5) show that participants with no background knowledge of Kokoda showed 0.47 points (4.7%) higher test scores using Kokoda VR compared to 360°. This means participants in the study have shown that using VRLEs can assist in mastery of the concepts. The same case with participants of Kokoda VR with background which showed increased test results of 0.6 points (6%) compared to the 360° video participants with background. In both cases there is an increased test results using Kokoda VR. With this we can conclude that using Kokoda VR there is a higher level of learning experience.

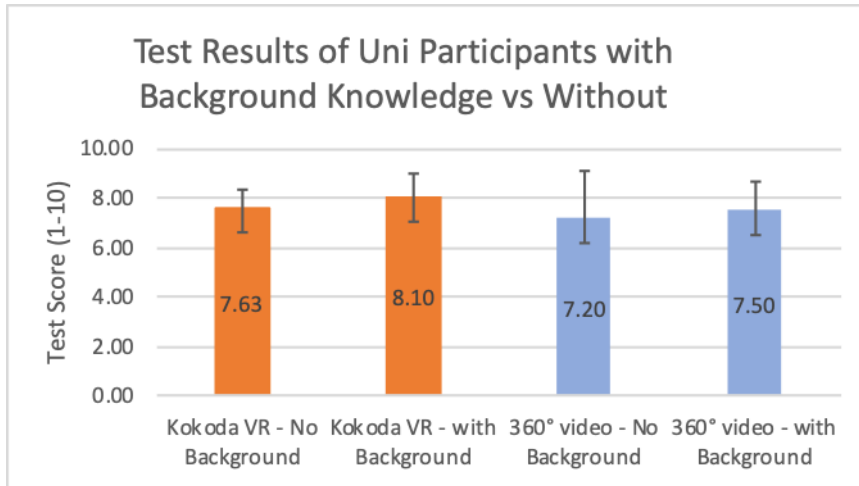


Figure 5. Graph Showing the Comparison of Test Results Between University Participants – Background Knowledge vs No Background Knowledge

5.3 Relationship between Empathy and Mastery Learning

The relationship between empathy and mastery learning in using VRLE is compared by using the empathy and test scores of participants. For both high school and university participants, high empathy scores also showed some high-test scores. The following graphs show the association between empathy and mastery learning in VR for participants (Figure 6) and 360° video participants (Figure 7).

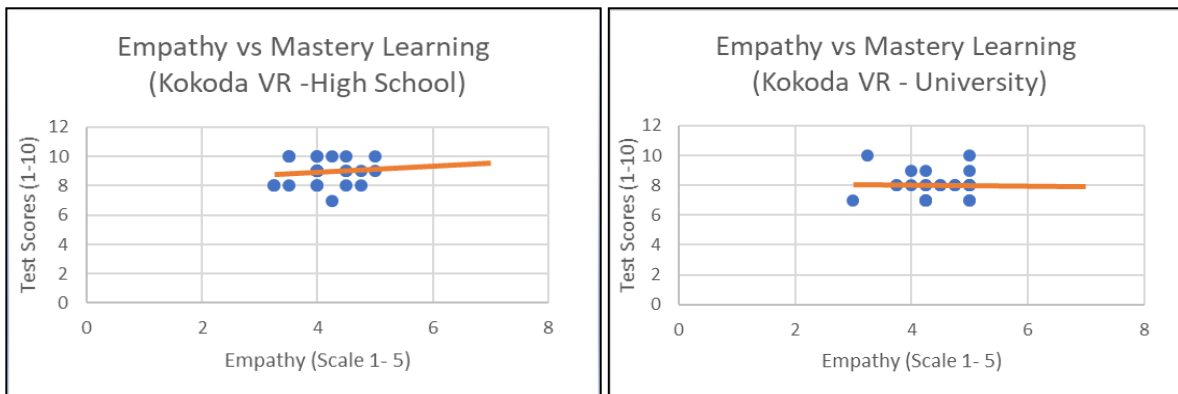


Figure 6. Association between Empathy and Learning for Kokoda VR participants.

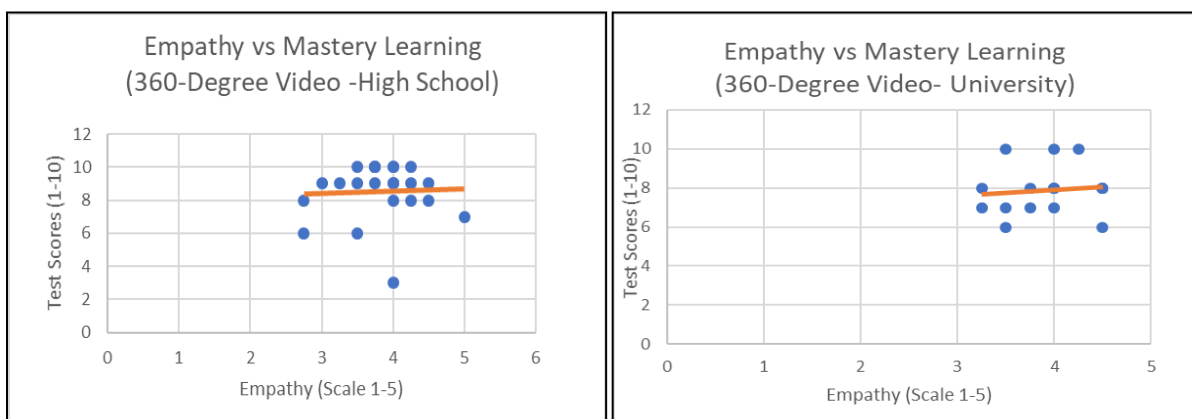


Figure 7. Association between Empathy and Learning for 360° video participants.

The coefficient of determination for Kokoda VR high school participants is 0.0144 while 0.0016 for 360° video participants. For university participants the coefficient of determination for Kokoda VR is 0.0004 while 0.0103 for 360° video participants. Although all showed positive association the relationship is weak, which is not sufficient to conclude that a high empathy leads to better mastery learning.

6. Conclusion and Recommendations

This study has shown how the use of an immersive VR has potential in improving the student empathy and mastery learning. Further studies are needed to identify relationships of other salient features of VRLEs (e.g., engagement, perception, immersion, motivation) to student learning, and find if there are causalities in these relationships. The results of the study show that students who use the Kokoda VR have higher empathy and mastery learning compared to those who use the 360° video. Although there is a positive association between empathy and mastery learning, the coefficient determination is low which is not enough to say that the two salient features have a strong relationship. The limitation of the study presented in this paper is that it did not include the effect of long-term recall and use of focus group to evaluate student's empathy. Although it was not included in the paper, a focus group where students were asked by their history teachers to talk about their experience were included in the study design. Students were also asked to take a quiz a month after the experiment to test the long-term recall. That will be reported in future presentations.

Acknowledgements

This research was supported in parts by funds received from the David A. Wilson Award for Excellence in Teaching and Learning, which was created by the Laureate International Universities network to support research focused on teaching and learning. For more information on the award or Laureate, please visit www.laureate.net

References

- Abadia, R., Calvert, J., & Tauseef, S. M. (2018). Salient Features of an Effective Immersive Non-Collaborative Virtual Reality Learning Environment. *10th International Conference on Education Technology and Computers*, Tokyo, Japan.
- Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, 35(11), 919-925.
- Anderson, R. E. (1992). Social impacts of computing: Codes of professional ethics. *Social Science Computing Review*, 10(2), 453-469.
- Bloom, B. (1968). *Learning for Mastery*. Formative and Summative Evaluation of Student Learning. Mc-Graw-Hill.
- Bhargava, A., Bertrand, J. W., Gramopadhye, A. K., Madathil, K. C., & Babu, S. V. (2018). Evaluating multiple levels of an interaction fidelity continuum on performance and learning in near-field training simulations. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1418-1427.
- Bonner, T. D., and Aspy, D. N. (1984, June). A Study of the Relationship Between Student Empathy and GPA. *J. Humanist. Educ. Dev.*, vol. 22, no. 4, pp. 149-154.
- Brennan, K., Daily, S. B., & Kaman, C. (2018). Empathy as a Foundation of Civic Engagement: Using Technology and Storytelling to Cultivate Perspective Taking Abilities. *Conference Paper*, Accessed at: <https://www.semanticscholar.org/paper/Empathy-as-a-Foundation-of-Civic-Engagement-%3A-Using-Brennan/d0782ab136467e72cd27d9f012f7f137021db42b>. On 28/11/2018.
- Bruno, P., Ongaro, A., and Fraser, I., (2007). Long-term retention of material taught and examined in chiropractic curricula: its relevance to education and clinical practice. *The Journal of Canadian Chiropractic Association*, 51 (1), 14-18.
- Butt, A. L., Kardong-Edgren, S., & Ellertson, A. (2018). Using Game-Based Virtual Reality with Haptics for Skill Acquisition. *Clinical Simulation in Nursing*, 16, 25-32. <https://doi.org/10.1016/j.ecns.2017.09.010>
- Buttussi, F., & Chittaro, L (2015). Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety. *IEEE Transactions on Visualization and Computer Graphics*, 21(4), 529-538. <https://doi.org/http://dx.doi.org/10.1109/TVCG.2015.2391853>

- Calvert, J., Abadia, R. and Tauseef S.M. (2019). Design and Testing of a Virtual Reality Enabled Experience that Enhances Engagement and Simulates Empathy for Historical Events and Characters. *IEEE VR 2019 (Poster Paper)*. Tokyo, Japan
- Chittaro, L., & Buttussi, F. (2015, April) Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety. *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 4, pp. 529–538.
- Cooper, B. (2011). *Empathy in Education: Engagement, Values and Achievement*. Bloomsbury Publishing, 2011.
- Decety, J., & Ickes, W. (2011). *The social neuroscience of empathy* (pp. 85-87). Cambridge, Mass.: MIT Press.
- Foster S., Halbstein D. (2014) Introduction. In: *Integrating 3D Modeling, Photogrammetry and Design. Springer Briefs in Computer Science*, Springer, London.
- Google for Education. (2019). Bringing virtual and augmented reality to school. Retrieved from https://edu.google.com/products/vr-ar/?modal_active=none
- Grühn, D., Rebucal, K., Diehl, M., Lumley, M., & Labouvie-Vief, G. (2008). Empathy across the adult lifespan: Longitudinal and experience-sampling findings. *Emotion* (Washington, D.C.), 8(6), 753–765. doi:10.1037/a0014123
- Kandaurova, M., & Lee, S. H. (2018). The effects of Virtual Reality (VR) on charitable giving: The role of empathy, guilt, responsibility, and social exclusion. *Journal of Business Research*. doi:10.1016/j.jbusres.2018.10.027
- Kuronen-Stewart, C., Ahmed, K., Aydin, A., Cynk, M., Miller, P., Challacombe, B., ... Popert, R. (2015). Holmium Laser Enucleation of the Prostate: Simulation-Based Training Curriculum and Validation. *Urology*, 86(3), 639-646.
- Lucas, J. (2018). Immersive VR in the construction classroom to increase student understanding of sequence, assembly, and space of wood frame construction. *Journal of Information Technology in Construction (ITcon)*, 23(November 2017), 179–194. Retrieved from https://itcon.org/papers/2018_09-ITcon-Lucas.pdf
- Mangione, G. R., Discepolo, T., Tore, P. A. D., Tore, S. D., Cozzarelli, C., & Corona, F.(2013). Measuring Empathy to Support Learning Design and Narrative Game: A Phenomenological Approach in 2013. *Seventh International Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 401–406.
- O'Brien, E., Konrath, S. H., Gruhn, D., & Hagen, A. L. (2012). Empathic Concern and Perspective Taking: Linear and Quadratic Effects of Age Across the Adult Life Span. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(2), 168-175. doi:10.1093/geronb/gbs055
- Oculus VR (2018). Announcing Oculus Education Pilot Programs in Taiwan, Japan, and Seattle. Retrieved from https://www.oculus.com/blog/announcing-oculus-education-pilot-programs-in-taiwan-japan-and-seattle/?locale=en_US
- Rehan Dar, F. (2016). Empathetic and Pro-social Awareness in Primary School Students: A Case Study. *Universal Journal Of Educational Research*, 4(10), 2394-2402. doi: 10.13189/ujer.2016.041019
- Riva, G., Baños, R. M., Botella, C., Mantovani, F., & Gaggioli, A. (2016, September 30). Transforming experience: The potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in Psychiatry. Frontiers Media S.A.* <https://doi.org/10.3389/fpsy.2016.00164>
- Schutte, N. S., & Stolinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, 41(6), 708-712. doi:10.1007/s11031-017-9641-7
- Shin, D. (2018). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior*, 78, 64–73. <https://doi.org/10.1016/j.chb.2017.09.012>
- Shin, D., & Biocca, F. (2017). Exploring immersive experience in journalism. *New Media & Society*, 20(8), 2800-2823. doi:10.1177/1461444817733133
- Smith, A. (2006). Cognitive Empathy and Emotional Empathy in Human Behavior and Evolution. In *Psychol. Rec.*, vol. 56, no. 1, pp. 3–21.
- Southgate, E., Smith, S. P., & Scevak, J. (2017). Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth. *IEEE Virtual Reality (VR)*. Los Angeles, CA, pp. 12–18.

Effects of game-based learning on informal historical learning: A learning analytics approach

Xuanqi FENG^{a*} & Masanori YAMADA^b

^a*Graduate School of Human-Environment Studies, Kyushu University, Japan*

^b*Faculty of Arts and Science, Kyushu University, Japan*

*dukefeng@mark-lab.net

Abstract: Game-based learning for informal learning has become an issue in digital game-based learning research. However, assessments from the observations of the learning process are difficult in the non-face-to-face situation of an informal settings. This research aims to evaluate the effects of a game-based informal learning system for history, “Hist Maker” integrating the external assessment with tests and the game-embedded assessment with the analysis of players’ gameplay log data. For the data analysis, we integrated the statistical model and learning analytics technology through cluster analysis. This approach allowed us to draw conclusions about the correlation between players’ behavior patterns and learning effects in the game. These conclusions show the potential of this approach to solve the observation problem in research on serious games for informal learning.

Keywords: digital game-based learning, serious game, informal learning, learning analytics, serious game analytics

1. Introduction

As entertainment media, digital games are considered engaging and appealing inspiring numerous attempts to integrate education and entertainment using digital games so that learners can be attracted to learning activities. Meanwhile, many scholars have taken an interest in how to support education through digital games, such as testing whether digital games can improve players’ cognitive skills (Greenfield et al., 1994) or figuring out the elements that make playing games fun and then trying utilizing these elements in education to enhance learners’ motivation (Malone & Lepper, 1987). In particular, after 2000, the terms “Digital Game-Based Learning” (abbreviated DGBL) and “Serious Game” (abbreviated SG) have been widely used as the names of research fields that concentrate on the use of digital games for educational purposes.

Many researchers have classified the features of digital games from the perspective of cognitive science. Based on the consideration of using different games in education, they found that the features of good games, such as interactive problem solving and adaptive challenges, are often also features of good learning environments (Shute & Ke, 2012). Thus, good games can improve skills and performance, support deep meaningful learning, and function as a revolutionary digital learning tool to create an effective learning environment. This approach is called Digital Game-Based Learning (Prensky, 2001). A “Serious Game” is defined as “a game that is not only for entertainment”. The concept of SG not only refers to a genre of games, but also to a wide range of issues such as the taxonomy of serious games, extensions of the concept, and the development and application of games to solve problems in education and society or the application of game technology (Sawyer & Smith, 2008). Now that games for learning are included in the category of SG, the design, development, implementation, and assessment of educational games are common topics of SG study.

More specifically, digital games for informal learning are one of the issues in research on DGBL and SGs. Informal learning means “the type of learning that is not organized and not in a structured learning environment” (Ainsworth & Eaton, 2010). A report of THE LIFE Center (Bank et al., 2007) shows that humans acquire various knowledge and skills in their lifetimes, and most of the places where they are acquired are informal. Therefore, the number of researchers on informal learning,

such as game-based learning environment design in an informal setting as well as on the support of learning in museums or libraries, has increased (e.g., Chang et al., 2008). However, digital games for informal learning present particular challenges for assessment. Because the features of informal learning include voluntary participation and a fluid time structure, and because they normally occur in a non-face-to-face situation without teachers, observation of the learning process is difficult (Squire & Patterson, 2009). Thus, the relevant research has mostly used external assessments like tests or questionnaires (Fujimoto & Yamada, 2013), which makes the learning activity a “black-box”. It is thus hard to explain the experimental results because of a lack of information on the learning process (Loh, 2011). This challenge makes it difficult to interpret the results of assessment and to use it as feedback to support the improvement of game design.

With the development of information technology, gameplay data collection, especially for real-time data, is becoming easy to implement. Therefore, a new assessment approach called “game-embedded assessment” has been proposed (Shute et al., 2009), whereby players’ operation logs are collected during game play to assess their activities in the game. This can be expected to solve the problem for informal learning research of observation in games. The basis of this method of assessment is Evidence-Centered Design (abbreviated ECD), which has been applied to various fields to obtain information on the learners’ learning situation without interrupting participation in learning activities such as gameplay behaviors in the SG (Shute, 2011). ECD is a framework to support assessment, aiming to combine learning demands with player actions (Mislevy et al., 2003). How to analyze these players’ operation logs so as to draw educationally meaningful conclusions has thus become a significant issue in game-embedded assessment. This issue can be considered from the perspective of Learning Analytics (abbreviated as LA) (Hauge et al., 2014). According to the NMC Horizon Report (Johnson et al., 2014), LA can be defined as “an educational application of ‘big data’”. LA research uses data analysis to help make decisions about education system. LA research involves not only dealing with the data generated by users through their interaction with the learning environment, such as players’ operation logs, but also analyzing the texts written by students on an online study system or learners’ social network (Romero & Ventura, 2013). In short, LA deals with all types of educational “big” data sets. However, some researchers suggest that the analysis of gameplay data from a serious game should be distinguished from LA because SGs have unique characteristics different from those of mainstream subjects of LA, and Serious Game Analytics (abbreviated SGA) needs more ubiquitous metrics (Loh et al., 2015).

There have been numerous recent studies of game-embedded assessment or SGA. Many of them concentrate on games that can improve skills or can be implemented in classroom settings. For example, Kang et al. (2017) used data mining technology to examine the problem-solving strategies of students in a formal setting, and Shute & Ke (2015) tested the impact of serious games on learners’ cognition skills with an in-game measurement. However, the analysis of knowledge learning games in informal settings is still relatively rare. Thus, this study utilizes LA or SGA to apply game-embedded assessment to a knowledge-learning serious game for informal learning so as to solve the observation problem of assessment in informal settings.

2. The Serious Game "Hist Maker"

To attain our research purpose, we developed a serious game called “Hist Maker”, which is intended to support learning of historical knowledge. The game has several stages, and each stage delivers historical knowledge about one era in one country. The game can be run on a computer with Windows Operating System or a smart phone with an Android System, and the game was published on the Internet to allow learners to access and play the game voluntarily, which makes it a learning activity in an informal situation. Even though the game has interfaces in Simplified Chinese, Traditional Chinese, Japanese, and English, this game is currently mainly developed as a Chinese-oriented game. In this study, participants as objects of the study are constrained so that only the data generated by mainland Chinese players will be analyzed. Interfaces in the game are shown in Figure 1.

2.1 Gameplay

The gameplay of “Hist Maker” is to some extent like a puzzle game. That is, the gameplay includes problem (i.e., puzzle) solving procedures that use the information collected from the game and players’ former knowledge (Kendall et al., 2008). The core gameplay of “Hist Maker” is based on the conception of a “Concept Map,” which is believed to encourage meaningful learning and serve as a good cognitive tool (Novak & Cañas, 2008). More specifically, we designed a “formula” mechanism as the core gameplay of the game. We introduced items called “elements” that display concepts and relationships in the concept map, and in the design of each stage of the game, historical knowledge is transformed for presentation in a concept map. Then, the concepts and relationships represented in the concept map are converted into a form whereby new “elements” are generated through the interaction between “elements”. A form in which two such “elements” are combined to obtain a new “element” can be regarded as the formula “element A” + “element B” = “element C.” The player has only a few “elements” when starting the stage. In the process of gameplay, players acquire “elements” through “synthesis” (combination). When they acquire new “elements,” text explaining the concepts and relationships appears. In this way, the player can explore the concept map presented in the form of a “formula” and study the concepts, knowledge, and relationships among the concepts.

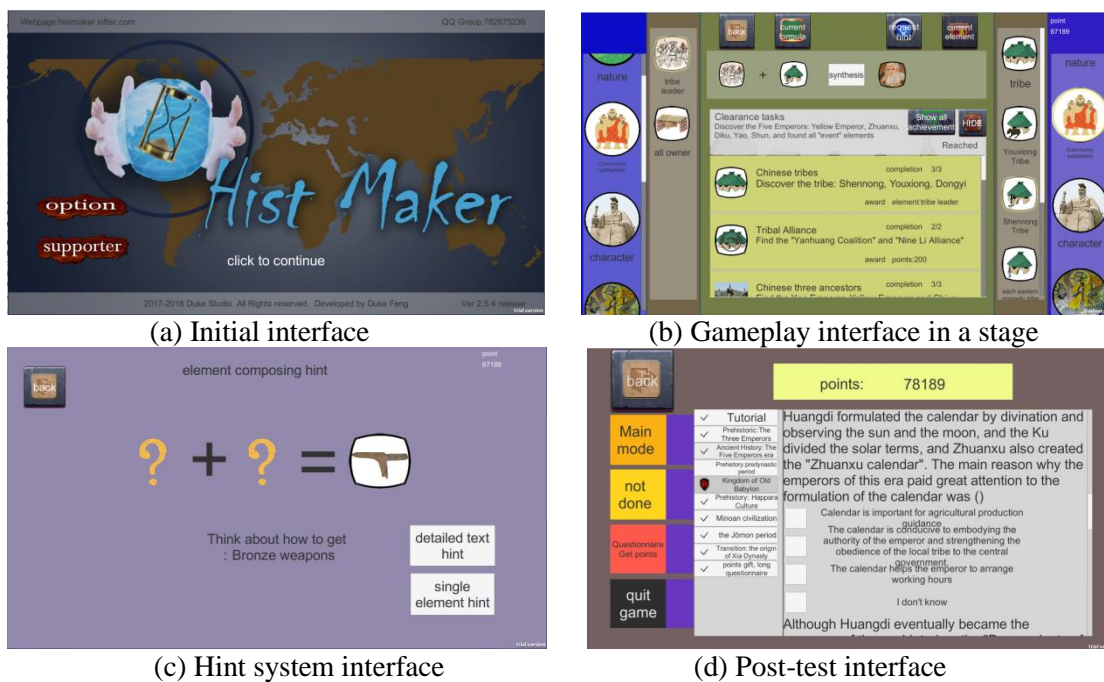


Figure 1. Interfaces in the game “Hist Maker”

2.2 Supporting Tools

The game includes tools to support players’ gameplay and improve the effects of the game. Malone and Lepper’s study (1987) showed that one of the features that makes games appealing is clarified goals, and Shute and Ke (2012) stated that a game should have rules to follow and goals to achieve to help players focus on what to do. Such goal-based scenarios can create a good environment for situation learning. In order to provide clarified goals, we developed the “Task List” tool in the game, which includes various tasks embedded in each stage and requires “elements” and “formulas” to be acquired as achievement conditions. The list is in the middle of the gameplay interface seen in Figure 1(b). In particular, a “clear task” is set for each of the stages. When a player completes this task, most of the content in the stage has been explored. The players can then continue on to the challenge of more difficult tasks. These tasks, as problems to be solved, serve as interactions between the player and the game, guiding the direction of the player’s thoughts and actions.

A good game has a balanced difficulty level, one set to match players’ ability. In the best game and learning environments, the challenge lies at the boundary of the learners’ ability (Gee, 2003). We accordingly developed a “Hint System” for the game, which is a function to present hints on problems such as what “element” can be obtained at that point or how to obtain it. This system lowers the level of difficulty and provides a moderate challenge to players who have difficulty in the game. As a scaffold,

hints can lead players to higher levels of knowledge. In particular, for those who lack historical knowledge, reading the hint is an important way to learn. When the player clicks the “Request Hint” button at the top of the gameplay interface, the hint system interface is shown as in Figure 1 (c). Moreover, as there may be too many “elements” and “formulas” in one stage, we developed two tools to reduce the cognitive load of players, “Show Acquired Elements” and “Show Acquired Formulas,” so that players can clearly see the “elements” and “formulas” they have acquired in the game at any time.

2.3 Test & Questionnaire Function

Since it is difficult to evaluate players’ achievement of knowledge learning from the gameplay behavior data, in this study we applied external assessment as well as game-embedded assessment. Since the players played “Hist Maker” voluntarily without monitoring by the researchers, we needed to develop the “Test & Questionnaire Function” embedded in the game to allow players to answer the questions of the test or questionnaire and submit the answers to us through the Internet. The function is embedded in the game, but as it is not shown while players are playing in a stage, it would not disturb the players’ mind flow as they concentrate on gameplay.

2.3.1 Test

To determine the changes in players’ knowledge, a pre-test was set at the first execution of the game, and after clearing each stage, a post-test was set. Each stage has seven single-choice questions about the historical knowledge presented in the “formula” form in the stage, and the questions on the pre-test and post-test are the same. In order to reduce the occurrence of situations where the player encounters a question whose correct answer he/she doesn’t know and randomly chooses the correct answer, we included an “I don’t know” option for every question. To improve the validity of the test, we referred to the history course guidelines of China and imitated the test items on university entrance exams and academic examinations in China, and then modified the item in consultation with an active history teacher in Chinese high school. The interface of the post-test is seen in Figure 1 (d).

2.3.2 Questionnaire

At the first execution of the game, players need to answer not only a pre-test, but also a pre-questionnaire. Since everyone can download “Hist Maker” on the Internet, it is difficult for us to specify the characteristics of players that are significant for educational research, such as gender or educational background. Furthermore, according to the lecture review part of Powers et al.’s meta-analyses (2013) of the effects of video game play on information processing, interest in games, experience playing games, and the type of the game are considered potential effective factors. Moreover, Uguroglu and Walberg (1979) pointed out that learning motivation could affect achievement. Therefore, the pre-questionnaire includes questions on the players’ gender, educational background, interest in games, experience playing games, self-considered amount of historical knowledge, and interest in learning history.

2.4 Game Telemetry

To implement the game-embedded assessment, we designed and developed game telemetry to record the gameplay log data of the players. Game telemetry is data related to a particular game event, game state, or other parameters that need to be recorded. The goal promoting the game telemetry collection is to develop meaningful evaluation methods from a integration of player behaviors and game states. Developed under guidelines for the design of game telemetry (Chung, 2015), the telemetry includes data recorded at the finest usable grain size within the context of the game situation, such as the current game state or the result of the action.

3. Data Collection and Analysis

3.1 Restrictive Condition

As the object of this study was restricted to mainland Chinese players, only the data from players who played the game in the simplified Chinese interface were analyzed, and this study only deals with the test and gameplay log of one stage: “The Five Emperors era.” The content of this stage is about the Legendary Era of Prehistoric China before the Xia Dynasty. Although there is not enough archaeological evidence for this era, certain amounts of ancient literature and documents exist. The knowledge presented in the game is based on the Records of the Grand Historian (Shiji) of Sima Qian, the book considered the most well-known source for the history of ancient China.

3.2 Procedure of Data Collection

Considering that players play the game without any supervision, to ensure the acquisition of data meeting the requirements of analysis, the procedure of data collection is embedded in the design of the test and questionnaire function and game telemetry.

The procedure is conducted below: First, when the game is first executed, the player is asked to answer the pre-questionnaire and pre-tests about the historical knowledge in all the stages. Unless the player submits the pre-questionnaire and pre-tests, he/she is not able to play the game. Next, the player has to play the tutorial stage to learn the gameplay of “Hist Maker” and become familiar with the interface, and then the player can choose one stage to play and learn knowledge from it. While the player is playing, the actions of gameplay are recorded by the game telemetry. The types of actions, time-stamps, and contextual information related to actions are all recorded. As soon as the player completes the “Clear Task” of one stage, the post-test of this stage will be unlocked, at which time the player can answer the test immediately or continue to the challenge of more difficult tasks in the stage and complete the test later. When the player submits the post-test, the gameplay log data will be sent at the same time, and once the player submits the post-test, the test will no longer be sent to us. This procedure guarantees that the pre-questionnaire, pre- and post-test, and gameplay data for a stage can be collected when a player submits the post-test.

3.3 Participants

Before we started to analyze the data, we have received data sets that met the requirement from 185 players with 196423 action records in total. These players played the game completely spontaneously without any intended recruitment. The result of the pre-questionnaire shows that, excluding one player who answered with blanks, there were 133 male (72.3%) and 51 female (27.7%) players, and 26 players (14.1%) from primary school, 38 players (20.7%) from middle school, 40 players (21.7%) from high school, 55 players (29.9%) from university or graduate school, and 25 players (13.6%) with other educational backgrounds.

3.4 Data Analysis

The central method of data analysis in this study is cluster analysis, which researchers have attempted to use in Serious Game Analytics research in recent years (Loh & Sheng, 2015). The concrete procedure is:

1. Since the questions in tests have an “I don’t know” option, which cannot be regarded as either a correct or incorrect answer, we referred to research that dealt with the same situation (White, 2012) and coded the change between the pre-test and post-test: “Positive” refers to a wrong answer or “I don’t know” answer for the pre-test question and a correct answer for the post-test. “Keep” refers to a correct answer for the pre-test question and a correct answer for the post-test. “Misunderstand” refers to an “I don’t know” answer for the pre-test question and wrong answer for the post-test. “Worse” refers to a correct answer for the pre-test question and wrong answer or “I don’t know” answer for the post-test. “Invalid” refers to the other situations.
2. To determine the proper parameters for the cluster analysis, we used a “parameter tuning” approach, which is a common approach in AI applications (Hutter et al., 2007). Concretely, in this study, we made a series of combinations of different actions’ frequency as parameter sets, then used each of these sets as input parameters for clustering. To handle the clustering results, we examined whether the numbers of “Positive” changes among clusters showed significant differences using Analysis of Variance

(ANOVA). If a significant difference exists, it means that different clusters had different learning effects, which is the expected result. If not, we have to change the selected parameters until the expected result appears, whereupon the clustering parameters are considered proper.

3. For clustering, we clustered the data with the K-means algorithm. The cluster analysis output numbers of clusters of players and each cluster represents a behavior pattern for gameplay.

4. To examine how different behavior patterns lead to different learning effects, we used ANOVA to examine the differences in the number of players showing all kinds of changes among clusters. Also, ANOVA was used to examine the behavior pattern of each cluster.

4. Results

4.1 Parameter Selection

After repeated attempts at multiple parameter sets, we found that the standardized frequencies of five actions could serve as proper parameters: “Close the panel of ‘Task Complete’” (“Task Complete” for short), “Require the detailed hint with an element in the formula” (“Require Element Hint” for short), “Close the Panel of ‘Acquired All Elements’” (“All Elements” for short), “Require the detailed hint with the instruction of the formula” (“Require Instruction Hint” for short), and “Click the task item in the list” (“Click Task Item” for short).

4.2 Clustering

Cluster analysis can group the samples by their similarities and without a rigid classification standard, and the K-means algorithm is an algorithm to implement such clustering.

However, for this algorithm, the number of clusters (k value) must be assigned before analysis, and there is an “elbow method” to help to determine the optimal k value. The specific method is to calculate the cost of a range of k values, then plot a graph of the cost for each k value; the point at which the downward trend slows sharply is the “elbow.” The corresponding k value is regarded as the optimal value. For the selected proper parameters, we decided on a k value of 3. For the given parameters and k value, the algorithm groups the players into three clusters. Because there are five parameters, we use the TSNE dimension reduction algorithm to display the result of clustering in two dimensions. The result is shown in Figure 2. Cluster 1 (red points) had 103 players, Cluster 2 (green points) 31, and Cluster 3 (blue points) 51.

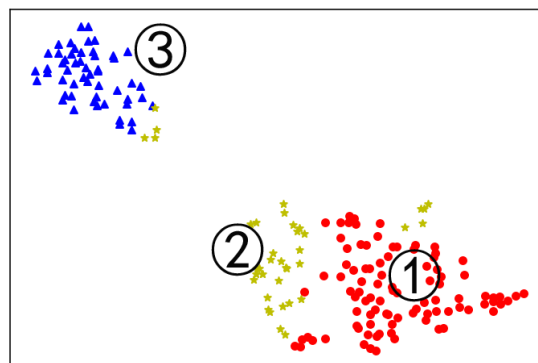


Figure 2. The result of clustering

4.3 ANOVA

4.3.1 Actions

To interpret the behavior pattern for each cluster, we examined the differences in the frequencies of actions by cluster using ANOVA. Because the actions selected to be parameters did not include actions about the tools “Show Acquired Elements” and “Show Acquired Formulas,” the actions we examined are the five selected actions and the two actions of using these two tools. There are significant

differences ($p < 0.05$) in the frequencies of all seven actions. To compare the differences between each pair of groups, we used post-hoc analysis. The results are shown in Table 1.

4.3.2 Learning Effectiveness

To determine the overall effectiveness of learning, not only was the number of “Positive” changes examined, but the other coded changes, including “Keep,” “Misunderstand,” “Worse,” and “Invalid,” were all examined. Moreover, since it is possible that prior knowledge affected the gameplay behavior, the number of correct answers was also examined. There were significant differences (with $p < 0.05$) for all examined items. We then used post-hoc analysis to compare the differences between each pair of groups. The results are presented in Table 1.

4.3.3 Players’ Characteristics

To determine whether the behavior patterns were influenced by characteristics that may affect the information process skills or achievements of players, we examined them from the pre-questionnaire. We used 5 point Likert scales to measure the degrees of such characteristics as interest in playing games so that these characteristics could be analyzed quantitatively. The Cronbach's α coefficient of these scales is 0.787. Gender and educational background cannot be measured on a Likert scale, so they were not examined. There were no significant differences for any of the examined characteristics.

Table 1
The results of ANOVA and post-hoc analysis

	Cluster 1 <i>M(SD)</i>	Cluster 2 <i>M(SD)</i>	Cluster3 <i>M(SD)</i>	F	P	Post-hoc comparison*
Task Complete	16.70(0.46)	15.61(1.52)	12.86(0.80)	371.858	0.000	C1>C2>C3
Require Element Hint	3.99(2.98)	8.61(6.20)	3.39(2.79)	22.760	0.000	C2>C1&C3
Require Instruction Hint	0.97(1.60)	6.00(3.29)	1.55(1.89)	73.335	0.000	C2>C1&C3
All Elements	1.00(0.00)	0.74(0.45)	0.00(0.00)	525.079	0.000	C1>C2>C3
Click Task Item	3.83(4.40)	15.55(11.86)	4.90(6.10)	38.055	0.000	C2>C1&C3
Show Acquired Elements	1.80(2.24)	1.65(1.80)	0.43(0.86)	9.274	0.000	C1&C2>C3
Show Acquired Formulas	2.32(3.12)	1.77(2.73)	0.71(0.92)	6.430	0.002	C1>C3
Positive Changes	2.13(1.36)	2.68(1.78)	1.84(1.03)	3.660	0.028	C2>C3
Keep Changes	2.92(1.71)	1.26(1.60)	2.29(1.71)	11.889	0.000	C1&C3>C2
Misunderstand Changes	0.52(0.92)	1.45(1.55)	0.53(0.92)	10.057	0.000	C2>C1&C3
Worse Changes	0.28(0.53)	0.23(0.50)	0.53(0.73)	3.737	0.026	C3>C1
Invalid Changes	1.15(1.03)	1.39(1.69)	1.80(1.33)	4.771	0.010	C3>C1
Correct Answers in Pre-test	3.20(1.73)	1.48(1.61)	2.84(1.58)	12.607	0.000	C1&C3>C2

* Probability of all post-hoc analytics: $p < 0.05$

5. Discussion and Conclusion

5.1 Interpretation of the Result

5.1.1 Behavior Patterns of the Clusters

ANOVA revealed significant differences among the clusters in the mean frequencies of the examined actions. The following discussion is based on the results of the post-hoc analysis.

The results show first that the players in Cluster 1 accomplished the most tasks, and all of them acquired all the elements in the stage. Simultaneously, they had relatively high frequencies of using the tools “Show Acquired Elements” and “Show Acquired Formulas.” Their high complementary results

mean they explored the concept map of the stage broadly, and thus we named this cluster the “Explore Cluster.”

Players in Cluster 2 required the most hints, whether hints on elements or hints on instruction, and they clicked the most task items showing the descriptions of the tasks. Thus, we named this cluster the “Hint Cluster.”

Finally, the players in Cluster 3 had the lowest complementary results for the stage and relatively low frequencies of using various kinds of tools. Moreover, none of them acquired all the elements in the stage. They seem to have had somewhat negative attitudes and were not willing to challenge the difficult tasks with the supporting tools, so we named this cluster the “Negative Cluster.”

5.1.2 Correlations Between Behavior Patterns and Learning Effects

With regard to learning effects, the ANOVA also showed significant differences among the clusters for every type of “code.” The following discussion is based on the results of post-hoc analysis.

The players in the “Explore Cluster” had relatively many “Keep” changes, as well as relatively many correct answers on the pre-test. They had relatively few “Worse” changes and “Invalid” Changes. The reason that they showed high complementary results in this stage is the wealth of knowledge they had of the stage, so that they easily handled the difficult challenges even though they hardly asked for hints on instruction. Since the challenges might have lain at the boundary of their abilities, they also might have demonstrated a positive learning effect (Gee, 2003).

The players in the “Hint Cluster” had the fewest correct answers on the pre-test, which means they had the most room for improvement. They had relatively many “Positive” changes but also the most “Misunderstand” changes, showing that the hint tools indeed supported the players with insufficient prior knowledge, and that by playing they could learn knowledge from the game. As to their having the most “Misunderstand” changes, that is inevitable since they had a greater lack of knowledge and more chances to make mistakes than players in the other two clusters. In summary, we tend to think of the overall effects for this cluster as positive.

Players in the “Negative Cluster” had quite many correct answers on the pre-test. However, they had many “Worse” and “Invalid” changes and relatively few “Positive” changes. The magnitude of the “Worse” changes suggests that some players might have chosen the correct answers by accident, and it seems likely that they quit playing before reaching challenges lying at the boundary of their abilities. The learning effects were apparently negative, showing that negative behavior patterns led to negative learning effects. Based on these results, trying to reduced the behavior patterns in the “Negative Cluster” by increasing the difficulty of “Clear Task” might be a feasible modification to improve the overall education effect of the game.

5.1.3 Correlations Between Behavior Patterns and Players’ Characteristics

ANOVA found no significant differences among the characteristics of the examined players by clusters, indicating that even if some characteristics of players might have influenced the cognitive skills related to the learning effects, they did not influence the behavior patterns directly. What should be noted is that the questionnaire merely investigated the players’ thoughts about themselves, which might not reflect the real situation. For example, the numbers of correct answers on the pre-test showed significant differences by cluster, but the answers to the question on the pre-questionnaire, “I have good knowledge of history,” showed no significant differences by cluster, suggesting that the players might not have known their own status well. That is a limitation of the pre-questionnaire. Psychological scale with higher validity and reliability should be used in the future.

5.2 Assessment Approach

The assessment approach used in this study integrated an external assessment and game-embedded assessment, while the analysis integrated a statistical model and LA technology. This approach constitutes the originality of this study. By this approach, conclusions about the learning effects and players’ behavior pattern could be drawn from the analysis results. These conclusions are considered feedback for modifying the game design. Thus, we consider this assessment approach to be able to

resolve the observation question for the DGBL in a totally non-face-to-face informal setting, which was the research question of this study. This assessment and study have some limitations discussed in the next part.

5.3 Limitations and The Future Research

There are three major limitations of this study that will be addressed in future work. First, in this study we only used the frequencies of the gameplay actions for Serious Game Analytics, even though the log data from the game telemetry provided plenty of details of players' behaviors. The narrow range of analyzed data could constrain the survey of the learning process. Currently, the timestamps of actions (Loh & Sheng, 2015), patterns of action sequences (Kang et al., 2017), and visualization approaches (Liu et al., 2016; Kaneko et al., 2018) have been used for Serious Game Analytics, and these methods will be used in future work to mine more meaningful results from the gameplay log.

Second, this study constrained the study object to only one stage, which means that the generalizability of this assessment approach needs to be examined. The results of parameter selection might not be appropriate for other stages. Therefore, the study object should be expanded to other stages that teach historical knowledge of different eras of different countries using different concept maps.

Lastly, even if this study showed that the learners could be assessed by the assessment that we developed, the assessment itself was not evaluated. As a formative assessment, the goal was to support modifications of the design of the learning environment, i.e., the game "Hist Maker" in this study. Modifying the game based on the conclusions of the assessment, and seeing if the modified game could be more effective for learning will be important tasks for future research.

Acknowledgement

This research is supported by JSPS KAKENHI Grants Nos. JP19H01716 and JP18K18657.

References

- Ainsworth, H. L., & Eaton, S. E. (2010). *Formal, non-formal and in-formal learning in the sciences*. Calgary (Canada): Onate Press.
- Banks, J.A., Au, K.H., Ball, A.F., Bell, P., Gordon, E.W., Gutiérrez, K.D., & Zhou, M. (2007). *Learning in and out of school in diverse environments: Life-long, life-wide, life-deep*. Seattle: LIFE Center and Center for Multicultural Education (University of Washington).
- Chang, C., Wu, S., Chang, M., & Heh., J.-S. (2008). Activity generator for informal learning in museum. In the Proceedings of the 7th WSEAS International Conference on E-ACTIVITIES, (E-ACTIVITIES 2008), Dec 29-31, Cairo, Egypt, 189-194.
- Chung, G. K. (2015). Guidelines for the design and implementation of game telemetry for Serious Games Analytics. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds. Cham (Switzerland): Springer International Publishing, pp. 59-79.
- Fujimoto, T., & Yamada, M. (2013). Review of educational games research for informal learning: Research frameworks and evaluation methods. *Japan Society for Educational Technology*, 37(2), 343-351.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave MacMillan.
- Greenfield, P. M., Brannon, C., & Lohr, D. (1994). Two-dimensional representation of movement through three-dimensional space: The role of video game expertise. *Journal of Applied Developmental Psychology*, 15(1), 87-103. doi:10.1016/0193-3973(94)90007-8
- Hauge, J. B., Berta, R., Fiucci, G., Manjon, B. F., Padron-Napoles, C., Westra, W., & Nadolski, R. (2014). Implications of learning analytics for Serious Game design. 2014 IEEE 14th International Conference on Advanced Learning Technologies. doi:10.1109/icalt.2014.73
- Hutter, F., Babic, D., Hoos, H. H., & Hu, A. J. (2007). Boosting verification by automatic tuning of decision procedures. *Formal Methods in Computer Aided Design (FMCAD07)*. doi:10.1109/fmcad.2007.4401979
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *The NMC Horizon Report: 2014 Higher Education Edition*. Austin, TX: The New Media Consortium.
- Kaneko, K., Saito, Y., Nohara, Y., Kudo, E., & Yamada, M. (2018). Does Physical Activity Enhance Learning Performance? : Learning Effectiveness of Game-based Experiential Learning for University Library Instruction, *The Journal of Academic Librarianship*, 44(5), 596-581. doi.org/10.1016/j.acalib.2018.06.002
- Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72, 757-770. doi:10.1016/j.chb.2016.09.062

- Kendall, G., Parkes, A., & Spoerer, K. (2008). A survey of NP-complete puzzles. *International Computer Games Association Journal*, 31(1), 13-34.
- Liu, M., Lee, J., Kang, & Liu, S. (2016). What we can learn from the data: A multiplecase study examining behavior patterns by students with different characteristics in using a serious game. *The Technology, Knowledge and Learning Journal*, 21(1), 33e57.
- Loh, C. S. (2011). Using in situ data collection to improve the impact and return of investment of game-based learning. In *Proceedings of ICEM-SIIE 2011, the 61st International Council for Educational Media (ICEM) and the XIII International Symposium on Computers in Education (SIIE) Joint Conference* (pp. 801-811). ICEM-SIIE.
- Loh, C. S., & Sheng, Y. (2015). Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies*, 20(1), 5-19.
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). *Serious Games Analytics: Theoretical Framework*. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, Cham (Switzerland): Springer International Publishing, pp. 3 - 29
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In *Aptitude, Learning and Instruction. Volume 3: Cognitive and Affective Process Analysis*, R. E. Snow and M. J. Farr, Eds., Hillsdale, NJ: Lawrence Erlbaum, pp. 223-253.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A Brief Introduction to Evidence-Centered Design*. Research Report RR-03-16. Princeton, NJ: Educational Testing Service.
- Novak, J. D., & Cañas, A. J. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition.
- Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*, 20, 1055-1079.
- Prensky, M. (2001). The games generations: How learners have changed. In *Digital game-based learning*, M. Prensky, Ed. . New York, NY: McGraw-Hill, pp. 35-51.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- Sawyer, B. & Smith, P. (2008) *Serious Games Taxonomy*. Presentation, Game Developer Conference 2008, San Francisco, CA.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In *Computer Games and Instruction*, S. Tobias and J. D. Fletcher, Eds. Charlotte, NC: Information Age, pp. 503-524.
- Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives*, D. Ifenthaler, D. Eseryel, and X. Ge, Eds. New York, NY: Springer, pp. 43-58.
- Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, 80, 58-67.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In *Serious Games: Mechanisms and Effects*, U. Ritterfeld, M. Cody, and P. Vorderer, Eds. Mahwah, NJ: Routledge, Taylor and Francis, pp. 295-321.
- Squire, K. & Patterson, N., (2009). *Games and Simulations in Informal Science Education*. Paper commissioned for the National Research Council Workshop on Gaming and Simulations, October 6-7, Washington.
- Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: A quantitative synthesis. *American Educational Research Journal*, 16(4), 375-389. doi:10.3102/00028312016004375
- White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, 96(3), 419-438.

Examining the Effect of Gamification in Information Science, Computer and Engineering Education: A Meta-analysis of Student Learning Performance

Shurui BAI^{a*}, Khe Foon HEW^b & Biyun HUANG^c
Faculty of Education, The University of Hong Kong, Hong Kong SAR
*shurui18@hku.hk

Abstract: Gamification is commonly defined as the use of game design elements in non-game contexts. Compared with the traditional lecture-based class, a gamified class is commonly seen as more engaging and joyful. Anecdotal reports suggest that students perceive the use of gamification positively. But does gamification really improve student learning performance? The conflicting results reported in previous studies make the decision to support (or to dismiss) the use of gamification in education difficult. This meta-analysis examined the overall effect size of gamification on student learning performance in the contexts of information science, computer and engineering education. Thirteen studies employing between-subject designs that compared the effects of gamified versus non-gamified courses on student learning performance constituted the current sample. Results using the random-effect model revealed a significant small effect of gamification on students' learning achievement (Hedges' $g = 0.36$, $CI = 0.006 - 0.714$, $p = 0.046$). Results of various moderator analyses were discussed. Suggestions regarding the selection and use of the game design elements were provided. We conclude that the use of gamification results in higher learning gains compared with the non-gamified courses within the contexts of information science, computer and engineering education.

Keywords: Gamification, information science, computer and engineering, learning performance, meta-analysis

1. Introduction

Gamification is widely described as the use of game design elements in non-game contexts (Deterding, Dixon, Khaled, & Nacke, 2011). Game design elements are the basic building components of gamification and the common game design elements include badges, challenges, leaderboard/rank, levels/unlock, storyline, points, progress bar (Sailer et al., 2017).

Gamification shows a natural relation with information science, which has developed from an entertainment-oriented technology only for productive aims into mainstream in information science domain (Koivisto & Hamari, 2019). Despite the increasing demand of the market, software engineering and computer science disciplines (e.g., programming) present a high dropout rate in schools (Narasareddygari, Walia, & Radermacher, 2018).

Hitherto, recent reviews show that empirical studies on gamification in information and computer science education have markedly surpassed other subject disciplines (Dicheva, Dichev, Agre, & Angelova, 2015; Subhash & Cudney, 2018). However, findings that emerged from the use of gamification in the field of information, computer and engineering education have been mixed. For example, on one hand, Marín et al. (2019) stated that the effect of gamification to teach C programming is encouraging since students earned higher marks when the gamified platform was used. On the other hand, De-Margos et al. (2017) reported that gamification failed to improve student learning compared to the control group. Students in the control (non-gamified) group obtained significantly higher scores in the final exams compared to the gamified group. This therefore leads us to an important question: "*Does gamification indeed improve students learning outcomes or it is merely a hype*"?

2. Literature Review

2.1 Previous Reviews

In this section, we briefly reviewed information, computer and engineering education literature in recent years (2010-2018). So far, only several narrative syntheses have been published.

Souza, Veado, Moreira, Figueiredo, and Costa (2018) conducted a systematic mapping study to identify what game-related methods had been used to support software engineering education. Leaderboards, points, and levels were found to be the most frequently used game elements. No quantitative analyses of effect sizes were reported.

Gari, Walia, and Radermacher (2018) examined 16 studies in computer programming and software engineering education. The researchers found that badges, points and leaderboard were the most commonly used game elements. The main findings of each study were summarized. However, similar to other previous reviews, this review did not present any quantitative analyses of the studies' effect sizes.

Narasareddygari et al. (2018) reviewed 16 empirical studies in software engineering and computer science education disciplines. The authors counted the frequency of most used game elements and summed up the benefits of gamified learning in each study, then some suggestions for the design of learning environment were provided. However, neither effect sizes to substantiate the effectiveness of gamification nor synthesis of key elements for a successful gamified class were presented. The review just summarized results from each study.

2.2 Contribution of the Present Study

We contributed the present meta-analysis examining the effectiveness of gamified class on student academic performance compared to non-gamified class. A meta-analysis, which integrates the results of several independent small studies, can provide a more precise estimate of the effect of an intervention. Meta-analyses can also determine the strength of the effect, establish whether the direction of effect is positive or negative, and examine possible sources of variation, or heterogeneity among studies. The research question was: “*What is the effect of gamified class in K-12 and university contexts on student academic performance in information science, computer and engineering education compared to traditional class?*”

3. Methods

3.1 Databases Searched

The study applied a systematic review at the beginning to select eligible papers for data extraction in meta-analysis. The process for paper selection was guided by the preferred-reporting of items for systematic reviews and meta-analyses (PRISMA) statement (Moher et al., 2015). We examined the online databases of ACM Digital Library, EBSCO, IEEE Xplore Digital Library, INSPEC, ProQuest, Scopus, Web of Science as they have been recommended in the fields of information, computer and engineering searching (Cavacini, 2015; Radjenović, Heričko, Torkar, & Živkovič, 2013). We included all published works in conference proceedings, journals, and full-text dissertations. The search string used in the literature search was: gamif* AND (education OR class OR course OR learning OR performance OR behavior OR outcomes OR evaluation OR impacts OR effects OR influence).

3.2 Inclusion Criteria

To be eligible for inclusion in the present meta-analysis, studies had to meet the following criteria:

- (a) The focus of the study was on the field of information, computer and engineering education;
- (b) The study compared student learning achievement between a gamified class versus a non-gamified class;

- (c) The outcome variable must be based on an objective assessment such as final exams, instead of student self-reported perceptions;
- (d) The study clearly described the game design elements used;
- (e) The study reported sufficient data for calculating effect size (e.g., sample size, mean, standard deviation);
- (f) The study had to be written in English, although no restrictions were imposed on the geographical locations in which the study was conducted.

3.3 Screening

A total of 13 out of 1365 papers met the inclusion criteria. Our initial search using the aforementioned databases yielded 1365 papers (Figure 1). 167 were found to be duplicate papers due to our use of multiple academic databases. Subsequent screening of paper title and abstract removed 1079 papers due to little relevance to the research topic. The remaining 119 papers were downloaded, and their full texts were carefully read. But 106 did not meet our meta-analysis inclusion criteria (see Figure 1). Finally, a total of 13 empirical studies employing the between-subject research design were included. Two coders independently examined the excluded papers and extracted the information from 13 articles. Intercoder reliability was 92%. The discrepancies were fully discussed and then resolved.

3.4 Effect Size Calculation

Key information such as the number of game elements used, the types of game elements, statistical data (e.g., sample size, mean, standard deviation), the setting of the interventions was extracted from each of the 13 papers. Two coders extracted the information and resolved any discrepancies of the coding. Effect sizes were computed using the Comprehensive Meta-Analysis software. All reported p-values were two-tailed. I^2 test was present for the test of heterogeneity. We used the random-effects model to compute the effect sizes since this model can account for variation in different study implementations (Raudenbush, 2009). Hedges's g , which is the adjusted standardized mean difference between two groups based on the pooled standard deviations, was used to report the effect sizes because it is particularly useful for the meta-analysis of studies with varying sample sizes (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016). One effect size was calculated for each study to meet the assumption of the independence of the effect sizes based on independent samples of students. If a study reported multiple assessments of a single course subject, we selected the most summative assessment, as suggested by Freeman et al. (2014). For example, we chose final examinations over other assessments. If a study reported multiple assessments (e.g., test 1, test 2, ...) Without the most summative assessment or overall result from the same set of participants, we computed a single combined mean effect size using the Comprehensive Meta-Analysis software. To identify the possible source of variance on effect sizes, moderator analyses were performed on several variables (see Table 1).

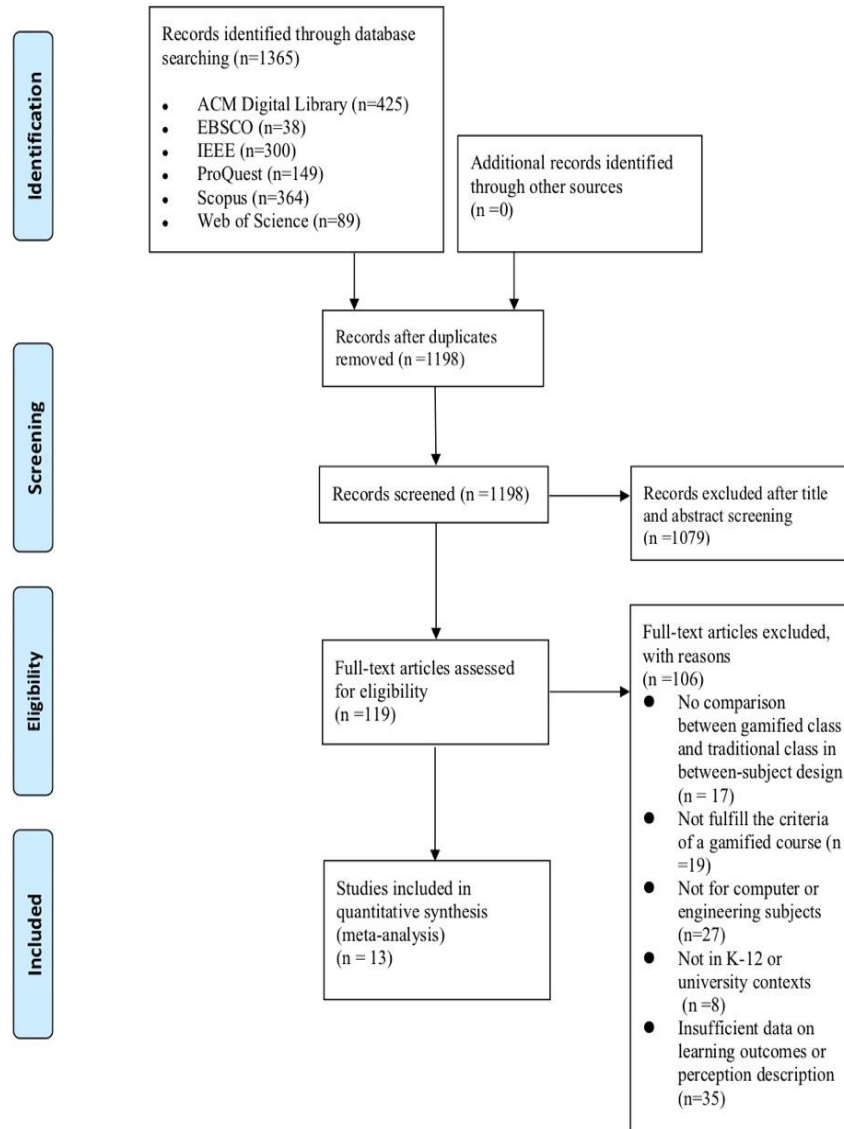


Figure 1. PRISMA flowchart of articles screening

4. Results

13 independent studies were selected for this meta-analysis with a total of 934 subjects in the gamified groups and 673 subjects in the non-gamified groups. An overall significant effect size (Hedges's $g = 0.36$, $CI = 0.006 - 0.714$, 95% confidence interval, $p = 0.046$) in favor of gamification under random-effects model (see Figure 2) was found. This suggests that the use of gamification results in higher learning gains compared with the non-gamified courses within the contexts of information science, computer and engineering education. A significant Q statistic ($p < 0.001$) indicated the presence of heterogeneity ($I^2=90.686\%$).

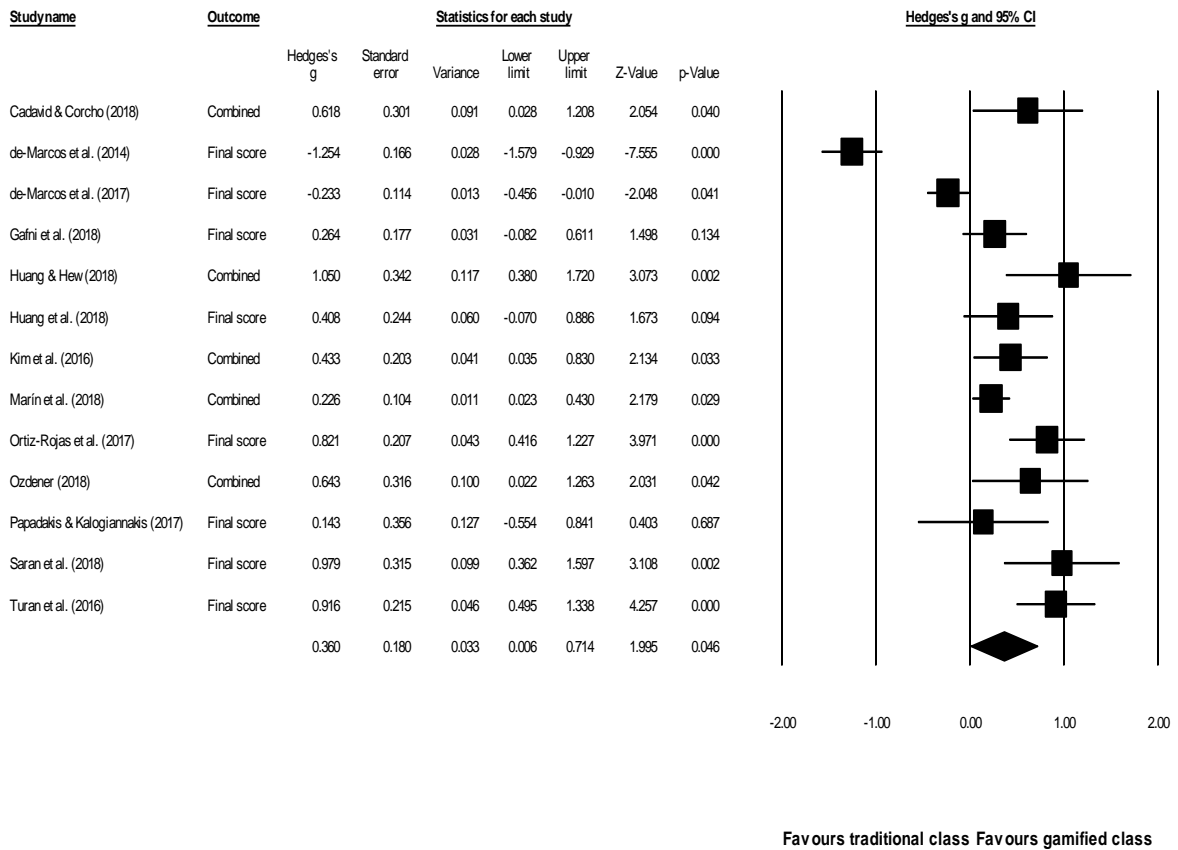


Figure 2. Forest plot of effect sizes (Hedges's g) using the random-effects model

4.1 Publication Bias

Publication bias refers to situations when authors deliberately publish significant-only results. To determine whether the results of our meta-analysis were affected by publication bias, we conducted the following tests: funnel plot, Begg and Mazumdar rank correlation, Egger's regression, fail-safe N test, and Duval and Tweedies' trim and fill. The funnel plot revealed a slightly asymmetrical position between negative and positive figures in Figure 3. However, Begg and Mazumdar rank correlation (Kendall's Tau with continuity correction) results of 0.269 (two-tailed $p = 0.2$), and Egger's regression intercept value of 3.64 (two-tailed $p = 0.121$) suggested that the overall mean effect size of 0.360 was not inflated by publication bias. Fail-safe N test showed 61 missing studies were required to bring the p -value over the alpha level (0.05). Given that we have used a broad search string, a large number of databases, as well as including both journal and conference publications, we believe that 61 missing studies would be an unreasonably large number of undetected studies with zero effect to bring the reported effect size of 0.360 to a statistically insignificant value. Due to these reasons, we believe that the overall mean effect size was not affected by publication bias.

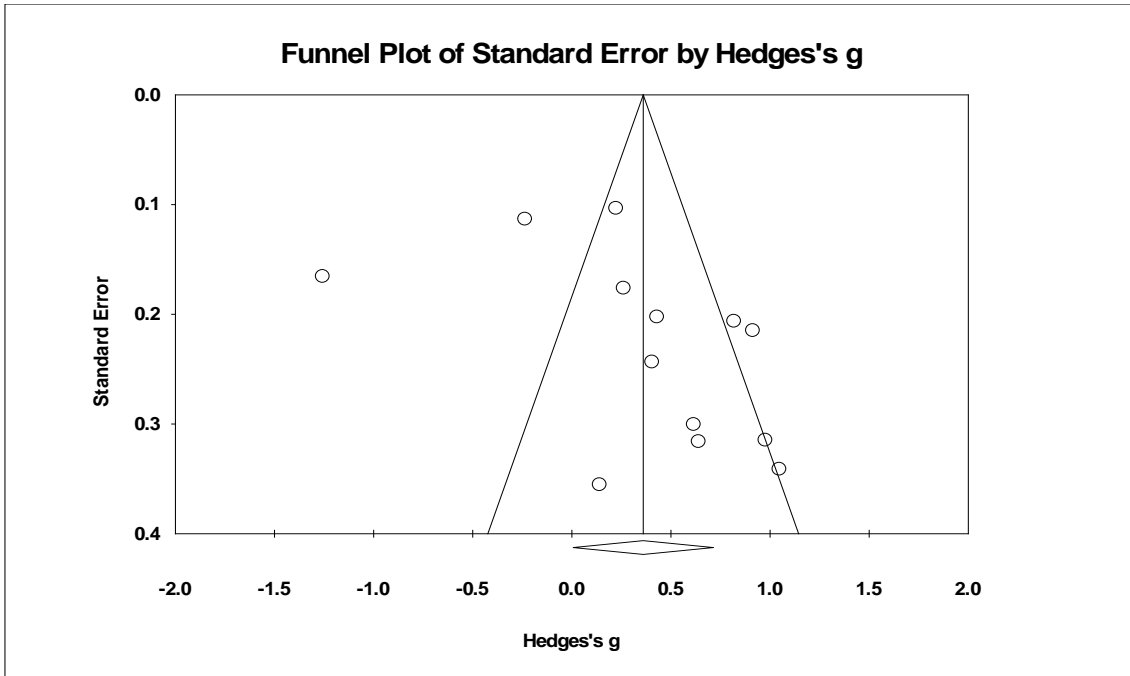


Figure 3. Funnel plot of standard error by Hedges's g under the random-effects model

4.2 Moderator Analyses

We performed several moderator analyses to explore the causal effects of student learning performance in gamified class. Three main categories of moderators were examined (Table 1): (a) controls of game design elements (the number and types of game elements used), (b) research design (research design type, student initial equivalence, and instructor equivalence), and (c) intervention condition (school setting, intervention duration and origin of study). The combination of badges, leaderboard/rank, level/unlock and points ($n=4$) was most commonly employed in the previous studies, followed by the combination of badges, leaderboard/rank, points ($n=2$). The origin of studies is listed in Table 1 and Figure 4. Spain, Hong Kong and Turkey showed high interests in gamification practices with regards to student learning performance measurement in information, computer and engineering science education.

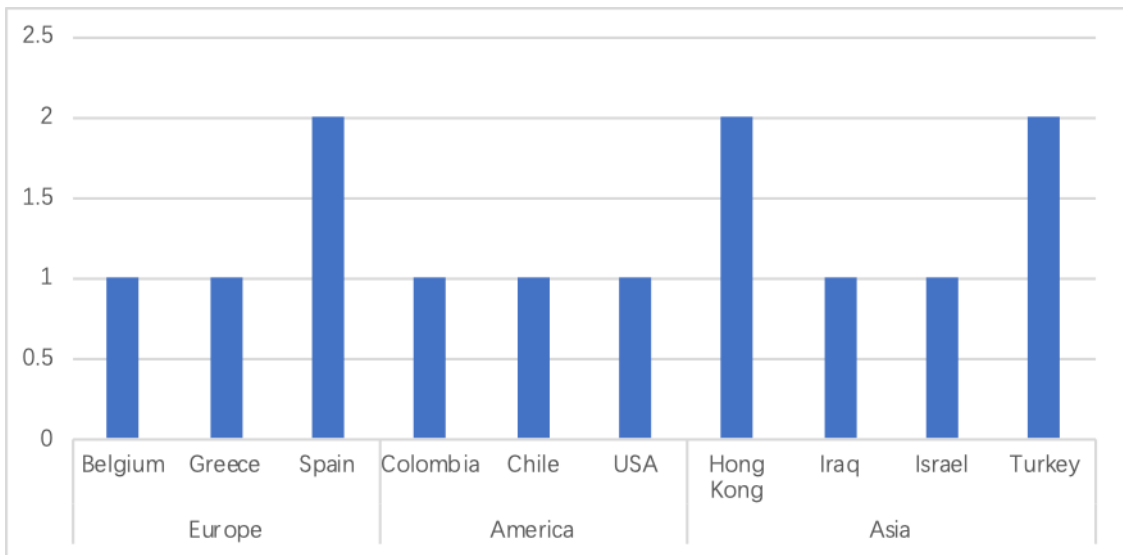


Figure 4. Origin of studies (continents and regions)

Table 1

Categories of moderator variables with numbers of studies

Moderator variables	Categories	Studies
Controls of game design elements		
Number of game design elements	1	1
	2	2
	3	4
	4	5
	6	1
	Type of game design elements	Badges + leaderboard/rank + level/unlock + points
Badges + leaderboard/rank + points		2
Badges + points + progress bar		1
Badges + leaderboard/rank + level/unlock + progress bar		1
Badges + leaderboard/rank + level/unlock		1
Badges		1
Badges + leaderboard/rank		1
Level/rank + points		1
Avatar + badges + leaderboard/rank + level/unlock + points + word notification		1
Research design quality		
Research design type	Quasi-experiment	9
	Randomization	4
Student initial equivalence (<i>i.e., whether students were indistinguishable before study</i>)	No statistical difference (pre-test)	8
	Randomized assignment	3
	No data reported	2
Instructor equivalence (<i>i.e., whether the same instructor taught the different groups</i>)	Different instructors	2
	Identical instructor	5
	No data reported	6
Intervention condition		
School setting	Elementary school	1
	High school	1
	High school + undergraduate + postgraduate	1
	Undergraduate	9
	Postgraduate	1
Intervention duration	A (one quarter of a term)	3
	B (two-quarters of a term)	3
	C (three-quarters of a term)	4
	D (more than one term)	1
	No data	2
Origin of study (continent)	America	3
	Asia	6
	Europe	4

Note: one academic term assumed to be 16 weeks (USA university standard)

We found no evidence of heterogeneity between studies that used more game elements and studies that used lesser game elements ($Q = 5.107$, $df = 4$, $p = 0.276$) (Table 2). However, results suggest a significant variation in terms of the types of game elements used ($Q = 16.007$, $df = 8$, $p = 0.042$). Effect sizes appear to be larger when badges + leaderboard/rank (Hedges's $g = 0.916$) or badges alone (Hedges's $g = 0.821$) were employed (Table 3). We observed an unexpectedly very large negative effect size (-1.254) in a single study that employed the combination of badges, leaderboard/rank,

level/unlock rendered a negative effect size (De-Marcos, Domínguez, Saenz-de-Navarrete, & Pagés, 2014), where the authors claimed the experimental group overemphasized skill acquisition in practical activities, leading to poorer scores on knowledge acquisition compared with traditional class. These results, nevertheless, have to be viewed with caution due to the very small number of studies involved.

Heterogeneity analyses indicated no significant variation when comparing (a) studies with different research design such as quasi-experiment, or true experiment ($Q = 0.02$, $df = 1$, $p = 0.889$); (b) studies that reported initial student equivalence, or not ($Q = 1.647$, $df = 2$, $p = 0.439$); and (c) studies with same or different instructors ($Q = 1.108$, $df = 2$, $p = 0.575$). There was also no evidence of heterogeneity when comparing (a) studies with different school setting ($Q = 1.892$, $df = 4$, $p = 0.756$); (b) studies with different intervention duration ($Q = 6.054$, $df = 4$, $p = 0.195$); and (c) studies conducted in different regions ($Q = 5.005$, $df = 2$, $p = 0.082$). Table 2 summarizes the results of various moderator analyses.

Table 2
Results of Q-test for heterogeneity for three main moderators

Moderator	<i>Q</i>	<i>Df</i>	<i>P-value</i>
Controls of game design elements			
Number of game design elements	5.107	4	0.276
Type of game design elements	16.007	8	0.042*
Research design quality			
Research design type	0.02	1	0.889
Student initial equivalence	1.647	2	0.439
Instructor equivalence	1.108	2	0.575
Intervention condition			
School setting	1.892	4	0.756
Intervention duration	6.054	4	0.195
Origin of study (continent)	5.005	2	0.082

* $p < 0.05$

Table 3
Effect sizes of type of game design elements

Type of game design elements	95% CI				
	<i>N</i>	Hedges's <i>g</i>	<i>SE</i>	<i>LL</i>	<i>UL</i>
Badges + leaderboard/rank + level/unlock + points	4	0.665	0.258	0.16	1.171
Badges + leaderboard/rank + points	2	0.129	0.348	-0.554	0.811
Badges + points + progress bar	1	0.264	0.474	-0.665	1.194
Badges + leaderboard/rank + level/unlock + progress bar	1	0.408	0.503	-0.578	1.394
Badges + leaderboard/rank + level/unlock	1	-1.254	0.47	-2.176	-0.332
Avatar + badges + leaderboard/rank + level/unlock + points + word notification	1	0.433	0.485	-0.517	1.383
Badges	1	0.821	0.486	-0.132	1.774
Badges + leaderboard/rank	1	0.916	0.49	-0.044	1.877
Level/rank + points	1	0.143	0.566	-0.966	1.252

Note: *n* number of studies, *SE* standard error, 95% *CI* 95% confidence interval, *LL* lower limit, *UL* upper limit

5. Conclusion and Discussion

This review provides a useful snapshot of the current quantitative studies on gamification with the contexts of information science, computer and engineering education. The results of this meta-analysis suggest that gamification can increase student learning performance compared to non-gamified courses. Nevertheless, it is important to note that the overall effect size of 0.360 reported here is considered a small effect (Cohen, 1988). In other words, gamification has a significant positive effect on student achievement, though the effect size is small under a random-effects model.

To our knowledge, this is the first meta-analysis conducted on gamification and its impact on student learning performance as measured using objective instruments such as student exam scores. The search string and databases were broad, thus allowing us to capture as many empirical studies as possible. The publication bias was calculated to avoid that only positive result studies were included, and no publication bias was found. Our comprehensive and careful selection of studies revealed only 13 such articles. Most of studies applied badges + leaderboard/rank + levels + points in combination. Badges and badges + leaderboard/rank remain top two for effect size figure, but this finding should be interpreted with caution due to the small frequency. Based on self-determination theory, humans possess three innate psychological needs to facilitate motivation to engage or not engage in an activity – autonomy, relatedness, and competence (Ryan & Deci, 2000). As one component of intrinsic motivation, the need of competence is more likely to be satisfied when students are given the autonomy to adjust difficulty levels. Therefore, badges and leaderboards can fulfil the most positive effects when students can obtain different badges/ranks when corresponding difficulty levels of learning tasks are accomplished. For game elements selection, progress bar, leaderboards and badges are often used to gain instant feedback, achievements and self-recognition. Students can make choices on learning tasks when levels and avatars are applied. Points, badges and progress bar can support goal-setting commitment.

We conclude by highlighting some follow-up research directions. One limitation of this review is that the search was limited to English and studies conducted within the contexts of information science, computer and engineering education. One follow-up research direction is to expand the search to include studies carried out in other disciplines. We noted that more high quality, rigorously designed experimental studies are needed in the field of gamification within the K-12 and university education settings. Carrying out such studies can help us better establish the causal effect of gamification on student learning. To date, most of the previous studies had only short intervention duration of less than one academic term (16 weeks). Short-term studies run the risk of novelty effect where the participants were interested to try out the intervention because it is something new and exciting. Longer-term longitudinal studies are needed to examine whether the effects of gamification on student learning hold over time.

References (* refers to studies included for meta-analysis)

- *Cadavid, J. n. M., & Corcho, A. s. F. P. (2018). Competitive programming and gamification as strategy to engage students in computer science courses. *Revista ESPACIOS*, 39(35), 11.
- Cavacini, A. J. S. (2015). What is the best database for computer science journal articles? , 102(3), 2059-2071. doi:10.1007/s11192-014-1506-1
- Cohen, J. (1988). *Statistical power ANALYSIS for the Behavioral sciences* (Vol. 2nd). Hillsdale: NJ: Lawrence Erlbaum Associates.
- *De-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education*, 75, 82-91. doi:10.1016/j.compedu.2014.01.012
- *De-Marcos, L., Garcia-Cabot, A., & Garcia-Lopez, E. (2017). Towards the Social Gamification of e-Learning: A Practical Experiment. *International Journal of Engineering Education*, 33(1), 66-73.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness: defining "gamification"*. Paper presented at the Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland.
- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study. *Educational Technology & Society*, 18(3), 75-88.

- *Eunsik, K., Rothrock, L., & Freivalds, A. (2016, 12-15 Oct. 2016). *The effects of Gamification on engineering lab activities*. Paper presented at the 2016 IEEE Frontiers in Education Conference (FIE).
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410-8415. doi:10.1073/pnas.1319030111
- *Gafni, R., Achituv, D. B., Eidelman, S., & Chatsky, T. (2018). The effects of gamification elements in e-learning platforms. *Online Journal of Applied Knowledge Management*, *6*(2), 37-53.
- Gari, M. R. N., Walia, G. S., & Radermacher, A. D. (2018). *Gamification in Computer Science Education: a Systematic Literature Review*. Paper presented at the ASEE Annual Conference & Exposition.
- *Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education*, *125*, 254-272. doi:10.1016/j.compedu.2018.06.018
- *Huang, B., Hew, K. F., & Lo, C. K. (2018). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments*, *1-21*. doi:10.1080/10494820.2018.1495653
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191. doi:10.1016/j.ijinfomgt.2018.10.013
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A Meta-Analysis of the Effects of Classroom Management Strategies and Classroom Management Programs on Students' Academic, Behavioral, Emotional, and Motivational Outcomes. *86*(3), 643-680. doi:10.3102/0034654315626799
- *Marín, B., Frez, J., Cruz-Lemus, J., & Genero, M. (2019). An Empirical Investigation on the Benefits of Gamification in Programming Courses. *ACM Transactions on Computing Education (TOCE)*, *19*(1), 1-22. doi:10.1145/3231709
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Reviews, P.-P. G. J. S. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *4*(1), 1. doi:10.1186/2046-4053-4-1
- Narasareddygari, M. R., Walia, G., & Radermacher, A. (2018). *Gamification in Computer Science Education: a Systematic Literature Re-view Gamification in Computer Science Education: A Systematic Literature Review*.
- *Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2017). *Gamification in computer programming: Effects on learning, engagement, self-efficacy and intrinsic motivation*. Paper presented at the Proceedings of the 11th European Conference on Games Based Learning, ECGBL 2017.
- *Özdener, N. (2018). Gamification for enhancing Web 2.0 based educational activities: The case of pre-service grade school teachers using educational Wiki pages. *Telematics and Informatics*, *35*(3), 564-578. doi:10.1016/j.tele.2017.04.003
- *Papadakis, S., & Kalogiannakis, M. (2017). *Using Gamification for Supporting an Introductory Programming Course. The Case of ClassCraft in a Secondary Education Classroom*.
- Radjenović, D., Heričko, M., Torkar, R., & Živković, A. (2013). Software fault prediction metrics: A systematic literature review. *Information and Software Technology*, *55*(8), 1397-1418. doi:https://doi.org/10.1016/j.infsof.2013.02.009
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In *The handbook of research synthesis and meta-analysis, 2nd ed.* (pp. 295-315). New York, NY, US: Russell Sage Foundation.
- Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, *55*(1), 68-78. doi:10.1037/0003-066X.55.1.68
- Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371-380. doi:10.1016/j.chb.2016.12.033
- *Saran, M., Al-Magsoosi, D., & Mohammed, D. (2018). *Gamification in e-learning: The Effect on Student Performance*.
- Souza, Mauricio R. de A., Veado, L., Moreira, R. T., Figueiredo, E., & Costa, H. (2018). A systematic mapping study on game-related methods for software engineering education. *Information and Software Technology*, *95*, 201-218. doi:https://doi.org/10.1016/j.infsof.2017.09.014
- Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior*, *87*, 192-206. doi:10.1016/j.chb.2018.05.028
- *Turan, Z., Avinc, Z., Kara, K., & Goktas, Y. (2016). Gamification and Education: Achievements, Cognitive Loads, and Views of Students. *International Journal of Emerging Technologies in Learning*, *11*(7), 64-69. doi:10.3991/ijet.v11i07.5455

Drug Defense: A Mobile Game for Prevention of Alcohol Abuse

Divine-kia T. TAN*, **Rodrigo Emmanuel ROY**, **Jose Rafael LASALA**, **Anton Gabriel YAP**, **Ma. Mercedes T. RODRIGO**, **Walfrido David DIY** & **Ma. Regina M. HECHANOVA**

Ateneo de Manila University, Philippines

*divinekiatagletan@gmail.com

Abstract: We present *Drug Defense*, a mobile tower defense game designed to prevent alcohol abuse. The game makes use of the Cognitive Behavioral Game Design (CBGD) methodology to encourage the retention of knowledge about the consequences of excessive alcohol use. The researchers compared used pretest and posttest to determine the game aided in greater retention. The researchers also assessed the playability of the game was also evaluated using a questionnaire adapted from the Heuristics to Evaluate the Playability (HEP) of Games. Data analysis showed that the players had significantly increased knowledge on alcohol and its use. Players also enjoyed the game's narrative and the need to strategize.

Keywords: Digital game-based learning, Alcohol knowledge, Health, CBGD

1. Introduction

The World Health Organization (WHO) reports that alcohol misuse caused three million deaths worldwide – 5.3% of all deaths – and 132.6 million disability-adjusted life years (DALYs) in 2016. More people died due to excessive alcohol consumption than people afflicted with tuberculosis, HIV/AIDS and diabetes combined. Out of all the deaths rooted in harmful alcohol intake in the same year around the world, 28.7% were due to injuries, 21.3% due to digestive diseases, 19% due to cardiovascular diseases, 12.9% due to infectious diseases and 12.6% due to cancers. Approximately 49% of alcohol-related DALYs were caused by noncommunicable and mental health conditions, and around 40% of DALYs are due to injuries (WHO, 2018).

Alcohol abuse has a wide variety of detrimental effects. Most alcoholics have a deficiency of Thiamine or Vitamin B1, which affects the cerebellum (National Institute on Alcohol Abuse and Alcoholism, 2004). Excessive intake of alcohol also can lead to alcoholic liver disease. The first stage of alcoholic liver disease is steatosis, or fatty liver, in which too much fat is present in liver cells (National Institute on Alcohol Abuse and Alcoholism, 2005). If alcohol use is not significantly decreased, then one is prone to alcoholic hepatitis, or liver inflammation (National Institute on Alcohol Abuse and Alcoholism, 2005). The most severe stage of the disease is cirrhosis, where the liver can no longer perform vital functions due to liver cells being replaced by scar tissue (National Institute on Alcohol Abuse and Alcoholism, 2005).

Alcoholism causes harm not only to the drinker themselves, but also to others, including their family or household, relatives and friends, and street counters. The harms are not limited to health (e.g. a family member's anxiety or depression), but also social (e.g. assault, community nuisance), and economic (e.g. damage to property, money) (Karriker-Jaffe et al., 2018).

Nations have established policies and interventions to lessen alcohol consumption and its consequences. One example is implementing drink-driving countermeasures which entail creating checkpoints by the police on public roadways to deter drunk drivers, penalizing those caught. Other than national policies, nationwide awareness-raising activities are being conducted. The availability of alcohol is also being regulated by controlling the production and sale of alcohol on a national level. In 2016, a national or subnational legal purchase age for on-premise beer and wine sales was observed in 152 countries, with 18 years old being the most common age limit (WHO, 2018). One of the most effective strategies for reducing the harmful use of alcohol is increasing its price (WHO, 2010). Another

technique for reducing the negative consequences of drinking is disclosing alcohol content on the labels of alcohol containers (WHO, 2018).

Another form of alcohol use prevention utilizes advertising and media. Studies show that adolescents' favorite alcohol brands are those most visible in music, film, television, and advertising (Moreno, 2011). A study estimated that reducing alcohol advertisements can significantly decrease adolescent alcohol use, and even binge drinking (Saffer & Dave, 2006). In addition to this, a longitudinal study found that those exposed to alcohol through media in early adolescence have an earlier onset of alcohol use (Ellickson et al., 2005). A systematic review on mass media campaigns established that the strength of media is in the recall of presented stimulus (Young et al., 2018). The use of media, primarily television and radio, also led to positive changes in knowledge, beliefs, and attitudes on alcohol consumption (Young et al., 2018).

Video games are one medium that can be used to educate people about the effects of alcohol abuse and persuade them to consume alcohol with moderation. Video games are persuasive. A literature review by Primack et al. (2012) concluded that video games have the potential to improve health-related outcomes. These changes, with the end goal of changing behavior, can also affect beliefs, attitudes, knowledge, skills, and risk perception (Thompson, 2012).

While the scope of the current build of Drug Defense is limited to alcohol abuse, the researchers hope to branch out to other addictive substances which are prevalent in the Philippines, such as methamphetamine and cannabis, to be a tool for educating the public. The game serves as a part of a larger program in substance abuse called *Katatagan Kontra Droga sa Komunidad* (KKDK). This program provides psychological and social support for low- to mild-risk drug users in the Philippines. Developing Drug Defense gives it a potential to be used in KKDK to alleviate observed monotony and increase engagement among participants.

2. Cognitive Behavioral Game Design

The researchers used Starks's CBGD (2014) to create Drug Defense, a game for alcohol use intervention. CBGD is the use of elements Bandura's Social Cognitive Theory (SCT) and Gardner's Multiple Intelligences (MI) theory for the purpose of creating games that induce real-world behavior change in the player (Starks, 2014).

Bandura's SCT claims that personal, environmental, and behavioral factors all affect one another into a resulting action (Bandura, 1999). An application of the theory (Bandura, 2004) states that there are five core elements that help facilitate behavior change and these are gamified through CBGD (Starks, 2014). These are knowledge, goals, outcome expectation, encouragement, and barriers. Knowledge refers to information that the game wants players to obtain (Starks, 2014). In Drug Defense, these are the facts about alcohol. Encouragement is facilitated in the game mainly through the doctor. The doctor praises the player through text, sound effects, and rewards whenever they answer correctly in the quiz.

The second set of elements in CBGD is based on Gardner's (1983) Multiple Intelligences (MI) theory, that combines learning with activities that engage students in accordance to their unique patterns of thought, or intelligences. The ideas of MI are concretized through correct expressions of various in-game assets, such as art and music (Starks, 2014). Logic/patterns are used to exercise reason (Starks, 2014). This can be seen through the consistent pricing of turrets and enemy behavior to help the give the player a sense of consistency throughout the game. The verbal intelligence of the MI theory was translated by Starks into words/languages in a game (2014). Drug Defense uses a dialogue box to convey texts needed in the game, such as the alcohol facts and the tutorial.

Beyond cognition, elements that heighten enjoyment are also present. The Enjoyment Process consists of engagement, challenge, flow, persistence, and mastery (Starks, 2014). Engagement is achieved when a player's attention is caught, leading to immersion in the game. Often, challenge refers to why players continue through the game (Starks, 2014). Flow is the state when the player is highly engaged and loses track of time (Csikszentmihalyi & LeFevre, 1989). Persistence is honed through the player's response to mistakes. Finally, mastery is what the player experiences when they win at a game and feels successful (Starks, 2014).

2.1 Gameplay

Drug Defense drew from the tower defense game design pattern. This pattern entails defending a base from incoming enemies using towers or turrets that the player can build and upgrade. Because the placement of the turret is important, players must use strategy in order to determine the best location for building the turret. Also, since building turrets cost limited in-game currency, the player must learn how to properly manage this resource. This combination of strategy and resource management helps form the foundation of the game's appeal.

Players took on the role of a medical apprentice. As an introduction, the doctor teaches the apprentice about the core mechanics of the tower defense gameplay. Afterward, players can enter the first stage. Each stage contains three to four waves, and each wave follows a similar flow. First, the player is presented a journal entry that tells the story of a day in the life of Miguel San Juan. Next, the doctor gives a fact pertaining to alcohol use. Then, the player undergoes the tower defense part of the game, which entails strategically positioning turrets in order to protect the liver from incoming alcohol. Afterward, the player is presented a multiple-choice quiz. The answer to said quiz should be known to the player if they had paid attention to the fact the doctor gave previously. This core gameplay loop repeats for eight stages, and the game takes approximately 35 to 45 minutes to complete.

In the tower defense part of the game, the player has to place turrets strategically along paths leading to the liver to avoid enemies from damaging it. They also have to manage and allocate resources, which can be used to buy turrets and upgrade them. Breaks between waves will provide player time to rest, re-evaluate their strategy, and upgrade the turrets as they wish. The moment that the enemies reach the liver, it will lose health. When the liver's health drops to zero, the player loses. The material of the liver gradually changes as it loses health, from healthy to that of one which has cirrhosis.

The emphasis of the gameplay is the defensive strategy the player employs. The player must understand how the enemy behaves and apply an effective strategy to stop them before they reach the liver. Some enemy substances are weak and fast, while others are strong but slow. The enemies take the form of an alcohol bottle. The alcohol bottles are designed according to the strength of each enemy. Strong enemies have more alcohol content in real life, while weak enemies have less alcohol content. Figure 1 shows an example of an attack.



Figure 1. Beer bottles attacking the liver.

A different enemy is spawned in waves. One wave emerges per time period. No new wave will be spawned until all enemies of the current wave are destroyed.

The narrative that accompanies the tower defense and strategy aspects of Drug Defense is intended to foster player attachment and sympathy. Each wave of enemies is prefaced by a journal entry, showing the main character's state. Each level starts easy then slowly becomes more difficult and

difficult. The last part of the game will be biased towards player victory, indicating the character's recovery. The game has a total of eight stages, 32 journal entries, and 31 waves.

The doctor figure in the game (Figure 2) serves as a general guide to the player. At the start of the game, the doctor administers a tutorial. Next, after showing the journal entry at the start of each wave, the doctor gives facts pertaining to alcohol use.



Figure 2. The doctor in the game.

Then, she administers a quiz (Figure 3) at the end of each wave in order to check if the player is retaining the alcohol use facts.

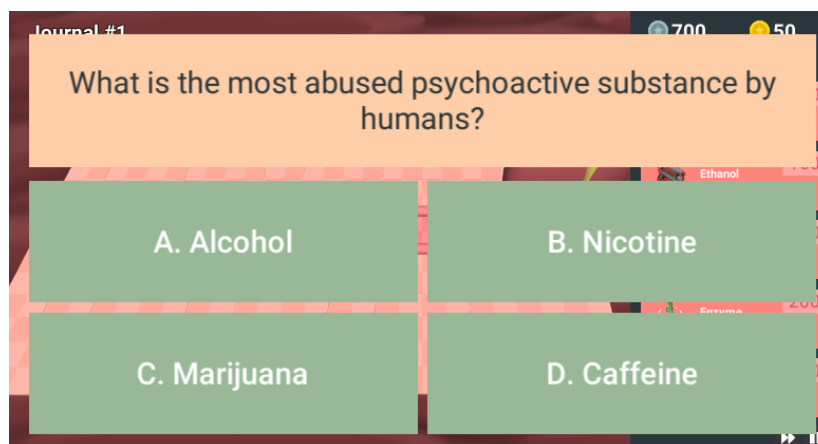


Figure 3. Quiz.

The doctor then responds to the results of the mini-quiz. Correct answers will result in an increase in an in-game currency that the player can use for building or upgrading turrets. This encourages the player to pay attention to the facts that are given throughout the game.

3. Testing

To determine the game's effectiveness, the researchers designed and conducted a field test testing the knowledge retention of participants who played Drug Defense. Ethics approval was obtained prior to the study.

To be eligible, participants had to have drunk alcohol at least once in their lifetime. Thus, participants recruited and sampled purposively from a university in Metro Manila. A total of 69 participants joined the study. Their ages ranged from 18 to 21.

Each participant selected a timeslot and was provided with a consent form. Each timeslot accommodated at most six participants.

At the start of the experiment, all participants completed the Alcohol Knowledge Test. The test consisted of 20 multiple choice questions about alcohol use facts, alcohol use disorder symptoms, as

well as the negative physical and cognitive effects of excessive alcohol use. This was followed by 10 true or false questions regarding protective behavioral techniques. Reliability analysis was conducted, and pretest Cronbach's alpha was 0.67 while posttest Cronbach's alpha was 0.71.

The participants were asked to play through the whole game, a total of 20 to 40 minutes. After playing, the Drug Defense participants were emailed a Google Form consisting of rating scale questions adapted from Desurvire's HEP (Desurvire et al., 2004) and open-ended questions relating to the game. In the latter, the players were asked what parts of the game they like the least, what parts they liked the most and if they have other comments. A total of 56 respondents answered the questionnaire. One week after each participant's timeslot, they were asked to complete the Alcohol Knowledge Test again.

4. Results

There is a significant difference between pre- and post-tests with an effect size of $d=2.54$, indicating an increase in knowledge about alcohol. In this regard, the game succeeded in achieving its goal of increasing player knowledge.

A total of 56 participants answered the questionnaire adapted the HEP. The highest possible score for each component is five. The average for gameplay, story, and mechanics are 3.92, 4.08, and 3.77 accordingly. Upon averaging all the heuristics, the game received a total score of 3.89/5. The findings imply that Drug Defense succeeded in providing players with an enjoyable gaming experience.

For the parts that the players liked the least, notable answers are the game was too easy, there was a bit a lot to read, and the mechanics were confusing at the first try. The last two were caused by the long text tutorial design. From this, the developers have improved the game to add a challenge mode and changed the tutorial from just a long text to an interactive one where the player can follow as each gameplay component is explained.

For the parts that the player liked the most, they mentioned that they enjoyed protecting the liver or strategizing, the dialogue and storyline, the knowledge the doctor gives them, and the quizzes. One of the most notable answers is that one player said that they genuinely wanted to know more.

In the other comments section of the survey, one player said that they found the game to be fun and that they learned a lot. Some respondents mentioned wanting to have the game released officially. In response to this request, one month after the experiment, the developers were able to put Drug Defense on Google Play Store and Apple App Store.

5. Conclusion

The researchers have developed Drug Defense, a mobile game designed using CBGD which aims to increase knowledge of alcohol. The game was tested for retention. Data analysis showed that the mobile game was successful in increasing student knowledge about alcohol. Game playability evaluation showed that although the game has its areas of improvement, the players found it playable and fun.

The researchers and developers consider expanding Drug Defense not only to cover alcoholism but also discuss the problem of other harmful substances and run similar experiments and game testing.

Acknowledgments

We thank the other members of the development team, Jan Michael Santos and Lance Hernandez for the 2D assets and Allen Ace Fortunato for the 3D assets used in the game. We thank the Ateneo Laboratory for the Learning Sciences and *Katatagan Kontra Droga sa Komunidad* for their support.

References

- Bandura, A. (1999). Social Cognitive Theory: An Agentic Perspective. *Asian Journal of Social Psychology*, 2(1), 21-41. doi:10.1111/1467-839x.00024
- Bandura, A. (2004). Health Promotion by Social Cognitive Means. *Health Education & Behavior*, 31(2), 143-164. doi:10.1177/1090198104263660
- Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, 56(5), 815-822. <http://dx.doi.org/10.1037/0022-3514.56.5.815>

- Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems - CHI 04*. doi:10.1145/985921.986102
- Ellickson, P. L., Collins, R. L., Hambarsoomians, K., & Mccaffrey, D. F. (2005). Does alcohol advertising promote adolescent drinking? Results from a longitudinal assessment. *Addiction, 100*(2), 235-246. doi:10.1111/j.1360-0443.2005.00974.x
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York, NY: Basic Books.
- Karriker-Jaffe, K. J., Room, R., Giesbrecht, N., & Greenfield, T. K. (2018). Alcohol's Harm to Others: Opportunities and Challenges in a Public Health Framework. *Journal of Studies on Alcohol and Drugs 79*(2), 239-243. doi:10.15288/jsad.2018.79.239
- Moreno, M. A. (2011). Media Influence on Adolescent Alcohol Use. *Archives of Pediatrics & Adolescent Medicine, 165*(7), 680. doi:10.1001/archpediatrics.2011.121
- National Institute on Alcohol Abuse and Alcoholism (2004, October). Alcohol's Damaging Effect on The Brain. Retrieved from <https://pubs.niaaa.nih.gov/publications/aa63/aa63.htm>
- National Institute on Alcohol Abuse and Alcoholism (2005, January). Alcohol Alert #64. Retrieved from <https://pubs.niaaa.nih.gov/publications/aa64/aa64.htm>
- Primack, B. A., Carroll, M. V., McNamara, M., Klem, M. L., King, B., Rich, M., . . . Nayak, S. (2012). Role of Video Games in Improving Health-Related Outcomes. *American Journal of Preventive Medicine, 42*(6), 630–638.
- Saffer, H., & Dave, D. (2006). Alcohol Advertising and Alcohol Consumption by Adolescents. *Health Economics, 15*(6), 617-637. doi:10.3386/w9676
- Starks, K. (2014). Cognitive behavioral game design: A unified model for designing serious games. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00028
- Thompson, D. (2012). Designing Serious Video Games for Health Behavior Change: Current Status and Future Directions. *Journal of Diabetes Science and Technology, 6*(4), 807-811. doi:10.1177/193229681200600411
- World Health Organization (2010). *Global strategy to reduce the harmful use of alcohol* [PDF file]. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/44395/9789241599931_eng.pdf
- World Health Organization (2018). *Global status report on alcohol and health* [PDF file]. Retrieved from <https://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639-eng.pdf>
- Young, B., Lewis, S., Katikireddi, S. V., Bauld, L., Stead, M., Angus, K., . . . Langley, T. (2018). Effectiveness of Mass Media Campaigns to Reduce Alcohol Consumption and Harm: A Systematic Review. *Alcohol and Alcoholism, 53*(3), 302-316. <https://doi.org/10.1093/alcalc/agx094>

The Effects of Virtual Reality System Applied to Shooting Training Course for Senior High School Students

Yi-Cheng Liao*, Hao-Chiang Koong Lin, Po-Chun Kuo & Kuan-Yu Lin
Information and Learning Technology, National University of Tainan, TAIWAN
*edcr@mail.nutn.edu.tw

Abstract: The main purpose of this study was to explore the effectiveness of virtual-reality shooting training system and traditional triangle aiming training on students' shooting performance, self-learning achievement and self-learning motivation for senior high school students. Second, the goal was to find out whether the relevance of shooting score to self-learning-performance and self-learning-motivation is positive or not. This study applied Game-Base Learning virtual reality shooting training system to construct human-computer interaction learning environment by utilizing 1:1 real-scale T91 teaching rifles with Bluetooth function, virtual reality shooting training application, which is suitable for Android platform, semi-Google Cardboard head, mounted viewer, TV streaming device with synchronous display function from mobile phone to projector. It combined the advantages of Immersion VR and Projection VR to achieve the purpose of synchronous learning. In this study, it took quasi-experimental design and divided 78 male students into two groups from two general classes in Tainan senior high school. The experimental group (40 students) adopted virtual reality teaching methods and the control group (38 students) adopted traditional triangle aiming training teaching method. The experimental courses were during for seven weeks, once a week, and 50 minutes for once. Finally, it analyzes the date including pre-test and post-test of shooting score, questionnaires, course recording video and interviews. The results showed that virtual reality shooting training system were better than traditional triangle aiming training on students' shooting performance, self-learning achievement and self-learning motivation.

Keywords: Virtual Reality, Game-Base Learning, Self-Learning Achievement, Self-Learning Motivation

1. Introduction

The main purpose of this study is to compare the traditional triangle aiming shooting method with the digital game-based virtual reality shooting training system. In the general high school national defense education curriculum, students' learning performance, achievement and self-motivation are explored in this study (Gong, Mei, Xiang, Mei, & Liu, 2017). This research dedicate to find whether there is a difference or correlation between academic performance and learning self-motivation. This study is expected to provide a reference and supplementary material when instructors teach high school national defense curriculum.

2. Literature Review

2.1 Game-Base Learning

The concept of Game-Base Learning is that using a variety of digital media to include the Internet, computers, corporate networks, satellite broadcasts, CDs, video tapes, audio tapes and interactive TVs. The range of this learning includes online learning, computer-assisted learning and collaborative learning of online digital classroom (Alden, 1998).

Game-based learning consists of three parts: learning, simulation, and games (see Figure 1). The Entertainment only consist of simulation and games, like ordinary video games in the world, purely is entertained-only. While Simulation Games consist only learning and games. It is like a simple kids

game and lacks the complexity of real world simulation. On the other hand, Simulation focus on learning and simulation, is like the military pilot training system or the training system of the medical profession (Martens, Diener, & Malo, 2008).

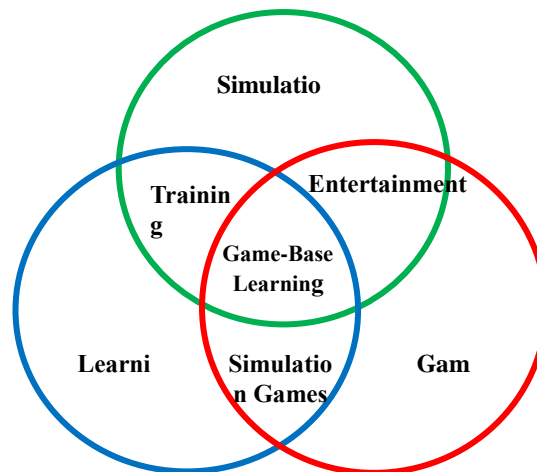


Figure 1. Three factors of Game-Base Learning

2.2 Serious Game

The game used for digital learning is called Serious Game. The term Serious Game was first invented in the United States and it was widely used for policy and management learning (Michael & Chen, 2006). Serious games can simulate the environment of the real world. It is not for entertainment which provide people a happy feeling, but mainly serve as a channel of educating and training operators. Militainment is one kind of Serious Game. It trains army force on the virtual battlefield in a game-like manner, communicate basic training information, and teach skills (Payne, 2014). The virtual reality shooting training system developed in this study belongs to this type of serious game, and the shooting skill of the T91 rifle can be learned through this game.

2.3 VR Military training

There are lots of researches or teaching software for digital information technology application in Taiwan, but are relatively rare for military training courses. On the contrast, the US Army is currently developing some VR Military training platform. Because the development of serious games requires joint cooperation by enterprises, academic institutions, and research institutions. The US Army's famous "America's Army" software is a good example. It is developed under the "US Army Training Guidelines", and allows players to choose between basic infantry or medical training with almost full range of current US equipment. Especially all weapons in the game are simulated using real data (Nichols, 2009). VR Military training can not only repeated exercises, but also to record the learner's training results in detail and to correct the individual's learning status. Similarly, that students can have the opportunity to learn repeatedly through computer simulation during trial and error process. In addition, they can learn independently and focus on the training course greatly.

3. Methodology

3.1 Research Process

This study is mainly to explore the how effective of virtual reality shooting training system compares to traditional teaching method on the students' academic performance, self-learning achievement and self-learning motivation in the national defense course of T91 rifle shooting training. In figure 2, this experiment used quasi-experimental design and organize the two classes of students with the same qualification as the experimental group and the control group. The experimental group were adopted the

virtual reality teaching method, while the control group adopted the traditional triangle aiming teaching method.

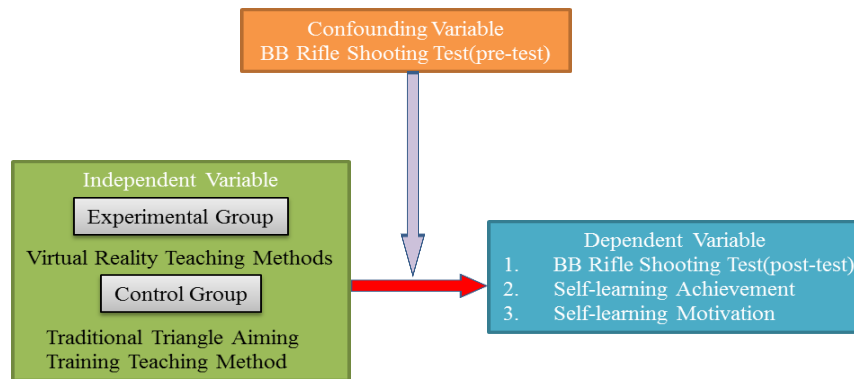


Figure 2. Experiment process

3.2 virtual-reality shooting training system

We used the Unity 5.3.1.f1 ver. to develop the virtual reality aiming shooting training APP Because Unity is a freeware, and there are abundant SDKs with templates (Kim et al., 2014). So we consider Unity ideal for beginners who have never developed any apps.

This research chose a shooting game called "Shooter Area VR", which is an authored free and modifiable template in the "VR Samples". We chose Android as the research platform and ran this VR app on mobile phone, because Android system is designed to adopt novice developers than the APPLE iOS platform (Ciesla, 2017). The components of this training system consist of floating marks and successfully shooting counters (Figure 3).

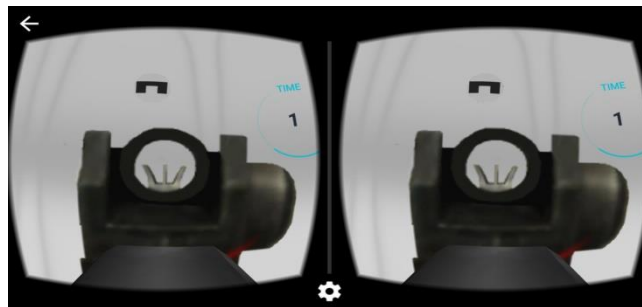


Figure 3. The marks and scores in shooting training system

In order to make students being immersed in the situation, and to get rid of the inconveniences of the wires, we use wireless headset VR BOX (Figure 4). The advantages of wearable device include to avoid suffering from environmental light which affect the aiming positioning, and no interference in the infrared aiming module. The other reason constraint the chosen of devices is the limited space in classroom, so the most economic benefit teaching method is wearable device like the VR headset (Alshaal et al., 2016). The combination of Bluetooth T91 training rifle and the wearable VR device, can be easily portable to any other class even outdoor environment.



Figure 4. VR BOX headset

One of the important factors affecting the accuracy of shooting is whether the posture of the gun holder obeying eight essentials of shooting, such as holding, resisting, holding, sticking, aiming, stopping, buckling, and reporting. Therefore, this research attaches the Bluetooth device with the 1:1 ratio of the T91 training rifle. The appearance, weight and sight of the rifle is identical to real T91 rifle, in which enables students to receive real shooting experience the essentials of operating the real T91 rifle when practically aiming. In addition, in order to enable the students to simultaneously trigger the shooting in virtual reality application, the researcher extends the button of the Bluetooth system remote control rear to the buckle of T91 teaching rifle (Figure 5). When the student buckles the trigger, the trigger forces the button of the Bluetooth remote control backwards, and execute the firing command of the training system (Espada et al., 2015). The advantages of the wireless virtual reality system is the nearly true experience of firing a rifle.



Figure 5. 1:1 real-scale T91 teaching rifles with Bluetooth function

Virtual reality generally require a display helmet or headset to allow its user to merge into an immersive experience. So the researcher used the synchronous display function of the AV bar to connect the current VR screen onto the projector. In this method, we synchronously project images and sound effects on the curtain in the front of classroom (Figure 6). And the operator and all other students in the classroom can learn at the same time. The other students waiting for shooting now could look at the operator's aiming posture and screen, so that they can learn the correct aiming method by observing the shooting score and failure of operator. Moreover, the other students might have verbal competition for the operator and increase the fun of learning (Ravyse, Blignaut, Leendertz, & Woolner, 2017). Other students can simultaneously learn the correct targeting method and increase the motivation in an asynchronous competition. In addition, the real-time display of results also increase the operator's desire to obtain higher scores and thus improve the learning achievement of correct aiming.



Figure 6. Synchronously project image to share VR screen with other classmates

4. Result

Comparing the differences between the two groups to understand whether the self-learning achievement and self-learning motivation are affected by different teaching methods, the independent sample T test is adopted. As shown in Table 1, in terms of self-learning Achievement, the control group who take traditional teaching have an average number 14.74, the standard deviation is 3.531, and the average value of standard error is .573. While experimental group who use virtual reality teaching method have an average 17.98, a standard deviation of 1.981, and a standard error of .313. The average number of experimental group (M=17.98) is higher than the average of control group (M=14.74). And in terms of self-learning motivation, the control group have an average number 10.29, the standard deviation is 2.977, and the average value of standard error is .483. While experimental group have an average 12.13, a standard deviation of 1.800, and a standard error of .285. The average number of experimental group (M=12.13) is higher than the average of control group (M=10.29). The T statistic of the two dependent variables reached a significant level, because both p-values were less than .05, indicating that different teaching methods have significant differences in self-learning achievement and self-learning motivation.

In the independent sample T test, if the T-value reaches a significant difference, the effect value can be further determined. The effect value represents practical significance, while the T statistic and p-value are represent statistical significance. The Eta square value η^2 is the effect value, indicating a highly correlated intensity between the groups and self-learning achievement (.279 \geq .14) as well as the groups and self-learning motivation are also a highly correlated intensity (.207 \geq .14). To judge whether the T value of the difference in two groups is significant or not, in addition to consider the value of the T value itself (two-tailed), we could observe the confidence interval of 95% area. If the interval estimate contains a value of 0, it must accept the null hypothesis. On the contrary, if the value of 0 is not included, the null hypothesis can be rejected and the opposite hypothesis can be accepted. In the term of the self-learning achievement, 95% CI are between (-4.545, -1.931), in which does not include 0. This indicate that there is a significant difference between the two groups. In the self-learning motivation, the confidence interval of the 95% are (-2.957, -.714). Without 0 inside this interval, it means a significant difference between the two groups too.

Table 1

Comparison table between self-learning Achievement and self-learning motivation of experimental group and control group

Factor	Group	N	M	σ	SEM	t	η^2	95% CI	
<i>self-learning Achievement</i>	<i>experimental</i>	40	17.98	1.981	.313	-4.960***	.279	-4.545	-1.931
	<i>control</i>	38	14.74	3.531	.573				
<i>self-learning motivation</i>	<i>experimental</i>	40	12.13	1.800	.285	-3.275*	.207	-2.957	-.714
	<i>control</i>	38	10.29	2.977	.483				

The SEM is equal to $\frac{SD}{\sqrt{N}}$

*p<0.05, ***p<0.001

5. Conclusion

According to the research results, digital learning method with virtual reality technology is efficient compared with the traditional triangle aiming shooting training method. It can improve the scores of learners and increases the motivation and achievement of students' self-learning. In addition, the head-mounted virtual reality shooting training system is not affected by light damage and wireless, instructors can teach in the general classroom. And those devices are easily to carry to another place and no longer suffer constraints of classroom venues. It can be integrated into a digital game-based teaching

system to enhance the effectiveness of teaching and the motivation of students (Woo & Yeom, 2012). When the creativity and technology are integrated into the teaching content, in addition to making the teaching activities would become more interesting and attractive. It also achieves the ideal state that learning by osmosis(Odom & Kelly, 2001).

5.1 Future Development and Suggestions

- 1.It is recommended that the future research add real shooting environment in the game. When students immerse in the actual shooting environment background (Lai, Hu, Cui, Sun, & Dai, 2017), they can adapt to the surrounding situation of the actually shooting experience earlier.
- 2.The actual aiming method is a series of combinations required coordination of both head and hands (Livingston, Ai, & Decker, 2018). So it is recommended to design a motion-sensing detector on the T91's rifle to simulate the actual shooting position of the gun.
- 3.There are no force feedback or shock response after shooting, they affect the accuracy of shooting. Another benefit of this kind of interacting rifle is that it can simulate the real recoil vibration in the practice, so that the learner can experience it and ease his mood when he shoot the real T91 rifle (Glassner, 2010).

References

- Alden, J. (1998). *A trainer's guide to web-based instruction: Getting started on Intranet-and Internet-based training*. Alexandria, VA: ASTD Press.
- Alshaal, S. E., Michael, S., Pamporis, A., Herodotou, H., Samaras, G., & Andreou, P. (2016). *Enhancing virtual reality systems with smart wearable devices*. Paper presented at the 2016 17th IEEE International Conference on Mobile Data Management (MDM).
- Ciesla, R. (2017). *Freeware Game Engines Mostly Codeless Game Development* (pp. 97-121): Springer.
- Espada, J. P., Díaz, V. G., Crespo, R. G., Martínez, O. S., G-Bustelo, B. C. P., & Lovelle, J. M. C. (2015). Using extended web technologies to develop Bluetooth multi-platform mobile applications for interact with smart things. *Information fusion, 21*, 30-41.
- Glassner, B. (2010). *The Culture of Fear: Why Americans Are Afraid of the Wrong Things: Crime, Drugs, Minorities, Teen Moms, Killer Kids, Muta*: Basic books.
- Gong, P., Mei, Z., Xiang, Y., Mei, C., & Liu, G. (2017). *Research on National Defense Students' Frustration Psychology and Its Management in Pre-service Training*. Paper presented at the International Conference on Man-Machine-Environment System Engineering.
- Kim, S. L., Suk, H. J., Kang, J. H., Jung, J. M., Laine, T. H., & Westlin, J. (2014). *Using Unity 3D to facilitate mobile augmented reality game development*. Paper presented at the 2014 IEEE World Forum on Internet of Things (WF-IoT).
- Lai, Z., Hu, Y. C., Cui, Y., Sun, L., & Dai, N. (2017). *Furion: Engineering high-quality immersive virtual reality on today's mobile devices*. Paper presented at the Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking.
- Liu, K., & He, Y. (2017). *Focusing on quality improvement and creating a new military education system*. Paper presented at the 2017 International Seminar on Social Science and Humanities Research (SSHR 2017).
- Livingston, M. A., Ai, Z., & Decker, J. W. (2018). *Human Factors for Military Applications of Head-Worn Augmented Reality Displays*. Paper presented at the International Conference on Applied Human Factors and Ergonomics.
- Martens, A., Diener, H., & Malo, S. (2008). Game-based learning with computers—learning, simulations, and games *Transactions on edutainment I* (pp. 172-190): Springer.
- Michael, D. R., & Chen, S. (2006). *Serious Games: Games that Educate, Train and Inform*. Boston, MA: Thomson Course Technology.
- Nichols, R. (2009). Target acquired: America's Army and the video games industry *Joystick Soldiers* (pp. 55-68): Routledge.
- Odom, A. L., & Kelly, P. V. (2001). Integrating concept mapping and the learning cycle to teach diffusion and osmosis concepts to high school biology students. *Science Education, 85*(6), 615-635.
- Payne, M. T. (2014). War bytes: the critique of militainment in Spec Ops: The Line. *Critical Studies in Media Communication, 31*(4), 265-282.
- Ravayse, W. S., Blihnaut, A. S., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: a systematic review. *Virtual Reality, 21*(1), 31-58.
- Woo, T., & Yeom, J. (2012). A Study on Medical Serious Game: Classification & Possibility. *Journal of Korean Society for Game, 12*, 2-12.

Analysis of Student Affect and Behavior while Playing a Mobile Game for English Comprehension

Ma. Mercedes T. RODRIGO^a, Jenilyn L. AGAPITO^a, Dominique Marie Antoinette B. MANAHAN^a

^a*Ateneo Laboratory for the Learning Sciences, Ateneo de Manila University, Philippines*
*mrodrigo@ateneo.edu

Abstract: We discuss a field study in which 30 public school students from grades 4, 5, and 6 play *Learning Likha*, a mobile-based game for practicing English comprehension. Using self-report questionnaires, a comprehension test, BROMP observations, and game interaction logs, we assessed the extent to which students understood the game's contents and enjoyed playing the game. We also tried to determine any relationships between student achievement, affect, and behavior. Self-reported feedback about the game was positive, with students reporting interest, enjoyment, and sufficient challenge. Students across all grade levels exhibited engaged concentration and on-task activity while playing the game. However, post-test comprehension scores were low, especially for the younger participants. On-task conversation and confusion were negatively correlated with achievement. Additionally, on-task behavior and engaged concentration were positively correlated with number of interface taps.

Keywords: mobile-assisted language learning, English comprehension, Philippines, Learning Likha, BROMP

1. Introduction

In the Global South, the biggest impediments to the integration of technology in education include the lack of computer hardware in schools, the absence or unreliability of electricity, and the absence or unreliability of Internet connectivity (Nye, 2015). Teachers and students in resource-constrained countries are then hard-pressed to take advantage of online educational materials (Wolfenden, 2017), social networks and virtual communities of practice (Tondeur, et al., 2016), and other such resources.

In the face of these barriers and the threat of exclusion, teachers and students have turned to mobile technologies. The primary computing platform of developing countries is the mobile phone. As of 2018, the developing world had an estimated 102.8 mobile phones for every 100 people (ITU, 2018). 61 out of every 100 people had mobile broadband access, and while this is compared favorably against fixed broadband subscriptions (10.4 out of every 100 people; ITU, 2018), it still implied that data access is limited. Furthermore, there is a lack of educational software that targets these platforms and contexts (Nye, 2015). This refers to education content written in the local language and is adapted to reflect local norms, cultures, situations, motivations, beliefs, and environments.

Mobile-Assisted Language Learning is an area within mobile learning that focuses on language learning areas such as grammar and vocabulary acquisition (Sung et al., 2015). Language learning is said to be effective if the learner is exposed to the target language in meaningful contexts. In this paper, we describe a field study in which we deploy and test *Learning Likha*, a narrative-based mobile phone game that helps students practice their English language comprehension skills. Our goals are to measure (a) the extent to which students are able to understand the game's contents, (b) the extent to which the game engages the students, and (c) any relationships between students' achievement, behaviors, and affective states.

2. Methods

The deployment and testing was conducted at AHA! Learning Center (ALC), a tutorial center in Makati, Metro Manila, Philippines that caters to Makati City public school students from elementary to high school. Staffed in large part by volunteers, ALC is committed to providing supplementary holistic education to underprivileged children.

The staff selected students from Grades 4, 5, and 6 and they were given informed consent forms that they and their parents had to complete in order to participate. During the study day, the participants were grouped by grade level. Each grade level had separate testing sessions because of the limited number of cellular phones and earphones. Upon settling down in the venue, they were oriented on the objectives. Then, they completed a demographic questionnaire to determine their level of access to mobile phones. The questionnaire also tried to determine their usage, attitude, and perceptions towards the English language by giving them 8 statements to which they would indicate their level of agreement (1 = Strongly Disagree to 5 = Strongly Agree). A sample statement is “I speak English at home.”

2.1 Learning Likha

After the questionnaire, each participant was lent a mobile phone with earphones. Each of them was asked to play Learning Likha for 20 minutes with the use of the earphones for full auditory attention.

Learning Likha is a narrative-centered English-language digital game intended for Filipino learners between 9 to 12 years old. Designed like an interactive storybook, it targets the literacy skill of noting explicit details through reading, listening and viewing. In the tradition of old-style adventure games, it follows the main character Likha, and her friend Taro the tarsier, as they help their bandmates prepare for town festivities by fetching various indigenous Filipino musical instruments from different shops in the town. As the game unfolds, learners are asked to interact with the game by selecting shops and instruments (Figure 1) based on the characters’ instructions. A text box on the upper right corner of the screen contains instructions and important details that may help the player find the correct shop or object. The text box at the bottom contains the dialogue. Each time the learner finds the correct instrument, the learner has an opportunity to play the instrument by tapping on its image within the game. The game is available free of charge from both the Google Play Store and the Apple App Store.



Figure 1. Map scene with the different shops and merchant scene with the instruments.

Learning Likha logs user interactions. It collects the participant’s user ID, the scene or context with which the participant is interacting, the coordinates of screen taps, and the time stamp. Because the participant cannot advance without correctly accomplishing the task of each scene, the participant’s last tap is always the correct move or answer.

2.2 BROMP Observations

Using the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP), a trained observer recorded the participants’ affective states and behaviors while playing Learning Likha using the Human Affect Recording Tool (HART) mobile application. BROMP is a protocol for quantitative field observations of participant behavior and affect (Ocumpaugh et al., 2015). Within BROMP, participants are repeatedly coded individually in a predetermined order. Both affect and behavior are recorded simultaneously, but separately. The affective states of interest observed were engaged concentration, confusion, frustration, and boredom. The descriptions (Ocumpaugh et al., 2015), are as follows:

- Engaged Concentration. This is manifested by focused attention to playing *Learning Likha*. It is shown through leaning closer to their phones and miming as they are reading through the dialogues. Those who were asking their peers what part of the game they were playing as if it were a competition on who finishes first were coded with this affective state.
- Confusion. This is manifested when a participant seems to have a difficulty understanding the game. Indicators include scratching the head, consulting with a peer or facilitator, and statements like “Why is this not working?” or “How do I get pass this?”
- Frustration. This state expresses feelings of distress or annoyance. Behaviors include heavy tapping on the phone screen, hair pulling, deep sighing, and statements such as “This is annoying!”.
- Boredom. This is often expressed as complete disengagement from the game. It is characterized by fidgeting, slouching, yawning, and statements such as “This is boring!” or “Are we done yet?”

The behaviors observed were on-task, on-task conversation, and off-task. The descriptions (Ocumpaugh et al., 2015) are as follows:

- On-Task. This refers to participants focused on completing the game tasks of *Learning Likha*.
- On-Task Conversation. This refers to participants who play while engaging in a conversation with a facilitator or his/her fellow participant. The conversation must be related to *Learning Likha*.
- Off-Task. This refers participants who are doing tasks other than playing the game such as browsing through other applications in the phone or using a different application altogether.

For each test session, at least one observation of affect and behavior was coded for each participant per minute in 5 to 8 second intervals. For instances where the participant seemed to exhibit two or more distinct states during his/her respective observation period, the first state was coded. If a participant finished before the 20-minute allotted time, a “no further activity” state was recorded by coding ‘?’ for affect and ‘\$’ for behavior in the HART application. This notation was agreed upon by the research team in anticipation of those who will finish early.

2.3 Post-Tests

After the participants completed the game, or when the 20-minute allotted play time had elapsed, a 21-item multiple choice post-test was administered to gauge how much they remembered and understood the details from the story. Below is a sample question:

1. What is Taro?
 - a. a monkey
 - b. a dog
 - c. a cat
 - d. a tarsier

The participants were asked to complete one final questionnaire composed of two sections, adapted from the Game-Based Learning (GBL) Engagement Metric (Chew, 2017) and the Intrinsic Motivation Inventory (IMI) (Ryan, 1982). The GBL questionnaire tried to determine how engaged students were with Learning Likha by posing 8 statements and asking participants for their levels of agreement (1 = Strongly Disagree to 5 = Strongly Agree). A sample statement is “I listened carefully to the instructions given to carry out the Learning Likha tasks.” The IMI questionnaire posed 9 statements that asked participants to indicate if the statements were not at all true (1), somewhat true (4) or very true (7). A sample statement is “I enjoyed playing Learning Likha very much.”

All procedures and materials used for this study were reviewed and cleared for implementation by the University Research Ethics Committee of Ateneo de Manila University.

3. Analysis

3.1 Participant Profile

A total of 32 students participated in the study: 12 from Grade 4, 12 from Grade 5, and 8 from Grade 6. Data from one student from Grade 5 and one student from Grade 6 were removed because they were incomplete. The remainder of the analysis will use data from the 30 remaining students only. Of the 30 participants, 19 were male, 11 were female (Table 1). Some students owned cell phones. The majority, especially those from Grades 5 and 6, played cell phone games, some of which were educational.

Table 1.
Demographics of the study participants.

	Grade 4		Grade 5		Grade 6	
	M	F	M	F	M	F
Sex						
N	8	4	7	4	4	3
Age Range	8 - 9		9 - 10		10 - 11	
Had their own cell phones	4 (33%)		10 (91%)		4 (57%)	
Played games on cell phones	5 (42%)		11(100%)		6 (86%)	
Played educational games on cell phones	3 (25%)		7 (64%)		3 (43%)	

When asked about their usage, attitudes and perceptions towards the English language, students said they did not speak English at home (2.7/5.0) or with their friends (2.5/5.0). They did like learning English (4/5) and reading in English (3.7/5.0). They expressed a desire to learn English (3.9/5.0), and agreed that it is important (4/5). In summary, the participants did not typically use English as a medium of communication, but enjoyed learning and were interested in improving their English language skills.

3.2 Incidence of Student Behavior and Affective States

The number of observations per grade level varied: We collected 15 observations per student for Grade 4, 14 for Grade 5, and 21 for Grade 6. To obtain the incidence of student behaviors and affective states, we divided the number of times a student was observed to exhibit a state by the total observations for that student. To obtain the overall incidence of a behavior or affective state, we took the average of the incidences of each state across all students.

Table 2 shows that students were predominantly on-task, working independently (66%) or in conversation with another student (11%). Off-task behavior was relatively rare (1%), though many students finished before the allotted time expired. Table 2 also shows that engaged concentration (65%) was the most often exhibited affective state, followed by confusion (10%). The students did not exhibit boredom often (2%). These findings are consistent with some prior studies that used BROMP observations. As summarized in Baker et al. (2010), the incidence of engaged concentration averaged 60%, confusion averaged 13%, and boredom averaged 5%.

Table 2.
Overall incidence of behaviors and affective states. Standard deviations are in parenthesis.

	Behaviors				Affective States			
	On-Task	On-Task Conver-sation	Off-Task	No Activity	Bored	Confused	Engaged Concen-tration	No Activity
Incidence	66% (16%)	11% (14%)	1% (2%)	22% (17%)	2% (4%)	10% (14%)	65% (18%)	22% (17%)

We used D'Mello's L (D'Mello & Graesser, 2012) as implemented by Karumbaiah (2018) to find out whether some of these states were more likely than chance to transition to each other. An L value of 0 means the transition occurs at chance. L values greater than 0 mean that a transition is more likely than chance to occur, while values less than 0 mean that a transition is less likely than chance.

Because the number of observations per grade level varied, we split the population by grade level and perform the computation on each subset of the observations. We found that engaged concentration and on task behavior tended to persist across all grade levels (Table 3). These findings imply that students attended to the application. When they were focused on it and actively using it, they tended to continue to remain so.

Table 3.
D'Mello's L means. Standard deviations are in parentheses.

Transition	Grade 4	Grade 5	Grade 6
Engaged Concentration to Engaged Concentration	0.28 (0.36), $p = .03$	0.41 (0.28), $p < .01$	0.60 (0.12), $p < .01$
On Task to On Task	0.25 (.30), $p = .03$	0.41 (0.10), $p < .01$	0.61 (0.13), $p < .01$

3.3 Game Engagement

Students self-reported positive feelings and attitudes towards the game. In the GBL questionnaire, they responded that they listened carefully to instructions (4.5/5.0) and tried their best to find solutions (4.2/5.0). They could relate what they were learning to prior knowledge (4.3/5.0) and could see links of the concepts shown to real-world skills (4.1/5.0). They said the game interested them (4.1/5.0), that they looked forward to completing the game (4.4/5.0), and that they were sufficiently challenged (4.1/5.0).

The IMI results were similar. They said that they enjoyed playing the game (6.4/7.0) and that the game was fun (6.4/7.0) and interesting (5.0/7.0). They confessed to being uncertain about how skilled they were in the game (4.8/7.0). Some thought that the game was boring (3.7/7.0).

3.4 Analysis of Affect, Usage, and Comprehension Scores

Despite the high level of engagement and on-task behavior, and despite the positive feedback about the game, comprehension scores were low. Grade 4 students' post-test scores averaged 48%, grade 5 students averaged 65%, while grade 6 students averaged 66%. A single-factor ANOVA showed that differences between groups were significant ($F(2,27)=3.91, p=.03$). A pairwise comparison showed that grade 4 students' post-test scores were marginally significantly different from those of grades 5 ($p=.06$) and 6 ($p=.07$). This implies that the older students understood more than the younger students did. This is something to be expected. Also, there was a very strong significant correlation between on-task conversation and confusion ($r=.91, p<.01$). This may be a consequence of the way that confusion was operationalized, i.e. asking for help or advice from a peer or facilitator. However, it is interesting to note that on-task conversation and confusion both had a moderately strong and significant negative correlation with score ($r=-.42, p.02$ and $r=-.40, p=.03$ respectively). This implies that conversation in this context may indeed be indicative of cognitive challenges. In terms of in-game behavior, the total number of taps had a weak, significant positive correlation with on-task behavior ($r=.37, p=.046$) and a moderately strong, significant correlation with engaged concentration ($r=.41, p=.02$).

4. Conclusions, Limitation, and Further Work

In this paper, we discuss the deployment and testing of Learning Likha, a narrative-based mobile phone game to help with practicing English comprehension. The participants found the it interesting and enjoyable. They professed to exert effort in understanding the contents and completing the game tasks. Observations concur that participants were on-task and engaged. These may be attributed to the contextualized and story-based nature of the game, as well as opportunities to simulate playing the musical instruments. Following the narrative and contributing to the progression of the story by completing game tasks may have enabled such favorable experiences.

Despite these positive self-reports and observations, the test scores were low especially for the younger participants. It is possible that these low scores were an artifact of the their backgrounds. They were not native English speakers. They were interested in learning English, but lacked practice as they did not use English at home or with their friends. One limitation of the study is that the research team did not have a baseline assessment of the participants' pre-existing English language skills. Future work should include an equivalent comprehension pre-test to enable measurement of learning gains.

Nonetheless, this work contributes an exploration of a narrative-driven mobile game specifically designed for Filipino learners. As previously mentioned, there is a dearth in the availability of mobile applications with contextualized education content (Nye, 2015). Learning Likha is an effort to address this barrier by providing young Filipino learners the opportunity to practice English comprehension in a platform accessible to them through a story-based game familiar to their context.

In the broader context, the correlations that we found between in-game behavior and affective states may assist with the development of automatic detectors of feelings and emotions within specific cultural contexts. Because culture influences the way affective states manifest, datasets from diverse populations can increase generalizability of detectors. These detectors contribute to personalization. They can prompt in-game interventions when student behavior indicates negative affective states, or can cue a teacher to approach a student who might be having difficulty.

Acknowledgements

We thank the Ateneo de Manila University, specifically the Ateneo Center for Educational Development, Areté, and the Department of Information Systems and Computer Science; the principals, teachers, volunteers, and learners of our partner public schools for their participation; Ingrid Yvonne Herras, Ma. Rosario Madjos, Monica Moreno, and Jessica O. Sugay; Shamyia Karumbaiah and Jaclyn Ocumpaugh of the Penn Center for Learning Analytics; Anton Andres and Samsung Electronics Philippines, Inc.; the Commission on Higher Education and the British Council for the grant entitled JOLLY: Jokes Online to improve Literacy and Learning digital skills amongst Young people from disadvantaged backgrounds; and John Carlo P. Fernando.

References

- Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
- Chew, B. S. (2017, December). An efficient framework for game-based learning activity. In 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (pp. 147-150). IEEE.
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145-157.
- ITU. (2018). ITU_Key_2005-2018 ICT_data_with_LDCs_rev27Nov2018.xlsx. Accessed 7 May 2019 from <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- Karumbaiah, S., Andres, J. M. A. L., Botelho, A. F., Baker, R. S., & Ocumpaugh, J. (2018). The Implications of a Subtle Difference in the Calculation of Affect Dynamics. In Yang, J. C. et al. (Eds.) (2018). *Proceedings of the 26th International Conference on Computers in Education*. Philippines: Asia-Pacific Society for Computers in Education.
- Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 25(2), 177-203.
- Ocumpaugh, J., Baker, R.S., Rodrigo, M.M.T. (2015) Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.
- Tondeur, J., Forkosh-Baruch, A., Prestridge, S., Albion, P., & Edirisinghe, S. (2016). Responding to Challenges in Teacher Professional Development for ICT Integration in Education. *Educational Technology & Society*, 19 (3), 110-120.
- Sung, Y. T., Chang, K. E., & Yang, J. M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational research review*, 16, 68-84.
- Wolfenden, F., Auckloo, P., Buckler, A. & Cullen, J. (2017). Teacher educators and OER in East Africa: Interrogating pedagogic change. In C. Hodgkinson-Williams & P. B. Arinto (Eds.), *Adoption and impact of OER in the Global South* (pp. 251-286). Retrieved from <https://doi.org/10.5281/zenodo.600424>

Core Gamification of Learning Activities through a Method based on Information Structure Manipulation

Pedro Gabriel Fonteles FURTADO^{a*}, Tsukasa HIRASHIMA^a & Yusuke HAYASHI^a

^a*Hiroshima University, Japan*

*gabrielpff2@gmail.com

Abstract: Gamification, the use of game elements in non-game contexts, is a popular method to design motivating, enjoyable learning experiences. It is often applied through external rewards like leaderboards and achievements. Research shows that these rewards can negatively affect intrinsic motivation. One way to avoid this issue is by applying the game elements to the internal design of the learning activity. This approach is called Core Gamification. However, there are no concrete methods to apply Core Gamification to learning activities. This paper presents a method for the Core Gamification of learning activities based on information structure manipulation. In this case, information structure manipulation is not done by the code. And they are also not just about the design of the activity. The structures are visible to the user. The manipulation is also done by the user. Case studies with verified learning effects is also shortly presented.

Keywords: Gamification, Gameful Design, Intrinsic Motivation

1. Introduction

Gamification has been associated with rewards. Most gamification approaches focus on a variation of leaderboards, points, or achievements. The goal of these rewards is to change behavior. Since these rewards are external to the activity, they can be interpreted as extrinsic motivation (Nicholson, 2015). Extrinsic motivation can be detrimental to intrinsic motivation (Deci and Ryan, 2002). Research shows that after the introduction of rewards, users become less likely to do the task when the reward is removed (Brandt, 1995).

A review of empirical studies by Zeng and Shang (2018) investigated studies published between 2013 and 2017. It states that there is a concern on the cognitive processes in the learning process. The review pointed out the work of Lee and Heeter (2017), which drew relationships between game behavior, skills, attention and comprehension. These types of work show a trend to go beyond rewards into the cognitive process that happen during gameplay. However, how to design experiences which optimize these cognitive processes? And at the same time, keeping the activities engaging.

The work of Deterding (2015) approaches this problem by defining a method to create gameful systems focused on intrinsic motivation. Gameful systems are systems which have gameful (game-like) characteristic. The method is grounded in designing systems through matching skills, motivations, and actions. The method is very broad, being usable in a wide variety of scenarios, attempting to influence experience-driven design in a general way. The work Yuan and Shang (2018) also presented a framework based on emotional design for game development. It can be used to improve or serve as guidelines for design, but it does not offer a more concrete method for designing gameplay.

While the work of Deterding does offer a general view and guidelines, there is a lack of more concrete, context-specific methods of achieving Gameful Design. When designing educational applications, we want to focus on the need for competence (Ryan and Deci, 2000). That is, we want to align the intrinsic motivation of mastering the learning target with the need of the users for competence. External rewards and systems would detract from this. To achieve this, game characteristics must not be added externally, but internally. The work of Wang (2016) defines this sort of internal gamification as Core Gamification, where the game elements are so integrated into the core mechanics that they are inseparable.

If we remove leaderboards and achievements from the discussion, what elements from games should be used in Core Gamification? The literature points out some elements like clear actions, rules, goals, and feedback (Deterding, 2015). Those are the elements that games use to afford experiences that satisfy competence needs. Furthermore, clear goals, sense of control, unambiguous feedback are all associated with the flow state (Jackson and Marsh, 1996), a state of high involvement and enjoyment. Core Gamification should focus on intrinsic motivation based on competence needs while promoting the flow state. It can manage to do so by clarifying and redefining the actions, rules, goals and feedback mechanics of the activity.

This paper proposes a method for Core Gamification based on defining information structures derived from the learning target and designing gameplay around them. It also includes the case study of applications developed using this method.

2. Core Gamification, Serious Games and Our Approach

The work of Deterding (2015) defines gamification as the means of using game design elements in non-game contexts. In this section, reward-based gamification is defined as the gamification approaches focused on leaderboards, points and achievements as a way to reward the player for their actions. Core gamification, in contrast, refers to the gamification approaches that focus on the activity design internally. While gamification is the means to introduce game design to non-game contexts, nothing is said about the result. The result might be a gameful system, a system that involves gameful design that is not necessarily a game. When gamification is applied to software development, the result can sometimes be defined as a serious game. A serious game is defined in the work of Susi et al. (2007) as a game made for more than entertainment. Whether or not the result is a serious game or not depends on the used definition of game and on the design of the activity. As such, the use of reward-based gamification or core gamification is not what determines if the result is a game or not. Both methods can be used in the design of both serious games and gameful systems.

The proposed approach in this paper aims for the result to be a serious game. As such, it is intended for it to be a method to design the core gameplay of a serious game. The relationship between gamification, serious games and gameful systems can be seen in Figure 1.

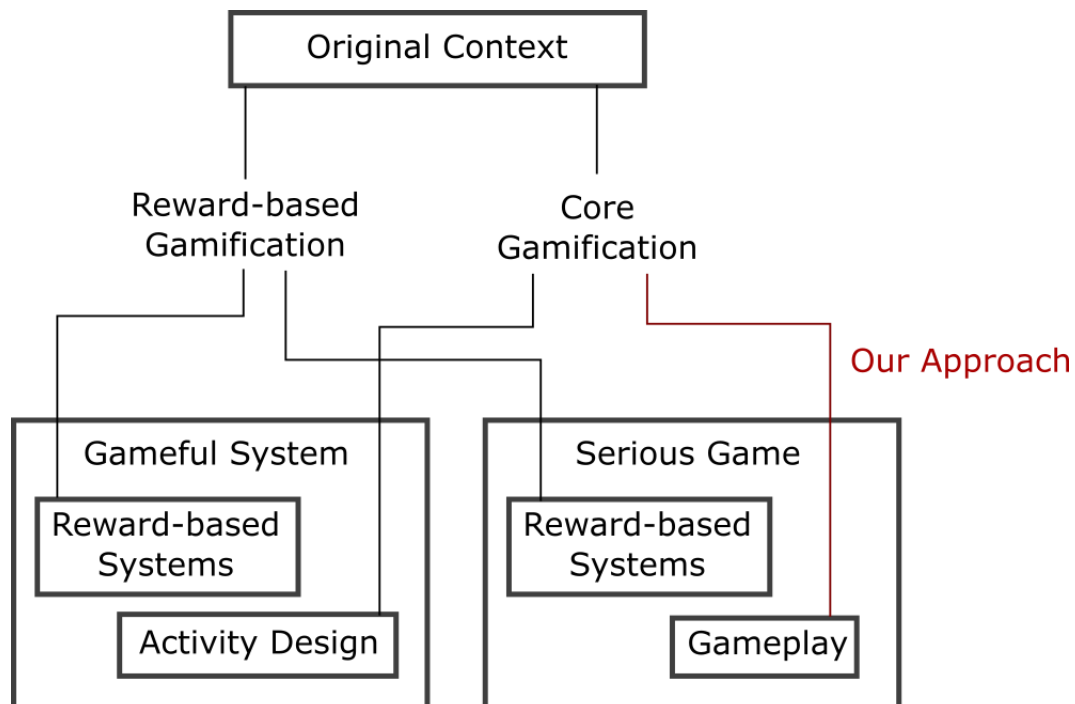


Figure 1. Diagram relating serious games, gamification approaches and the approach in this paper.

3. Proposed Method

Below is an outline of the proposed method:

- Original task analysis
 - Select subset of the original task
 - Define broad structure
 - Decompose structure
 - Define constraints
- Gameplay Design
 - Define task
 - Define goal
 - Define problem-specific constraints
 - Define end states
 - Define state space
 - Define initial state
 - Define feedback

3.1 Original Task Analysis

In this step, the original task will be analyzed so the appropriate information for gamification can be obtained. Firstly, a subset of the task will be chosen. A subset is chosen because classical learning tasks often contain parts which are not directly related to learning the target. Then, an information structure will be defined based on the learning content and on the subset of the original task. This information structure must be well chosen because the entirety of the gameplay is defined based on it. Next, we decompose that structure into pieces. These pieces are what the player can interact with. Finally, we define constraints for how the pieces can be combined. Those constraints are based on the properties of the information structure. They are closely related to the content that must be learned.

3.2 Gameplay Design

In this step, the rules of the game will be defined. While an order has been proposed for the sub-steps, the designer is likely to go back and forth during the process. First, a goal is defined. The goal represents what the player should strive for. Then, possible problem specific constraints can be defined. This allows for the creation of problems focused on a single aspect of the activity. End states are then defined. The end state defines one or more application states where the problem is considered cleared. The end states are closely related to the goal and to the constraints. Then, the state space is defined. The state space is all possible combinations of states for a given problem. Next, the initial state is defined. This is the application state that is presented at the start of a problem.

4. Case Studies

4.1 Monsakun

Monsakun is a sentence integration software for arithmetic problems (Hirashima et al., 2007) which was developed through the presented method. In Monsakun, users pose problems by choosing and ordering three sentence cards. The use of Monsakun has been associated with an increase in correctness during sentence integration. One experiment found a significant increase ($p < 0.001$, $\eta^2 = 0.410$) between pre-test scores ($M = 1.33$, $SD = 0.79$) and post-test scores ($M = 3.33$, $SD = 0.70$), where the tests measured correctness of sentence integration problems.

The design of Monsakun is framed and explained through the proposed procedure:

- Original task analysis

- Select subset of the original task: The subset we use in Monsakun is only the integration portion of understanding as described by Mayer et al. (1985);
- Define broad structure: The main structure in Monsakun is the structure of a single problem. A problem is composed of multiple sentences such that each sentence can be mapped to a number in a mathematical expression. An increase type problem would be described in three sentences as;
 - ✧ John has three apples;
 - ✧ John buys two apples;
 - ✧ John has five apples.
- Decompose structure: A piece is defined as a single sentence which has attributes like value, referenced objects, type, etc.;
- Define constraints: Most of the constraints are intrinsic to the original problem. One extrinsic constraint is that all problems are defined by a maximum of three sentences. Another extrinsic constraint is that all problems describe additions or subtractions. Example of intrinsic constraints in Monsakun:
 - ✧ Object constraints: All sentences only reference one object and they all reference the same object;
 - ✧ Number constraints: The third sentence contains a value that is the sum of the values of the two previous sentences;
 - ✧ Sentence type constraints: The first sentence and the third sentence must describe the existence of an object. The second sentence must describe an object's value increasing.
- **Gameplay design**
 - Define goals: The overarching goal of Monsakun use is the understanding of arithmetic problems. The goal of a single task is for the student to construct a problem satisfying several constraints;
 - Define task: In a Monsakun task, the user must choose three pieces from the available pool and order them so that they satisfy the constraints;
 - Define task-specific constraints: Other than the constraints previously defined, each task in Monsakun provides the user with a mathematical expression and a story type. This results in two task-specific constraints:
 - ✧ The three sentences must be mappable to the provided mathematical expression;
 - ✧ The story type of the resulting problem must be the same as the provided story type.
 - Define end states: The end state is when the three correct phrases are in the correct order;
 - Define state space: Tasks in Monsakun provide six sentences for the user. This means that three of the sentences are dummy sentences that won't be used. The dummy sentences are designed in a way that they violate their use will result in the violation of a specific constraint. The motivation for this is to make students must think about specific constraints to solve the problems. Another motivation is to be able to gather data on what constraints give student trouble. Given the six sentences, the search space for a Monsakun task contains 120 possible combinations.
 - Define initial state: All six sentences are available, and none are chosen;
 - Define feedback: Message telling users if their problem is correct or incorrect. Constraint-based feedback is also an option.

More information on the underlying structure of Monsakun can be found in the work of Hirashima et al. (2014).

4.2 Kit-build

Kit-build is a closed concept map building tool which was developed through the presented method. Further information on the design of Kit-build can be seen in the works of Hirashima et al. (2015). In Kit-build, learners build maps from provided pieces. The use of Kit-build has been associated with improved reading comprehension and with improved retention for contention. For example, the work of Alkhateeb et al. (2015) found that Kit-build users showed better performance ($M=79.06$, $SD=12.08$) in foreign language delayed comprehension test than traditional concept mapping users ($M=66.12$, $SD=20.26$) in an ANOVA test, $p < 0.01$.

- Original task analysis
 - Select subset of the original task: The subset of concept map construction we use in Kit-build is only making the connections using the links and nodes, without inputting the labels;
 - Define broad structure: The main structure in Kit-build is a concept map defined by the teacher. The map is made up of links and concepts.
 - Decompose structure: Maps are made up of multiple links and concepts. Links and concepts are treated as separate pieces which are not interchangeable.
 - Define constraints: The constraints in Kit-build are defined by the expert map designed by teachers. The constraints are the connections between the nodes and links;
- Gameplay design
 - Define goals: The overarching goal of Kit-build is to improve understanding of a text or class;
 - Define task: In a Kit-build task, users are tasked with associating concepts through links;
 - Define task-specific constraints: In Kit-build a node can only be associated to another node through a link. Furthermore, each connection in the expert map is a constraint of correctness. Violating these constraints would reduce the score of the map;
 - Define end states: There are two possible end states in Kit-build. One of them would be the student building the expert map. The other end state is the student building a map he is satisfied with. In the second case, even a submitted empty map would be a possible end state.;
 - Define state space: Assuming N nodes and L links, each link can connect two nodes. They can also be connected to no nodes at all. As such, the following equation defines the number of states in the space: $(N + 1)2L$
 - Define initial state: No links and no nodes are associated at first;
 - Define feedback: One automated feedback approach of Kit-build is based on showing the differences between the user map and the expert map back to the user, as in the work of Furtado et al. (2018).

5. Discussion and Conclusion

More research is necessary in gamification beyond reward-based systems. Core gamification refers to gamification methods that are more concerned with the internal design of the transformed activities. However, there is a lack of concrete method on how to perform core gamification of learning software. To address this, this work proposes a method of Core Gamification based on students manipulating information structures. Our proposed method addresses Core Gamification needs in the following ways:

- Actions
 - Actions serve to manipulate the information structure. They should be made clear by the interface design, employing intuitive and easy to use mechanics, such as drag-and-drop. Users will build or modify structures by using sets of pieces that make up the information structure. As actions are primarily used to modify these structures, every action is relevant to the learning target;
- Rules
 - The rules of gameplay are based on the intrinsic rules of the learning target, but clearly defined in terms of the information structure. As such, the rules of the gameplay are highly related to the learning target. This means that becoming a better player should imply in becoming a better learner. The large state space of the activities serves to discourage other learning behaviors that do not contribute to learning;
- Goals
 - The goal of the original activity is replaced with the goal of satisfying clearly defined constraints. These constraints are defined in terms of the information structure. They need to require understanding of the information structure to be solved. The constraints should be clearly specified to the players, so that their effort in solving the activities can go towards comprehending the information structure;
- Feedback

- Feedback is designed based on the constraints. A given state either violates or satisfies a constraint. As such, this information can be clearly shown to the player in the form of “correct” or “incorrect”. More complex feedback can be given by relating the information structure to reading material or by giving more detailed information on which constraints were violated and which were satisfied.

Evidence of the learning effect of applications that use this method has also been presented. This study presents, as far as researched, the first concrete method for designing gameplay of gamified applications.

For future work, the method will be expanded to include clearer interface recommendations. It is also necessary to quantify the effects of the applications regarding motivational metrics.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 17H01839 and 15H02931.

References

- Alkhateeb, M., Hayashi, Y., Rajab, T., and Hirashima, T. (2015). Comparison between kit-build and scratch-build concept mapping methods in supporting efl reading comprehension. *The Journal of Information and Systems in Education*, 14(1):13–27.
- Brandt, R. (1995). Punished by rewards. *Educational Leadership*, 53(1):13–16.
- Deci, E. L. and Ryan, R. M. (2002). *Handbook of self-determination research*. University Rochester Press.
- Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. *Human– Computer Interaction*, 30(3-4):294–335.
- Furtado, P. G. F., Hirashima, T., and Hayashi, Y. (2018). Reducing cognitive load during closed concept map construction and consequences on reading comprehension and retention. *IEEE Transactions on Learning Technologies*.
- Hirashima, T., Yamamoto, S., and Hayashi, Y. (2014). Triplet structure model of arithmetical word problems for learning by problem-posing. In *International Conference on Human Interface and the Management of Information*, pages 42–50. Springer.
- Hirashima, T., Yamasaki, K., Fukuda, H., and Funaoi, H. (2015). Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, 10(1):17.
- Hirashima, T., Yokoyama, T., Okamoto, M., and Takeuchi, A. (2007). Learning by problemposing as sentence-integration and experimental use. In *AIED*, volume 2007, pages 254–261.
- Horiguchi, T., Imai, I., Toumoto, T., and Hirashima, T. (2014). Error-based simulation for error-awareness in learning mechanics: An evaluation. *Journal of Educational Technology & Society*, 17(3):1–13.
- Jackson, S. A. and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of sport and exercise psychology*, 18(1):17–35.
- Lee, Y.-H. and Heeter, C. (2017). The effects of cognitive capacity and gaming expertise on attention and comprehension. *Journal of Computer Assisted Learning*, 33(5):473–485.
- Mayer, R., Sternberg, R., and Freeman, W. (1985). Human abilities: An informationprocessing approach.
- Nicholson, S. (2015). A recipe for meaningful gamification. In *Gamification in education and business*, pages 1–20. Springer.
- Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68.
- Susi, T., Johannesson, M., and Backlund, P. (2007). Serious games: An overview.
- Wang, G. (2016). Game design for expressive mobile music. In *New Interfaces for Musical Expression*.
- Yuan, M. and Shang, J. (2018). Values and design strategies of emotional design in educational games. *Proc. of ICCE2018*, pages 583–588.
- Zeng, J. and Shang, J. (2018). A review of empirical studies on educational games: 20132017. *Proc. of ICCE2018*, pages 533–542.

A Usability Study on Shape Shape Hooray: an Adaptive Educational Game Associating 3D Geometric Shapes to Daily Objects

Winna Mia Victoria D. BUENVIAJE^{a*}, Ma. Anniela B. DELA CRUZ^a & Ingrid Marie Therese P. FADRIQUELA^a

^a*Graduate Studies, Ateneo de Manila University, Philippines*

*winna.buenviaje@obf.ateneo.edu

Abstract: Situated learning theory argues that learning is embedded within an activity, context, and culture. It posits that students are more likely to learn if they have an exposure to the authentic context of the learning environment. Based loosely on this theory, Shape Shape Hooray is an adaptive educational game that aims to teach basic 3D geometric shapes by allowing basic education students associate 3D shapes to daily objects. As an adaptive game, this paper discusses the paths developed for different kinds of players (no prior/low prior, average, and high prior knowledge). A usability test was conducted to which a generally positive score was acquired. Various kinds of metrics (task success and task indicators) and the Systems Usability Scale (SUS) were also tested to find out the game's impact on user experience and to evaluate possible design directions and improvements for Shape Shape Hooray. We found that respondents were highly impressed with the game's usability, scoring 81.5 on the SUS. Future work will include educators' testing on the game's usability, testing of learning to students, and consultations with programmers for expert validation.

Keywords: adaptive learning, educational game, stealth assessment, situated learning theory

1. Introduction

In the discussion of shape recognition and properties as a skill by Virginia Kindergarten Readiness Program, it is stated that learning shapes is an essential building block of a child's early stages of development which makes it very important as children tend to notice objects by shapes rather than by color. A learner who is able to understand shapes specifically and accurately can connect with the objects around them, and use them more effectively in the context of their environment.

Situated learning means creating meaning from the real activities of daily living (Stein, 1998). It posits that learning is embedded within an activity, context, and culture (Lave, 1989). Experiments show that situated learning occurs best when done in the context of authentic tasks (Oppermann, 2006). This allows for an increase in learning effectiveness, efficiency, and satisfaction as well as the replicability of learning results.

Shape Shape Hooray is an educational game designed for students in grades 1-2 of the Philippine K-12 curriculum. Players drag digital 3D shapes into containers and the game will issue correctness feedback. When players are able to correctly drag chosen shapes to their proper containers, the objects would rotate to show and demonstrate their dimensions in 3D space. In addition to this mechanic, it was designed to be adaptive such that students have control over the pace with which they interact with in-game objects and finish levels. Features to engage active participation as well as goal-directed storylines were included. As the narrative of the game unfolds, student users are afforded the capacity to experience local culture by interacting with digital representations of their local vicinity embedded within the game. The game covers basic 3D geometric shapes i.e. cube, cylinder, cone and sphere. A drag-and-drop mechanism was implemented.

This paper presents a usability study as well as metric tests to see the games impact on user learning experience (R1). Testing was also conducted to uncover improvements that may be implemented in future work to enhance user learning experience (R2).

2. Shape Shape Hooray

2.1 Storyline

The importance of narrative in game-based learning is that it provides a cognitive framework for problem-solving (De Freitas, 2011). In Shape Shape Hooray, the game depicts a story of how the character follows the story of a student life, or specifically the last day of school. When a player chooses a character, the character will explain how from preparing from home, he or she heads to school, and then eats after a long day, and then eventually goes for vacation. This story is the framework of Shape Shape Hooray, making the locations the levels of the game itself.

This storyline is a derivative of common Philippine living. Living in a tropical country, students and their families head to the beach during the school vacation. This idea was brought to life in the game so that students may be able to relate it in their lives embedding the situated learning theory by creating meaning from real activities of daily living.

2.2 Adaptive Paths

Shape Shape Hooray would adapt based on the player's learning progress before playing and during the game. Before the actual game starts, the game will test the students of the following: 1) knowledge of the shapes, 2) spelling of the shapes, 3) association of shapes in real life objects in order to determine whether a player's knowledge about the shapes is high, average or no prior knowledge at all.

A total of ten questions are given and the player will be directed to three different paths depending on the scores they acquire. The player needs to answer all of the questions and that is the only time the scores will be generated. Pretest scores will be the key indicator as to which path the players will be led to as a starting point. Generally, the target users of Shape Shape Hooray are students with low or no prior knowledge of the shapes.

Classification of players will depend on the paths as follows:

1. High prior knowledge path - for players who gets a score of 8-10 points in the pretest. Game will automatically set on a hard level with all the shapes possible to be learned in the game. Confusing and more distracting objects are also added.
2. Low prior knowledge path - for players who get a score of 5-7 points in the pretest. Players on this path will be on the average level of difficulty. Designated shapes to learn per location in the map will be followed. Cube—Bedroom (Level 1) (See Figure 1), Cylinder—Classroom (Level 2), Cone—Diner (Level 3), and Sphere—Beach (Level 4).
3. No prior knowledge path - for players whose scores ranges from 0-4 points, manifesting that the player has no background knowledge at all. This will be on easy mode and players will be familiarized with a 3D shape following rote learning or memorization through repetition, showing only the shapes needed for a corresponding level.



Figure 32. Level 1: Bedroom (Cube) sample.

Once the game determined which path the player takes, the map will appear for the player to proceed in the first level. In each location, the player should find the specific shapes needed to surpass a certain level. When an object is clicked and dragged to the backpack, the object will zoom in and rotate to emphasize the object's 3D shape for the visualization of the player.

2.3 Assessment

Some features of the game are designed in order to evaluate the player's learning progress:

1. Bonus rounds throughout the game will serve as the stealth assessment—a quiet, powerful process by which learner performance data are continuously gathered during the course of playing (Shute et al., 2009). A bonus round (See Figure 2) will appear after 2 levels and players of Shape Shape Hooray will be asked to identify shapes of real life objects presented. This round is to test whether the player is able to associate the shapes the game is trying to teach in real life objects. If the player shows an improvement, the game would adapt and proceed to the higher or lower path from originally the player was based on the score he got on the bonus round.
2. The character's energy in the form of a battery will determine whether the player is using the trial and error method or merely guessing. Players are only allowed to make mistakes four times or else the game resets on the particular level it was on. This is to ensure that the players are still trying to play in terms of the objectives of the game.
3. Hints are in the game in order to give an idea to the players who are unsure of what the shapes look like. Every time a player clicks on the hint button, it will show an instruction to look for the objects that look like the shape the level is trying to teach. To give a general idea, an outline of the shape will be shown.

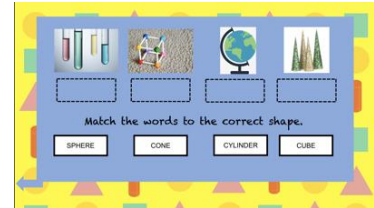


Figure 33. Bonus round sample.

3. Methodology and Limitations

In this section, target participants, metrics used for usability, platform of the prototype, testing, limitations and methods used in conducting the study will be discussed thoroughly.

3.1 Target Participants and Testing

Shape Shape Hooray is intended for children with ages from 7-8 years old or Grades 1-2 students. Since the game is adaptive and has different paths set to their prior knowledge, the main focus in this study are those players with low to no prior knowledge regarding the basic 3D shapes. The game is ideally to be conducted in schools to test both its usability and learning, but due to time constraints, the researchers conducted the testing with



Figure 34. Prototype

students individually, meeting them up at their preferred time and place.

An animated version of the prototype was made using an existing software application. Since the focus of the research is Shape Shape Hooray's usability to its potential respondents and not its effectiveness, we gathered respondents with ages 5-8 years old due to time constraints of the approval to test in schools.

The testing started by presenting the game to the players and letting them play without the intervention until asked. After the pretest, the results will determine which path of the game will the respondent be on and shall continue playing the game until after the posttest. The researchers, on the other hand, observed the players and took down notes through data collection sheets with regards to their behavior and comments during testing through the use of Task Success and Tasks Indicator Metrics. After finishing the game, they were asked to answer the SUS and a debriefing interview.

3.2 Limitations

- a. *Participants* - The researchers ended up testing 5-8 years old players since the researchers were not able to gather a great number of participants that will fully represent the age range of 7-8 years old. Moreover, the usability of the game, and not its learning outcome, was to be tested so it was possible for Shape Shape Hooray to be played by children within that age range.

- b. *Time Constraint* – Researchers have always intended to conduct the study to more students and players but if were given a sufficient amount of time, teachers can also be asked to play the game and gather feedback from. Getting an approval to test 7-8 years old children in schools has also been a limitation to the researchers as the schools are mandated by law to practice the data privacy act.
- c. *Limited Programming Knowledge and Equipment* - The researchers wish to disclose that the team does not have enough background in advanced programming. As a result, an animated prototype using a computer was used. This may have also affected players’ perception about the game features were not as robust in production.

4. Results, Analysis of Data, and Design Mock-Ups

4.1 Task success

To measure the success rate of three major tasks the game is asking to do which are: 1) locating objects, 2) using hints, and 3) map navigation, a task success metric was created in order to determine the easiest and hardest task for the players. Looking at Table 1, most players were able to complete tasks successfully without the need of assistance from the conductors.

- Locating objects is the easiest task of the game. All of the players were able to locate the objects needed at a certain level but there were two players who needed assistance to successfully locate other objects.
- Asking for help with the use of hint option was not totally utilized as some players did not see the need for help during the game. However, two of them asked for guidance on how to use the hint and one was able to discover on his/her own.
- Two of the players were aware of the narrative that the game is trying to tell. Although, one player was not given the chance to navigate the map because he was directed to the high-prior path where there is no need to explore the four different places and levels unlike those who were directed to the average and no-prior path.

4.2 Task Indicators

The researchers used a Task Indicators Metric containing more specific tasks that players may be able to do. Composing a name, answering the pretest and posttest questions, navigating the map, using hints and clicking all the objects needed per level are some of the tasks indicated in the metric. Most of the tasks were fully accomplished and utilized except the use of hint button.

It was observed that players found it easy to compose a name, answer the posttest, locate sphere and cube objects without seeing any assistance. On the other hand, the use of the hint button, although helpful, was not utilized.

4.3 Debriefing

After the game testing and posttest questions, a debriefing interview was conducted to the players. When asked if the instructions were clear and understandable, all of them agreed and said that the game was very easy to understand and the instructions were clear and presented in an uncomplicated manner. Generally, the players did not have much of a hard time moving from one location to another. However, not all of them were aware of the narrative of the game. Some students wanted to click and go to the beach first as maybe because the game was conducted during their summer break and was set in their mind that they were actually going to the beach. When asked if the game was confusing, most players said that it is not, but some objects in certain levels were confusing because of the similarity in shapes which results to giving second thoughts on whether they think that it is the right answer or not. According to the players, the illustrations are very colourful and attractive and commented that the

Player	Locating Objects	Hints	Map Navigation
P1	1	1	1
P2	2	2	2
P3	2	1	2
P4	1	0	2
P5	2	0	0

Indicators: 2- User was able to perform the task; 1- Needed Guidance; 0- Not able to perform the task

Figure 35. Average of raw data for task success.

avatar cheering them up made the game more fun and encouraging. However, one player said that the voiceover of the avatar was scary.

Aside from conducting the interview, players were asked what are the shapes of physical things that were found around the house. For example, a tumbler or a box of tissue was raised and asked the player what shape it embodies and most of the players answered correctly despite not having prior knowledge on 3D shapes before playing the game. This shows that players learned and can already identify the said shapes after playing Shape Shape Hooray.

4.4 Systems Usability Scale (SUS)

SUS is a Likert scale (Brooke, 1996) which is done after the player made use of the system. System Usability Scale (SUS) contains ten questions about user experience of Shape Shape Hooray and the players will have to rate their agreement or disagreement on a 5-point scale, 1 being the lowest. This is to determine what the users think of the system's design and ease of use and data will also be used as a basis for future improvements of the game.

Indicator	S1	S2	S3	S4	S5	Overall Average
Shape Shape Hooray is easy to use.	4	4	4	4	4	
Shape Shape Hooray is easy to navigate.	3	3	1	4	1	
Instructions were clear and understandable.	4	2	4	4	4	
The drawing and graphics are presentable and appropriate.	2	4	3	4	4	
The music and sounds of the game are engaging.	4	3	2	4	1	
There were some parts of the game that is difficult to understand.	4	2	4	3	4	
I needed guidance in playing the game.	4	3	4	4	4	
I was able to learn about geometric shapes	4	2	4	4	2	
I needed to learn a lot before playing the game.	4	4	4	4	4	
I would play Shape Shape Hooray to learn and have fun at the same time.	2	1	1	2	4	
Average	87.5	70	77.5	92.5	80	81.5
Remarks	Excellent	Good	Good	Excellent	Good	Excellent

Figure 36. Systems Usability Scale for SSH based on John Brooke.

To determine the score of the SUS, first, the sum of the score per item needs to be computed and for the odd number questions (1,3,5,7,9), 1 should be subtracted from the score and as for the even number questions (2,4,6,8,10), 5 should be subtracted from the score. Afterwards, the sum of the scores should be multiplied by 2.5 to get the total value of the System Usability.

Since Shape Shape Hooray received good feedback from the players and got a total average of 81.5 that falls under an excellent-good category. It was found out that it is easy to navigate and really helpful if you want to know about geometric shapes. The instructions given were not difficult to understand and they liked it that if they would be given a chance, they will play it again to learn about shapes and have fun at the same time.

5. Conclusion and Future Work

This paper presented the usability study of an adaptive educational game that teaches 3D geometric shapes, Shape Shape Hooray. In this work, impact or impression of the target users in the experience of the game(R1) and possible improvements for the game(R2) are answered.

Based on the results of the systems usability scale, respondents were highly impressed with the game's usability. Conducting the SUS revealed that the game falls under the excellent-good category(R1). However, this is only indicative of its usability and not conclusive of the learning which is the main goal of the development of Shape Shape Hooray.

Through debriefing, respondents gave recommendations such as: varying hints, indicators not only on audio but also on the visuals, and using age-appropriate and phase of learning appropriate graphics used(R2). The group who conducted the study filtered these enhancements based on the frequency of comments done when the prototype testing was happening. Early projected design mock-ups were also created in order to implement it on the game to enhance both the gaming and learning experience.

Although the development of the game is still in the works and continuously progresses, the feedback that were collected in the study conducted were duly noted for the enhancement of this educational game. Educators' feedback on the game's usability, implementation of feedback, and testing of learning to students in the association of the shapes in real life objects would be conducted for further study. Further research and consultations with programmers and developers will also be part of the future work to consult expert validity of the game.

Acknowledgements

We would like to thank our faculty advisers for guiding us throughout this experience. We also thank the people who supported and helped us in our prototype making and testing.

References

- Bangor, Kortum and Miller. (n.d.). *Determining what individual SUS scores mean : Adding an adjective rating scale.* Retrieved from <https://www.trymyui.com/sus-system-usability-scale?fbclid=IwAR3hvXDycFEhDihB0mDla9cujep0POavqSGuxsDQqfdPE18HQQVW-qgwnrI>
- Brooke, John, (1996). *SUS - A quick and dirty usability scale.* United Kingdom : Redhatch Consulting Ltd.,. Retrieved from <https://hell.meiert.org/core/pdf/sus.pdf>
- Chapron, B. (n.d.). *The Importance of learning shapes in preschool.* Retrieved from <http://www.cedarwoodschool.com/the-importance-of-learning-shapes-in-preschool/>
- David, L. (2017). “*Situated Learning Theory (Lave)*” in *Learning Theories* [web log]. Retrieved from <http://www.learning-theories.com/situated-learning-theory-lave.html>
- De Freitas, S. (2011). *Digital games and learning.* London, UK: Continuum International Publishing Group.
- Druin, A. (1999). *The Design of children’s technology.* San Francisco, CA: M. Kauffman Publication.
- Lave, J. (1989). *Cognition in practice: mind, mathematics, and culture in everyday life.* Cambridge, UK: Cambridge University Press.
- Mensa Education and Research Foundation. *Lesson Plan: Shapes.* Retrieved from http://www.mensaforkids.org/MFK2/assets/File/Teach/LessonPlans/Lesson_Shapes.pdf
- Northern Illinois University, Faculty Development and Design Center. *Situated Learning.* Retrieved from www.niu.edu/facdev/_pdf/guide/strategies/situated_learning.pdf?fbclid=IwAR1jw3tPPGtjzRfckcfybwTn9C5
- Oppermann, R. (2006). Situated learning in the process of work. In D. Hung & M. S. Khine (Eds.), *Engaged learning with emerging technologies* (pp.69-89). Dordrecht: Springer Netherlands.
- Stein, D. (1998). *Situated learning in adult education.* *ERIC Digest.* Retrieved from <https://www.ericdigests.org/1998-3/adult-education.html>
- Tobias, S. (2011). *Computer games and instruction.* Charlotte, NC: Information Age Publishing, Inc.
- Verdine, B. N., Lucca, K.R., Golinkoff, R.M., Hirsh-Pasek, K., & Newcombe, N.S. (2016). The Shape of Things: The Origin of Young Children’s Knowledge of the Names and Properties of Geometric Forms. *Journal of cognition and development: official journal of the Cognitive Development Society*, 17(1), 142-161. Doi:10.1080/15248372.2015.1016610

Integrating Computational Thinking with Digital Storytelling to Enhancing Expression Ability

Chih-Kai CHANG^{a*} & Xin-Yu Dai^b

^a*Department of Information and Learning Technology, National University of Tainan, Taiwan*

^b*Department of Information and Learning Technology, National University of Tainan, Taiwan*

*chihkai@mail.nutn.edu.tw

Abstract: Cultivating computational thinking has become an important key to the curriculum of science and technology learning area. When students use interactive media, they can no longer be just receiver, but use interactive media to self-expression. We will let students in grades 3-4 learn ScratchJr, because most students in grades 5-6 learn Scratch at Tainan city. The student use ScratchJr to create the animations, stories and games which interest them. We use creation as a motivation, and children can learn how to solve problems by creating their projects over and over. Children can cultivate their own computational thinking by this way. Digital storytelling can enhance cognitive ability, and make things specific and clear. This is a good way to make everyone to understand the abstract knowledge. Let students use ScratchJr to make digital storytelling on the iPad mini in order to cultivate students' computing thinking. The main goal of this thesis is to improve students' expressing ability to perspectives in the three-dimensional framework of computational thinking. Twenty-seven students in third grade participated in this study. They used ScratchJr to make digital storytelling on the iPad mini. The researcher designed a series of training curriculum. All training curriculum were taught by researcher. Students completed three projects in the basic, intermediate and advanced three-phase training curriculum. Students present and save their projects by the way of YouTube in this thesis. Teaching students to use the iPad mini for video recording, and providing students the interview questions. Arrange students to work in groups of two, use iPad mini to interviews with each other, and then record the interview process. In the end, the qualitative analysis method was used to analyze the students' projects after the three-phase training curriculum and the video of the interview. It will explore the objective of this study: Enhance students' expression ability by the integration of computational thinking into digital storytelling teaching.

Keywords: Science and technology learning area, computational thinking, ScratchJr, digital storytelling, expression ability.

1. Introduction

1.1 Learning by making games

Digital game-making activities can engage students to enhance computational thinking skills. Many findings of relevant studies are very positive (Clark, Tanner-Smith & Killingsworth, 2016; Hayes & Games, 2008; Kafai, 2012; Kafai & Burke, 2015). However, the process of creating a digital game is an ill-structured problem. Many students seem to face difficulties in higher-order thinking skills such as problem-solving, computational thinking, communication, and cooperation. During the last decade, a lot of tools have been developed to scaffold students rapidly build their games such as Scratch, Kodu, GameMaker, and AppInventor. Those tools or programming language did help students and teacher overcome initial development issues. Unfortunately, students tend to spend too much time and energy in becoming competent in building games using specific authoring tools. As a result, students do not have enough opportunity to develop a systematic perspective about software development life cycle and practice other higher-order skills, such as communication and collaboration.

In this paper, a pair programming curriculum is presented to promote the acquisition of multiple higher-order thinking skills at the same time via digital storytelling activities. In other words, students

in a pair can systematically design digital storytelling projects with the goals of increasing their communication, collaboration, motivation and improving their computational thinking skills. The pair programming approach of game making not only can enhance students' social skills (communication and collaboration), but also can match the requirements of digital storytelling. This pair programming approach seems not only promote computational thinking, but also very appropriate for digital storytelling activities (Chang, Tsai & Chin, 2017).

1.2 Digital Storytelling

Digital storytelling is a short form of digital media that allows people to share aspects of their stories. Individuals can use it to tell a story or an idea. The most important characteristic of digital stories is that they no longer conform to the traditional narrative mode, because they can combine static images, moving images, sound and text, as well as non-linear and interactive features. The expressive power of technology provides a broad base for integration. It enhances the experience of the author and the audience and allows for greater interaction. Education workers generally believe that the advantage of digital storytelling is that students can choose their creative expression through a range of technical tools. Learners begin to use these tools to create meaningful content. One form of digital storytelling is microfilm, which is "a very brief presentation that lasts from a few seconds to no more than five minutes. It allows the cashier to combine personal writing, photographic images or video, narrative, sound and music. Many people, regardless of their skill level, can tell their stories through images and sounds and share them with others.

1.3 Pair Programming

Pair programming, often used in computer programming courses, is a collaborative concept in writing programs. The effects of pair programming are obvious in the way it enhances students' performance, reduces programming errors, and produces high-quality programs. Meanwhile, students' ability to write a specified program is also improved, while their confidence is enhanced (Gorla & Lam, 2004; Han, Lee & Lee, 2009; McDowell, Werner, Bullock & Fernald, 2006). Pair programming is defined as the activity that two programmers complete one specific programming task together. Each pair of programmers play two roles: one programmer writes the program, and the other examines whether it is correct or not. Different from individual programming, pair programming can produce a shorter program and a better quality, more logical thinking process. During the programming process, team members share their knowledge by discussing their views on the topics, including problem-solving skills and the programming concepts. Based on the same activity, pair programming produces better quality coding faster than individual programming. Very often, pair programming brings several benefits: a better understanding of a logical argument, a stronger motivation to learn, and more joyful, better learning experiences that come from sharing in a classroom (Forte & Guzdial, 2005). Thus, in this study, pair programming is adopted for grouping in digital storytelling activities.

2. Curriculum Development

ScratchJr is a derivative of Scratch, used by more than 10 million people around the world (Faber, Wierdsma, Doornbos, van der Ven & de Vette, 2017; Kalogiannakis & Papadakis, 2017). However, write code to the basic reading skills, therefore, developers need to another language, it will provide a simplified way, when they were younger learning code, without any reading. ScratchJr is a visual programming language designed to introduce programming skills to children between the ages of 5 and 7. By creating projects in the canoe, children can learn to think creatively and rationally, even though they can't read. It's free for iOS, android and chromebooks. The code is created by dragging blocks to a coding region and combining them. All blocks are based on ICONS (no text), the language children can use before reading. Blocks are connected from left to right, just like words. The user interface is much simpler than Scratch. The number of categories of blocks and the number of blocks in each category are reduced, so only the most basic categories remain.

Researchers design and teach ScratchJr's lessons. The course lasts 16 weeks, one session a week for 40 minutes, and the researchers designed an 18-page lecture themselves. Table 1 shows the outlines of the course. After each student has completed the course, after each stage model, they produce an animation story that USES ScratchJr to create their own. The animation stories created by these students themselves will serve as the basis for the research objectives. ScratchJr's lessons allow students to learn from building block programs to think in math. In addition to being effective in science, technology,

engineering, and mathematics, these learning algorithms can be applied to many fields, such as social science, writing, and music. ScratchJr is designed with three modes: basic, intermediate, and advanced. ScratchJr's six types and 28 building block programs are divided into three modes of difficulty. Four lessons in the each pattern is, the former three classes are used to the use of the teaching building program, the last section is to allow each student to make a work, each work with LonelyScreen video, finally on YouTube, for later research.

Table 1. Syllabus of course model.

MODE	ACTIVITY	TIME
PREPAPRATION	Code.org course	One session
INTRODUCTION	Introduction to the basic usage of iPad and ScratchJr interface.	One session
BASIC	Choose roles, add or remove roles, change backgrounds, add or delete pages Learn the basic programs by building blocks Be familiar with the following functions: Start on Green Flag (B11), Move Right (B21), Move Left (B22), Move Up (B23), Move Down (B24), Turn Right (B25), Turn Left (B26), Hide (B35), Show (B36), Set speed (B53), End (B61), Go to Page (B63) Produce a multi-page, multi-character animation	Four sessions
INTERMEDIATE	Learn to touch, repeat, modify roles and backgrounds, and add text Learn the intermediate programs by building blocks Be familiar with the following functions: Start on Tap (B12), Hop (B27), Go Home (B28), Say (B31), Grow (B32), Shrink (B33), Reset Size (B34), Wait (B51), Stop (B52), Repeat (B54) Produces a multi-page, multi-character, and text-based animation	Four sessions
ADVANCED	Learn control flow, infinity, add sound Learn advanced programs by building blocks Be familiar with the following functions: Start on Bump (B13), Start on Message (B14), Send Message (B15), Pop (B41), Play Recorded Sound (B42), Repeat Forever (B62) Produce a full story with multiple pages and characters and a written narrative	Four sessions
PEER ASSESSMENT	Watch classmates' work	One session
REFLECTION	Students interview each other and use the iPad mini to record videos	One session

3. Data Collection And Evaluation

This study collects three ScratchJr projects completed by each student after three stages: basic, intermediate and advanced. Therefore, each student should have three works, one for each of the basic, intermediate and advanced stages. However, due to the absence, only 24 students have three complete works, two students have only one work and one student has only two works. Finally, the students interviewed with each other. The two groups used the iPad mini as a tool to interview each other's questions, and then recorded the interview results. When students interview with each other, the researcher designed the following questions to provide as interview guidelines:

What's your story about?

What are the blocks your character uses? Why do you choose those bricks?

Which part do you think you are doing best?

Do you think there are any other ways to play the role's building block arrangement?

Does your story go according to your meaning?

Do you think it's difficult to use blocks to make character movements?

If you have more time, what would you like to add?

Do you think it's interesting to tell a story with ScratchJr? What's more interesting?

Because of the absence, only 25 interviews were recorded. This means that the total number of students who have comprehensively analyzed the three works is 24, and the video record of the analysis and discussion is 25.

In 28 basic ScratchJr projects, the blocks selected by students can be shown as Fig. 1. At first it was the basic stage mode, so the students only used the building block program B11 in the yellow startup building block program. Blue moving blocks program taught B21 to B26 these six blocks, but there are students from beginning to end in B21 a building program, and it is the students have been using B27 and B28 hasn't taught the building blocks of the program, because blue mobile building

blocks, program is relatively simple building blocks. Only B35 and B36 were taught in the stage of purple building blocks, but some students have already used B31 and other building block programs. The building block program for orange control only teaches B53, so some students use it. As for B54, which has not been taught by students, it is wrong to examine its usage, so it will not be included in the project of computational thinking. The final red end of the block program, because to use the screen to switch to the next page, so most students are useful to B63. In fact, if the program sequence is complete, B61 will be used, but because it does not affect the results, many students are useless.

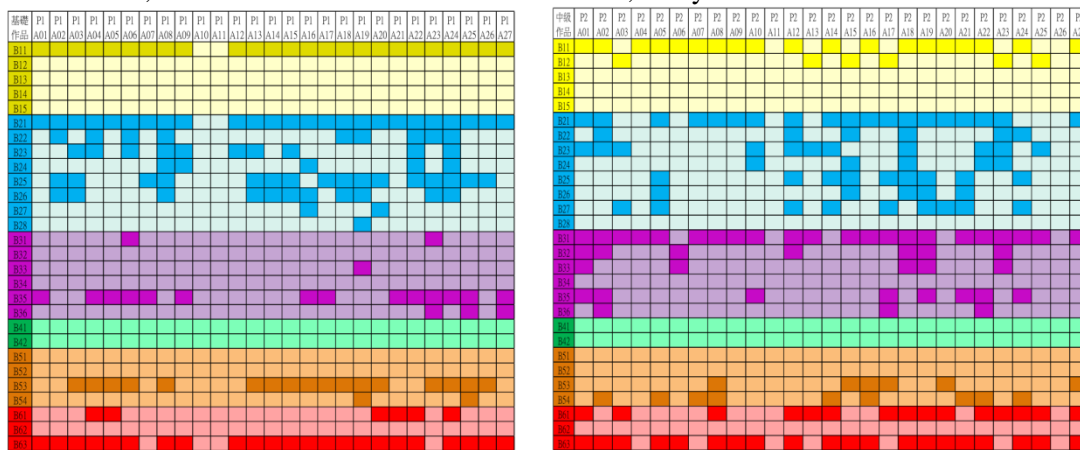


Figure 1 & 2. Distribution table of blocks in basic projects (left) and intermediate stage (right).

What ScratchJr building block programs do students use in intermediate works, as shown in Fig. 2. In this stage, we taught B12 in the yellow startup building block program, so the student's work no longer begins with B11. Blue mobile blocks program taught B27 and B28 this two blocks program, but students don't use B28, because in the operating interface is B28 button, so students don't have to add it to the program in the sequence. The purple building blocks taught B31, B32, B33, and B34 at this stage. Students prefer to use B31, which is useful for most students. The orange-controlled block program teaches B51, B52, and B54, and because B54 has examples to learn, many students know how to use this important block program. The red end of the building blocks, program and no protestant blocks, but because is the next page so B63 still a lot of people in a row, though students know B61 not doesn't matter, but in order to the integrity of the program sequences, there are more students use the building blocks of the program.

The advanced mode teaches B13, B14, and B15 in the yellow startup block program. All three are important building block programs, especially B14 and B15, not only important but probably the most difficult building block programs in ScratchJr as a whole. More time was spent teaching, and examples were used to aid the explanation. Then there's the green sound building block program, which teaches B41 and B42. The two building block programs are simple, but the environment can be a distraction, so not many students use the green sound building block program. Finally at the end of the red building program taught in B62, to ask students to pay attention to the difference of B62 and B54, but really want to use this building program in the story is not much, so rarely used by students. The analysis results are shown in Fig. 3. Because B14 and B15 this two blocks programs involve wide, the most available to 6 kinds of color change, if you use the two blocks, and use it correctly, that has learned of the parallel operation thinking concepts, events and conditions, namely C13, C14 and C15. There are also repetitive use and blending and abstraction and modularization in operational thinking practices, namely C23 and C24. Take the advanced work of student A25 for example, P3A25. He wrote a very long program number, but only used B11, B14, B15, B31, B61 and B63 building block programs. Blue and green didn't work, and purple did. But the whole process is correct, so that the role of the next action is continuous and smooth, indicating that his thinking logic is very clear.

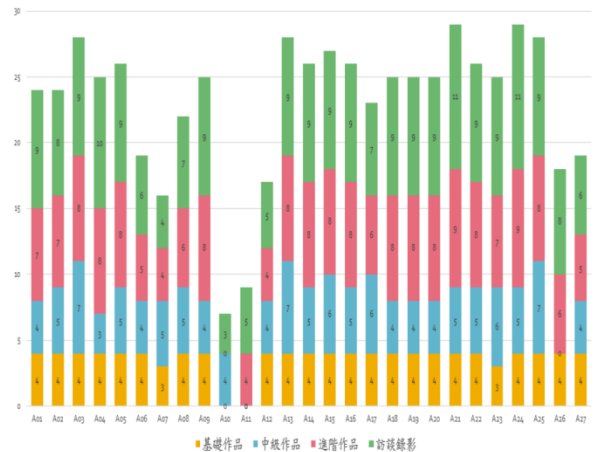
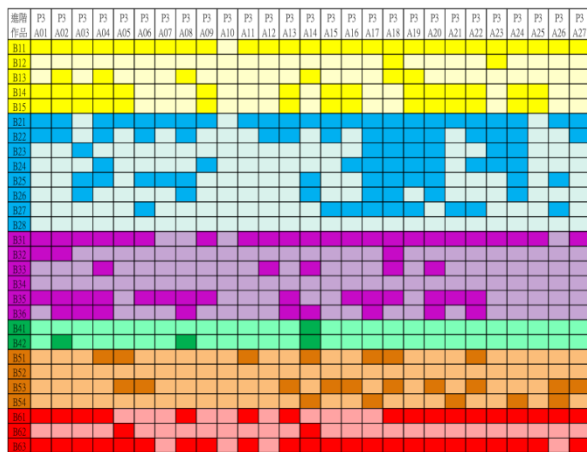


Figure 3 & 4. Block distribution in advanced mode (left) and students' CT ability in three stages.

After finishing the basic, intermediate and advanced works, the students were asked to use the iPad mini as the interview tool to interview each other's questions. Record the interview process using the iPad mini as analysis material. Take the interview process of student A16, P4A16 as an example:

- [00:00] Interviewer: now, would you please tell me the outline of your story?
 [00:02] Presenter: one day a rabbit and a turtle are racing to see who can run fast. The rabbit was far ahead. He thought the turtle was too slow, so he took a nap under the tree. Still undaunted, the tortoise ran on, and passed the hare. When the hare reached the end of the race, the turtle was startled to know that he was first.
 [00:36] Interviewer: what are the building blocks for your character? Why?
 [00:40] Presenter: I used orange blocks because turtles are slower and rabbits are faster. You can adjust the speed.
 [00:51] Interviewer: which part do you think is the best part?
 [00:54] Presenter: the action of a character.
 [00:59] Interviewer: Do you think there is any other way to the character's building block arrangement?
 [01:05] Presenter: You can also use the blue block, you can go to later let him move.
 [01:15] Interviewer: You have ...
 Did your story go according to your meaning?
 [01:20] Presenter: Yes!
 [01:22] Interviewer: Do you think it's difficult to use blocks to make character movements?
 [01:27] Presenter: No.
 [01:30] Interviewer: If you have more time, what would you like to add?
 [01:35] Presenter: Make the story a little bit longer, with a little more content. [01:41] Interviewer: Do you think it's interesting to tell a story with SCRATCHJR? Where's the fun?
 [01:48] Presenter: Interesting, because you can create a variety of roles.

Student A16 used long sentences to make his story very clear. He uses the B35's invisibility to create an effect of distance, using two characters in the car representing turtles and rabbits. In his story, B53 is not only used to control how fast the characters move, but also B14 and B15 are used to control the order of dialogue. The best part is that he uses ScratchJr to express common stories. But it's important that the last sentence of the interview, ScratchJr, allows him to create all kinds of characters. That is, ScratchJr allows him to express his designs and ideas.

An ability map, illustrated as Fig. 4, was made of the number of abilities in the computational thinking framework reflected in the three-stage works and interview videos of 27 students. It can be seen from the graph that the students have more and more projects in the computational thinking framework after the basic, intermediate and advanced courses and interviews, and almost all the expression skills of the operational viewpoints in the framework of computational thinking is achieved. It can be found that only four students did not reach 50%, and two of them only completed one work. As for the other two students who did not reach 50 percent, one of them had a learning disability. Therefore, it is feasible to integrate computational thinking into the teaching of digital storytelling to improve the expression ability.

4. Conclusion And Future Work

Because this research is to use the National small information technology curriculum to carry out, but the computational thinking emphasizes that when the computational thinking is truly integrated into the human activities of the whole, computational thinking is as an effective tool for solving problems, everyone should master, everywhere will be used. Therefore, it is suggested that the future of the different disciplines can be integrated into the calculation of thinking, training and training. For example, the mathematics curriculum can take the four steps of computational thinking: Disassembly problem, pattern recognition, abstraction and algorithm, as the direction of the problem-solving process. It should not be confined to information technology courses. This study uses both the iPad Mini and the ScratchJr iPad app. But in this era of rapid information technology, there are a lot of things like the iPad Mini and ScratchJr iPad app, and probably a better learning platform than the iPad Mini, which is more convenient than the SCRATCHJR iPad app for visual programming. Statement Therefore, it is suggested that different tools can be used in the future to achieve better learning results. But computational thinking is a universal way of thinking and basic skills that should be actively studied and used by all. Therefore, it is suggested that teachers of different courses should be taught to design teaching courses, to carry out learning activities, or to be a team, with teachers from different disciplines, to design teaching courses together. This interdisciplinary subject integration teaching course is in fact the direction of future curriculum reform.

Acknowledgements

The research reported in this paper has been supported by the Ministry of Science and Technology in Taiwan under research project numbers MOST 104-2628-S-024 -001 -MY3 and 105-2511-S-024 -007 -MY3.

References

- Chang, C. K., Tsai, Y. T., & Chin, Y. L. (2017, July). A visualization tool to support analyzing and evaluating Scratch projects. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 498-502). IEEE.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research*, *86*(1), 79-122.
- Faber, H. H., Wierdsma, M. D., Doornbos, R. P., van der Ven, J. S., & de Vette, K. (2017). Teaching computational thinking to primary school students via unplugged programming lessons. *Journal of the European Teacher Education Network*, *12*, 13-24.
- Forte, A., & Guzdial, M. (2005). Motivation and nonmajors in computer science: identifying discrete audiences for introductory courses. *IEEE Transactions on Education*, *48*(2), 248-253.
- Gorla, N., & Lam, Y. W. (2004). Who should work with whom?: building effective software project teams. *Communications of the ACM*, *47*(6), 79-82.
- Han, K. W., Lee, E., & Lee, Y. (2009). The impact of a peer-learning agent based on pair programming in a programming course. *IEEE Transactions on Education*, *53*(2), 318-327.
- Hayes, E. R., & Games, I. A. (2008). Making computer games and design thinking: A review of current software and strategies. *Games and Culture*, *3*(3-4), 309-332.
- Kafai, Y. B. (2012). Learning design by making games: Children's development of design strategies in the creation of a complex computational artifact. In *Constructionism in practice* (pp. 87-112). Routledge.
- Kafai, Y. B., & Burke, Q. (2015). Constructionist gaming: Understanding the benefits of making games for learning. *Educational psychologist*, *50*(4), 313-334.
- Kalogiannakis, M., & Papadakis, S. (2017, August). Pre-service kindergarten teachers acceptance of "ScratchJr" as a tool for learning and teaching computational thinking and Science education. In *proceedings of the 12th Conference of the European Science Education Research Association (ESERA), «Research, practice and collaboration in science education»*. Dublin City University and the University of Limerick, Dublin, Ireland (pp. 21-25).
- McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2006). Pair programming improves student retention, confidence, and program quality.
- Williams, L. A., & Kessler, R. R. (2000). All I really need to know about pair programming I learned in kindergarten. *Communications of the ACM*, *43*(5), 108-114.

A mind-mapping guide based on Unified Modeling Language for developing educational role-playing games

Ahmed TLILI ^{a*}, Mouna DENDEN ^b, Fathi ESSELMY ^b, Mohamed JEMNI ^b, KINSHUK ^c,
Nian-Shing CHEN ^d, Maiga CHANG ^e & Ronghuai HUANG ^a

^aSmart Learning Institute of Beijing Normal University, China

^bResearch laboratory of Technologies of Information and Communication & Electrical Engineering (LaTICE), Tunis higher school of engineering (ENSIT),
University of TUNIS, Tunisia

^cUniversity of North Texas, 3940 N. Elm Street, G 150, Denton, TX, 76207, USA

^dDepartment of Applied Foreign Languages, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan

^eSchool of Computing and Information Systems, Athabasca University, Canada

*ahmed.tlili23@yahoo.com

Abstract: Designing an educational game can be a difficult task for individuals who do not have appropriate background towards game developments. Several game genres exist in the literature which can make beginners confused about which are the appropriate or required game elements and functionalities for a particular game genre. At the same time, many studies highlighted the effectiveness of mind-mapping techniques in enhancing learning and thinking task. Therefore, this study deploys the powerful graphical representation of Unified Modeling Language (UML) class diagram to present a mind-mapping guide for people who are developing educational role-playing games. A total of 113 students used the developed mind-mapping guide in a public Tunisian University, and the obtained results showed that the guide can: (1) significantly enhance beginner students' level of knowledge regarding developing role-playing educational games; and, (2) promote the creative thinking when it comes to imagining educational role-playing game scenarios. Also, these results showed that students have a favorable attitude towards using the mind-mapping guide to learn and design their games.

Keywords: Educational games, Mind map, Unified Modeling Language, Role-playing, Game design

1. Introduction

Various game generators and engines are available in the literature that have made the process of developing new games, including educational ones, easier. They significantly help in reducing the effort required for the development of games by including in game engines basic functions that are common to a number of games (Gelman, 2018; Juang, Hung & Kang, 2011). Consequently, beginners of game development, including educators and students, will be able to create their own games without having any particular programming skills. However, designing a good game still need appropriate background knowledge and skills about games. Many game genres with different elements and functionalities exist in the literature. Thus, it becomes important to collect enough information about a specific type of the game in order to effectively develop a better game of that type. In this context, Baranowski et al. (2016) stated that an effective game design research must examine the appropriate game features to be included in the game in order to engage users and facilitate game transfer to real-life behaviors. Besides, each game genre contains different features based on its characteristics (Paavilainen, Alha & Korhonen, 2017). Therefore, providing a guide, which clearly defines game elements,

functionalities and features of a given game genre, would help beginners in the game development process. The proposed guide also aims to support students' creative thinking in providing original game scenarios that could be implemented as an educational game. A game scenario is a set of activities that the players have to go through in order to achieve the game objective.

De Troyer, Janssens and Vandebosch (2013) pointed out that mind-mapping tools are more suitable to structure ideas and concepts. Additionally, these tools provide a structure for thinking in a nonlinear manner. Specifically, this study adopts Unified Modeling Language (UML) class diagram to present a mind-mapping guide to support the development of educational role-playing games (RPGs). To the best of our knowledge, there were no examples, models or guides found in the literature which support beginners for developing role-playing educational games.

The rest of the paper is organized as follows: Section 2 presents the developed mind-mapping guide for beginners. Additionally, it presents the motivation behind the need for it. Section 3 describes the used method to validate the game design guide, while Section 4 presents the results. Finally, Section 5 discusses these results, concludes the study with a summary of the findings and potential research directions.

2. The development of Mind-Mapping guide

To develop the mind-mapping guide, “educational role-playing games” element (noted as RPG in figure 1) was the starting point which will be related to the other elements and components. Specifically, to provide a realistic generic guide, all the game elements, features and functionalities presented in this guide are based on various educational role-playing games in the literature. Thus, a comprehensive literature search was conducted to identify these games and use them to create the guide using various keywords, such as “educational games”, “role-playing educational games” and “game-based learning”, in different electronic databases, such as IEEE Xplore and ScienceDirect. In particular, this review was limited to only journal papers because they contain more descriptions about the implemented educational games that could be used in the developed mind-mapping guide.

As shown in Figure 1, all the game elements are represented as classes (□). The game functionalities on the other hand are represented with different relations between classes (which are game elements). For instance, all educational role-playing games offer the students a character to control and a virtual environment to learn and play within it. Therefore, the relation between the RPG class, the virtual environment class and the character class in the proposed guide is “composition” (—◆) (i.e., the game environment and character classes are mandatory in RPGs). In addition, some game elements are optional and not always found within the virtual environment, thus they are related to the virtual environment class using the “aggregation” relation (—◇). Furthermore, different missions are found in RPGs which can be in the form of combats or challenges, hence the mission class and both combat and challenge classes are related using the “inheritance” relation (—▷). In particular, various types of challenges are found in RPGs namely time, memory/knowledge cleverness/logic, endurance and dexterity (Feil & Scattergood, 2005). Thus, the feature “enumeration” is used with the challenge class. Finally, the game activities which are used to transmit the learning content to students during the learning-playing process, as shown in table 1, are presented using “association” relations (—→). These relations highlight how the classes (game elements) communicate together.

As shown in Figure 1, the developed mind-mapping guide highlights the functionalities that beginners can design with the game elements (defined using classes and relations). Consequently, they are all well-informed about the game elements and functionalities that can be included within their role-playing games. Particularly, the internal architecture that this guide would present about these games should help beginners know the different game elements to implement within their game. Furthermore, the developed guide can support them to think and generate their game scenarios by referring to the way the game elements (represented as classes) communicate with each other (represented as relations). However, attributes and methods are not specified in this proposed guide because they are beyond its main objective highlighted above. It should be noted that this developed mind-mapping guide was refined and validated by fifty one international game developers before being used with students, as highlighted in (Tlili, Essalmi, Ayed, Jemni, & Kinshuk, 2016).

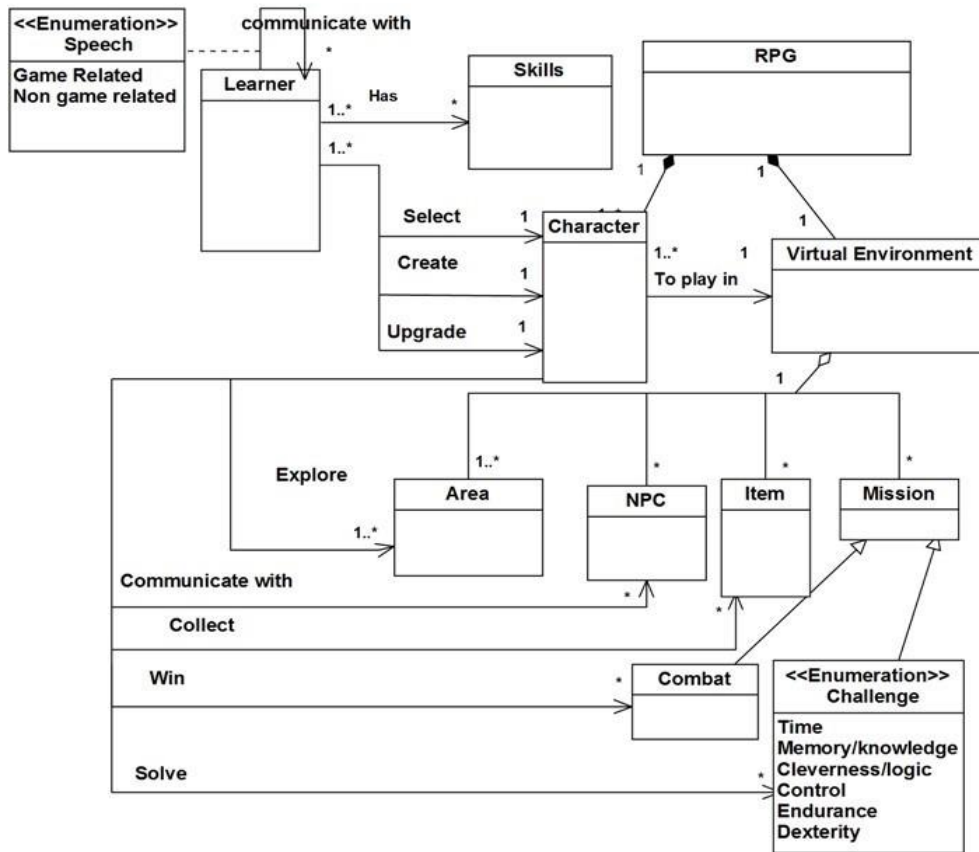


Figure 1. The developed mind-mapping guide for game development.

3. Method

This section investigates the effectiveness of the proposed mind-mapping guide in supporting beginner students design educational role-playing games. Specifically, this study investigates the following research questions.

1. Does the developed mind-mapping guide significantly enhance the students' level of knowledge towards designing educational role-playing games in comparison with the traditional learning method?
2. Does the developed mind-mapping guide enhance the students' creative thinking when it comes to designing educational role-playing game scenarios in comparison with the traditional learning method?
3. Do the students have a favorable attitude towards using the developed mind-mapping guide?

3.1 Participants

One hundred and thirteen students participated in this experiment in a public Tunisian University after the Institutional Review Board (IRB) approval was obtained. These students, aged between 21 and 23, were undergraduate students majoring in computer science. All of them are enrolled in the "game development" course which was recently added to the curriculum. Additionally, they all had taken UML courses earlier last year. Furthermore, they had all reported that they had never taken any game design or development courses (public or private courses) before and were considered as the beginners of game development.

3.2 Procedure

Due to the University context limitation, it was very difficult to divide students to two groups, namely control and experiment groups. Therefore, a within-subjects design experiment also known as repeated

measures design was applied (Ellis, 1999) to validate the effectiveness of the mind-mapping guide in supporting beginners to design their educational role-playing games. In this experiment, the same group of participants serves in more than one learning method to learn designing role-playing games. In this context, during the second semester, one hundred and thirteen beginner students studied the design of educational game development using the traditional learning method namely lectures (first learning method) for six weeks (one hour and half per week). These students then took a written test, namely the learning and creative thinking test, which includes questions regarding various knowledge they learned and question asked them to imagine an educational role-playing game scenario that can be implemented.

After that, as a second learning method, the students used the proposed mind-mapping guide to learn about role-playing games for six weeks as well (one and half hours per week). Each of the students (113 students) was given the proposed mind-mapping guide to get to learn about role-playing games and its possible functionalities, features and elements. The instructor then has started drawing on the board this guide (step by step) and going with the students through each branch and each element (class) by analyzing it and explaining each part. Also, he answered their questions and discussed their suggestions to ensure that this proposed mind-mapping guide is fully understandable. Finally, the students took the same learning and creative thinking test regarding the design of educational role-playing games. Furthermore, these students gave their opinions regarding learning with the mind-mapping guide through unstructured interviews.

3.3 Data collection

To collect data from students while learning, the following three methods are used.

- *Level of knowledge questions:* The learning test is a paper-based written test, aims to investigate the impact of both learning methods (lectures and the developed mind-mapping guide) on the students' level of knowledge regarding the design of educational role-playing game. The students were asked to answer questions regarding various knowledge, such as game elements, functionalities, that developers should know while designing their educational role-playing games.
- *Creative Thinking test:* Inspired by various thinking and creativity instruments, such as Guilford (1967) and Torrance (1966), in the test there is a question aims to investigate the effectiveness of both learning methods (using lectures and the developed UML guide) in simulating creative thinking within students to generate educational role-playing game scenarios. In this context, the students were asked to imagine an educational role-playing game scenario that can be implemented and write it down. To validate these proposed game scenarios, a group of experts, including game designers and developers, were selected to review these scenarios and validate them based on the following scoring: (1) Three points on Elaboration - amount of details given by students regarding their educational game scenarios; (2) Three points on Originality - Degree to which the proposed game scenario is original and not similar to an already developed games in the literature/game market; and, (3) Three points on Implementation - Degree to which the proposed game scenarios can be implemented in the future. A maximum of nine points can be obtained by students on their proposed educational role-playing game scenarios.
- *Unstructured interviews:* It was used to collect the opinions of students regarding the proposed guide. This instrument is very flexible and allows participants to give researchers their most honest and direct opinions. In the 15-minute interview, students can talk freely about the importance of the mind-mapping guide.

4. Results

4.1 Impact of the mind-mapping guide on students' level of knowledge

The results of the two learning tests were analyzed using the paired samples t-test method. This method is appropriate in this context because the same group of students answered both learning tests (after learning with lectures and mind-mapping guide). In this case, the null hypothesis is that there is no

difference between the mean score of the two level of knowledge tests. Table 1 presents the obtained results. As shown in Table 1, the mean of the level of knowledge test results after using the mind-mapping guide was higher than the test results after taking lectures. This is seen where the mean difference is equal to 7.65. Particularly, this difference was significant since the p value is equal to .01 and less than 0.05. Consequently, the null hypothesis is rejected. Therefore, the mind-mapping guide as a learning method has significantly enhanced the students' knowledge regarding designing educational role-playing games, compared to the lectures learning method.

Table 1. Paired samples t-test results of the level of knowledge test

Pair 1	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Dev	Std. Error Mean	95% CID				
				Lower	Upper			
Test 2 - Test 1	7.65	3.95	.37	6.92	8.39	20.69	112	.01

4.2 Impact of the mind-mapping guide on students' creative thinking

Experts considered that scenarios which got 5 points or more are considered valid. The number of valid game scenarios using the first learning method (using lectures) is very limited and only has SIX scenarios. However, after using the mind-mapping guide, the number of valid game scenarios is improved to have FORTY-SIX scenarios. This is seen when almost half of the students were able to write a valid game scenario. Thus, it can be deduced that the internal architecture this guide presents about these games can help students to know the different game elements to implement within their game and how these elements to communicate to generate a game scenario that could be implemented.

Furthermore, the creative thinking test results were analyzed using the paired samples t-test method. The null hypothesis is that there is no difference between the mean score of the results. As shown in Table 2, the mean value of the results after using the mind-mapping guide is higher than the results after taking lectures. This is seen where the mean difference is equal to 13.12. Particularly, this difference was significant since the p value is equal to .00 and less than 0.05. Consequently, the null hypothesis is rejected. Therefore, the mind-mapping guide as a learning method has significantly enhanced the students' creative thinking regarding imagining educational role-playing game scenarios, compared to the lectures learning method.

Table 2. Paired samples t-test results of the creative thinking test

Pair 1	Paired Differences					t	df	Sig.(2-tailed)
	Mean	Std. Dev	Std. Error Mean	95% CID				
				Lower	Upper			
Test 2 - Test 1	13.12	1.95	.21	.67	1.64	4.85	112	.00

4.3 Impact of the mind-mapping guide on students' opinions

Students were asked to freely give their opinions about the proposed guide, as a learning material; did they find it effective and helpful? Did they find it interesting or not? This helped to further evaluate it through their given feedback. As a result, eighty students found the mind-mapping guide helpful, nine students found it hard to understand, ten students did not give any feedback and finally fourteen students gave different feedback (e.g., the mind map guide would be efficient if it contains some descriptions). Based on these statistics, we can conclude that the students have a favorable attitude toward the proposed mind-mapping guide.

5. Discussion and conclusion

Buzan and Buzan (1996) mentioned that the human brain is drawn to graphics more than words, thus mind-mapping should always relies on graphics. Also, they mentioned that the use of graphics and no linear branches stimulate the entire brain. This study presented a mind-mapping guide to support the design of educational role-playing games for beginners. Based on the obtained results, the presented mind-mapping guide conceived in graphics is very effective for students learning the design of educational role-playing games.

Boley (2008) found that the learning outcomes of nursing students who have used mind-mapping are much better than those who have not used mind-mapping. The obtained findings also showed that the developed mind-mapping guide can help beginners enhance their knowledge regarding educational role-playing games compared to the traditional method. Additionally, Wu and Chen (2018) stated that mind mapping can be used as a learning tool to promote thinking in both the left and right brain hemispheres of a student. Spencer, Anderson and Ellis (2013) stated that mind-mapping can enhance creative thinking and knowledge attainment. For instance, Wang, Ding, Xu, Wei and Dilinar (2014) found that mind-mapping supports student's divergent thinking and their innovation skills in learning of medical immunology. Based on the obtained results, the provided mind-mapping guide does support the thinking or the imagination task regarding educational game scenario within students compared to the traditional method. On the other hand, some limitations are found in this study which should be acknowledged and further researched. For example, the conducted experiment did not include control and experiment groups. Additionally, the proposed mind-mapping guide requires a short learning session about UML, specifically the class diagram and its different types of relations. This helps to better understand the guide and make full use of it.

References

- Baranowski, T., Blumberg, F., Buday, R., DeSmet, A., Fiellin, L. E., Green, C. S., ... Young, K. (2016). Games for Health for Children-Current Status and Needed Research. *Games for health journal*, 5(1), 1–12. doi:10.1089/g4h.2015.0026
- Boley, D. A. (2008). Use of premade mind maps to enhance simulation learning. *Nurse Educator* 33 (5), 220–223.
- Buzan, T., & Buzan, B. (1996). *The Mind Map book: How to use radiant thinking to maximize your brain's untapped potential*. New York: Plume.
- De Troyer, O., Janssens, E., & Vandebosch, H. (2013). How to kick off the development process of a serious game?. In *CHI workshop: Lets talk about Failures: Why was the Game for Children not a Success*.
- Ellis, M. V. (1999). Repeated measures designs. *The Counseling Psychologist*, 27(4), 552-578.
- Feil, J., & Scattergood, M. (2005). *Beginning game level design*. Course Technology PTR.
- Gelman, G. M. (2018). U.S. Patent No. 9,875,617. Washington, DC: U.S. Patent and Trademark Office.
- Guilford, J.P. (1967). Creativity: yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14.
- Juang, J. R., Hung, W. H., & Kang, S. C. (2011). Using game engines for physics based simulation-A FORKLIFT. *Journal of Information Technology in Construction*, 16, 3-22. Access: https://www.itcon.org/papers/2011_2.content.00898.pdf
- Paavilainen, J., Alha, K., & Korhonen, H. (2017). A Review of Social Features in Social Network Games. *Transactions of the Digital Games Research Association*, 3(2).
- Spencer, J. R., Anderson, K. M., & Ellis, K. K. (2013). Radiant thinking and the use of the mind map in nurse practitioner education. *Journal of Nursing Education*, 52(5), 291-293.
- Tlili, A., Essalmi, F., Ayed, L. J. B., Jemni, M., & Kinshuk. (2016). Towards a generic UML model to support designing educational role playing games. In *2016 IEEE 16th International Conference on Advanced Learning Technologies*, 153-157.
- Torrance, E. P. (1966). *Torrance tests of creative thinking: Directions manual and scoring guide*. Personnel Press, Incorporated.
- Wang, S., Ding, J., Xu, Q., Wei, X., & Dilinar, B. (2014). Application of Mind Map in Teaching and Learning of Medical Immunology. In *Frontier and Future Development of Information Technology in Medicine and Education* (pp. 2091-2094). Springer Netherlands.
- Wu, T. T., & Chen, A. C. (2018). Combining e-books with mind mapping in a reciprocal teaching strategy for a classical Chinese course. *Computers & Education*, 116, 64-80.

Dancing a treasure: A videogame to motivate young audiences toward Spanish dance

Alejandro ROMERO-HERNANDEZ^{a*}, Manuel GONZALEZ-RIOJO^a, Meriem EL-YAMRI^a, Borja MANERO^a

^aComplutense University of Madrid, Spain

*alerom02@ucm.es

Abstract: The increasing in the average age of performing arts audiences is a worldwide problem. If we focus on Spanish dance, the problem becomes more serious. Several studies alert that the dance situation in Spain is critical: the number of spectators in this discipline has been decreasing in the last ten years, reducing its numbers by a half in this period. On the other hand, there is an incipient current that advocates the use of new technologies -especially, videogames- to attract young audiences.

This article details the creation process of "Dancing a treasure", an educational videogame aiming to motivate the youngsters towards the Spanish dance. First, we discuss the Application Domain Research Document (ADR Document), a design document to help new researchers by shortening designing times in similar projects. Next, we detail the videogame design decisions based on the lessons learned from the ADR phase. Finally, we argue the most important points of the development and production phase.

This work aims to set a precedent for the design and development of videogames based on performing arts, simplifying the work of other developers facing a similar project.

Keywords: Serious games; Engagement; Dance; m-learning

1. Introduction

According to the latest Yearbook (2017) of the General Society of Authors and Publishers (SGAE, 2017), the dance situation in Spain is critical. According to this report, the number of spectators in this discipline has been decreasing in the last ten years, reducing its numbers by a half in this period (1.65 million in 2007 compared to 0.85 million in 2016).

This is not an isolated report. It is also demonstrated by the data collected by the latest survey on cultural habits in Spain, produced by the Spanish Ministry of Education, Culture and Sports (Ministerio de Educación, 2015), which shows that only 7% of the people surveyed have attended a dance show between 2014 and 2015. This situation is aggravated if we focus on Spanish dance, a big representative of the folk culture. This type of shows is attended only by 17.9% of the regular audience of dance shows, which results in 1.25% of the total number of respondents. These data pose a risk to a very representative part of the Spanish culture, our most typical dances.

A good way to keep this type of shows alive in the medium and long term is to bring them closer to young people, in line with the thought: "You only can love what you know". We live in the age of digital natives and it is very difficult to reach an audience that receives stimuli in form of information from many sources simultaneously (Piscitelli, 2008). To solve this problem, in the last decades a current has emerged: *serious games*. This type of games has a lot of applications. Different studies indicate that they can be very useful when it comes to increasing student interest (Sung & Hwang, 2013), their motivation to learn (Dickey, 2011; Hwang, Wu, & Chen, 2012; Van Eck, 2007) and the engagement with culture (Shapiro & Hall, 2017) using specific techniques (Wouters & Van Oostendorp, 2017).

Currently, most studies based on serious games are being conducted in purely technical fields, such as mathematics or computer science. This is known as STEM (*Science, Technologic, Engineering, Mathematics*), a term that encompasses all these fields. Despite this, there are American researchers developing tools outside these limits -and that is also reaching countries like Spain- to the creation of a new term: STEAM. "A" introduces the concept of Arts within the set of disciplines of educational videogames. Arts based videogames have already been created, for example, based on classical theater

plays, which have demonstrated the effectiveness of these tools to motivate young people (Hernandez, Perez, Riojo, Morata, & Manero, 2016; Manero, Fernández-Vara, & Fernández-Manjón, 2013).

The National Ballet of Spain (from now on BNE) is a public dance company. Its fundamental task is to preserve and spread the rich Spanish choreographic heritage as an intangible heritage that is a sign of cultural identity, collecting its stylistic plurality and its traditions, represented by its different forms: Ballet, *Danza Estilizada*, *Folclore*, *Escuela Bolera*, and *Flamenco* (España, 2018). Until now, the BNE has tried to alleviate the problem of the public influx with campaigns to recruit young people through traditional formats with limited results. The book "Bailando un Tesoro" (Azucena Huidobro; Mercedes Palacios, 2016), written by Azucena Huidobro, artistic director of this institution, is the best example of this. For this reason, the BNE has decided to test the use of new technologies to attract young audiences.

This article details the design of the videogame "Bailando un Tesoro", a tool based on Spanish dance, which objective is to motivate young people towards Spanish dance. We focus the design to children from 9 to 11 years old. This document can be useful for future researchers who want to carry out similar projects. This project, as we will see later, has been carried out in close collaboration with the BNE.

This article is structured as follows: In section 2 we talk about the domain research process prior to the videogame design. Section 3 illustrates this design phase, both narrative and interaction. Section 4 details the development and production of the tool. In section 5, we describe the conclusions drawn from this project. Finally, section 6 illustrates the long way to go with the presented tool and future related projects.

2. Domain of application phase

As in many occasions, when it comes to creating educational tools for other areas than computer science, the big problem we face is the lack of information that the designers have about the specific area of knowledge. Given this lack of knowledge, the first step was to know this area and study if there is any precedent tool in this field. This project puts dance and computers science together, two disciplines that share a small common ground at first. Thus, before starting the design, we had to familiarize ourselves with the BNE and the Spanish dance. It is very important to bear in mind that this investigation did not find us any similar videogame in the market, so we had to start from scratch. We started by creating an Application Domain Research Document (ADR Document), a document that details the steps to follow in this first phase (See Figure 1). In this graphic, the different phases that were included in said document are detailed.

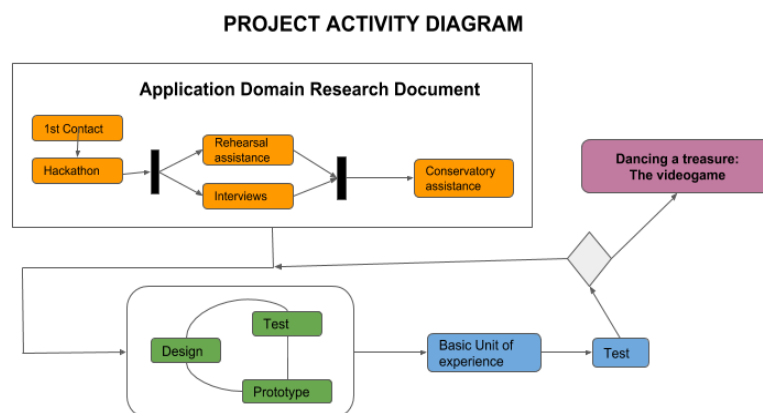


Figure 1. Research of application field diagram contextualized.

As reflected in our ADR Document, the first step was the organization of a hackathon. The lack of similar games forced us to place ourselves in a phase of divergent thinking (the main goal was to produce as many ideas as possible without any restrictions) in which to compile different ways of creating a videogame of these characteristics. This hackathon would work as a brainstorming phase but with real prototypes. During this weekend event, the participants had the assistance of dancers from the company itself, who acted as technical consultants in the field of Spanish dance.

After collecting all the ideas in the hackathon phase, during the first month and a half of the project, the designers went regularly to the BNE headquarters, in order to see trials and learn about the work that is done there. The team attended all the different rehearsals performed by the company dancers throughout the day, as well as general rehearsals of functions that were to be performed soon. At the same time, interviews were prepared with different members of the company. The goal of all these interviews was to cover the entire spectrum of ballet figures. All interviews followed the same script: 1) Personal history from the beginning of their dancing careers, 2) their day to day in the BNE, and 3) issues that are considered basic to become a dancer.

All the dancers and the interviewees agreed on the importance of beginnings. Therefore, it was fundamental to know how these beginnings were and that those future professionals would tell their experience in the first person. For this reason, the staff of the Royal Mariemma Professional Dance Conservatory was able to help. With their help, we attended to several classical ballet and Spanish dance classes for children between 8 and 12 years old. Furthermore, we were able to speak directly with several of these children. They gave their vision of the process they were going through.

3. Design phase

3.1 Narrative design

Thanks to the information collected in the ADR phase, especially in the interviews with the dancers, the weight of the dancer's story in the life of the dancers was clear. In addition, by comparing the stories of the different interviewees, one could clearly observe a pattern in their life stories, where their beginnings in the dance world are a fundamental part.

The book "Dancing a treasure" (Azucena Huidobro; Mercedes Palacios, 2016) tells the story of 4 children who love to dance. The 4 styles of Spanish dance (*Flamenco, Folclore, Danza Estilizada and Escuela Bolera*) are told through these kids. Because this book is written by a former dancer, you can see similarities between this narrative and the beginnings of a dancer. These similarities made the design team decide that this world created for the book was the perfect starting point for the game story. The goal of this phase was to find the best way of - knowing the data of the characters offered by the book - giving life to it and creating a context that included all the data collected in the domain of application research phase.

The first phase of the design was focused on the design of the book worlds and the gamification (i.e. the stages and how the game progresses). Thanks to interviews and data collected in the ADR document, we realized that there are always three basic stages in a professional dancer's development: 1) the beginnings, 2) the conservatory; studying the dance in depth and 3) becoming a professional. For this reason, it was decided to give the narrative these three phases: 1) "Neighborhood" (tutorial) school where we try to represent the beginnings of the dancers. 2) Conservatory, where the in-depth study of dance is represented. In this phase, the four styles of the Spanish dance are explained and put into practice. 3) The National Ballet of Spain, where the last leap to professionalism is represented. In this phase there are exams or auditions to dance with the company. Finally, a small representation is done in the virtual Zarzuela Theater, a theater located in Madrid and one of the main theaters in Spain.

The second design decision was to choose how the main characters were going to be. In the BNE case, and in the dance world in general, there is a strong division by gender. Following the recommendation of the BNE, and to favor the identification of the player with a character thanks to the gender, it was decided that there would be two possible characters: a boy and a girl. It was decided not to give names to either of them to avoid distancing the characters from the user, being able to identify with the characters directly as if the game main characters were themselves.

On the other hand, there are NPCs (Non-Playable Characters), who would be the ones who would offer all the information to the players during the story. As in the book mentioned before, we have characters with whom the users will discover the different styles and key components of the national dance. It was decided that each of the characters would identify with one of the videogame phases (explained more in detail in the previous point). Therefore, we needed 6 NPC characters that would serve as teachers.

At the same time, the book that contextualizes this narrative, as explained at the beginning of this point, has 4 children representing each of the 4 styles of dance. In the book, they explain what is

important in each one of the styles and what is special about it. This was translated directly to the videogame, making these children act as support during the conservatory phase.

Preparation for the conservatory

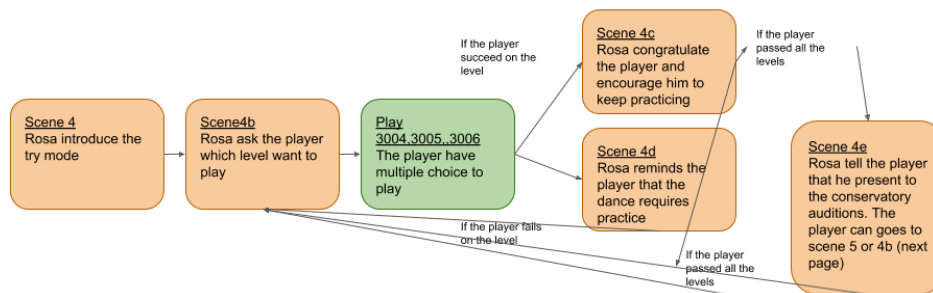


Figure 2. Scene flux diagram.

3.2 Design of videogame interactions

From the data extracted in the ADR phase, the main keys in Spanish dance were obtained: 1) movement and 2) rhythm. Therefore, based on the first point, we decided from the beginning that the main character was going to be in the center of the scene. Having the focus on it, the user must act to influence the dance. To do this, and taking rhythm as a reference, it was proposed to integrate different types of rhythmic interactions -buttons, collectibles, etc.- on the sides of the screen, in order not to hide the dance. This idea seemed to work since most of the music or dance games are focused in the same way; for example, Guitar Hero or OSU!

The main problem regarding this point was the choice of the best interactions that would be developed. Based on data collected in ADR phase, many possible types of interactions were designed in order to try and check which of these works better. Subsequently we began to prototype all these ideas to get tests and try to find the most appropriate ones. This process was repeated iteratively testing different options, varying them and discarding many of them.

Finally, we obtained the two following interactions to be implemented in the development phase:

- Tap: rhythmic button that appears following the music rhythm. With this button appears a surrounding circle that narrows as it approaches the moment of pressing it. With this interaction, we represent simple or individual movements in the dance.
- Follow the path: interaction consisting of the appearance of a back comb (with which the user interacts) that must be dragged over a "path" drawn by roses -collectables- in order from closest to farthest. This button represents movements that require displacements or that are more stylized.



Figure 3. Follow the path (left) and tap (right)

4. Development and production phase

In a first phase, we developed a Minimum Viable Product (MVP). This is a prototype of the game, as close to the final version as possible, to be used as a model for the entire game development. This strategy aims to test a version that offers a complete experience, although much shorter, to receive feedback at the lowest possible cost.

To achieve this goal, two scenes were developed: 1) a story scene, with test dialogues and 2) a dance scene where the selected interactions could be tested, the different shuffled possibilities of the 3D animations and different possible feedbacks to the user.

As already specified in the design phase, the two keys to Spanish dance are movement and rhythm. Therefore, we considered crucial from the beginning that: 1) the movement of the character had to be the most technically correct and 2) the design of the interactions had to be designed in a way that would fit both the choreography rhythm and the music.

To achieve the first objective, we used a motion capture suit, recording the BNE director dancing the different choreographies. Later, we transferred those movements to our game's main character. This recording was not exact in detail due to the low accuracy of the motion capture suit, so it was important to refine the dances together with the professional dancers.

As a design decision, 3 versions of the same dances were recorded with 3 levels of technical accuracy: 1) a perfect dance 2) a dance with certain technical failures and 3) a badly danced one. The objective with these recordings was to mix them during the gameplay in order of accuracy according to the player's good (improve the dance) or bad (the dance get worse) performance. It was intended that the user's interactions would have a direct impact on the character's dance. These recordings were finally discarded due to the BNE need that the dances must always be perfectly performed. They did not want to represent even slightly non-technically exact dances- much less some badly done ones-.

Finally, we developed a tool to configure the interactions. It tells the system what, where and when the interactions should appear on the screen. This configurator was developed to allow people without technical knowledge to modify choreographies, such as dancers, allowing them to adjust this interaction to the rhythm.

5. Conclusions

The main objective of this paper is to explain the detailed design of a motivational videogame. The aim of this tool is to alleviate an important problem, that is, the low attendance to classical Spanish dance performances, demonstrated by the MECED cultural habits survey (Ministry of Education, 2015) and the last Yearbook (2017) (see Introduction).

The main conclusion drawn is that we managed to put together two very distant fields: computer science and dance. In the first steps of the project, the members of the design team used a completely different language than the one used by the BNE team. Mainly, the game designers have difficulties to understand part of the BNE main points to represent the dance. This meant that the team had to interview the BNE team more times than expected. To alleviate this, a continuous study of the area was needed, requesting the BNE to give small workshops where all these key points could be explained to the design team: in particular, multiple workshops on rhythms and tempos, terms completely unknown by the design team. With this information, we developed tools to make the game design process more precise. The conclusion of the designers is that, following the steps represented in the ADR document, the design process is optimized.

During the process several false steps were taken. Recording regular and bad dances in order to provide feedback to the player seemed to be a great idea to improve gamification. When included in the game, the BNE members couldn't admit showing not accurate dances. This misunderstanding led us to lose a great development effort.

However, if we had the ADR, this could have been avoided, since the idea was not tested correctly. The continuous testing and the inclusion of the company in the design process was key in the success of the project.

It is our hope that the design process detailed in this paper will help other educational game designers to tackle projects involving performing arts.

6. Future work

This project seeks to have some results and an academic impact. For this reason, a massive experiment will be carried out with the videogame in order to verify if it really works and fulfils what it has been created for: the increase of youth interest in Spanish dance. This experiment will be presented with a quasi-experimental design, with a control group and an experimental group. For this experiment we have the support of the National Ballet of Spain, which will allow all groups to go to a private rehearsal of this company. With the analysis of the results, we will try to verify statistically the effectiveness of the videogame. This support will also allow us to measure how the game influences future spectators.

Finally, the possible applications of all this material and learning in environments of new interfaces, such as virtual reality and augmented reality, are being considered. Currently, the research team is involved in several projects with virtual reality, such as the one financed by the BBVA Foundation for the improvement of public speaking skills. Currently, one of the authors of this article is working on adapting this project to a virtual environment.

Acknowledges

This project has been partially funded by BBVA foundation (ComunicArte project: PR2005-174/01), and Ministry of Science, Innovation and Universities of Spain (Didascalias, RTI2018-096401-A-I00)

References

- Azucena Huidobro; Mercedes Palacios. (2016). *BAILANDO UN TESORO* (INST. NACIONAL DE ARTES ESCENICAS Y DE TEATRO, Ed.). Retrieved from <https://www.casadellibro.com/libro-bailando-un-tesoro/9788490412046/4689433>
- Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, 42(3), 456–469. <https://doi.org/10.1111/j.1467-8535.2009.01032.x>
- España, B. N. de. (2018). Ballet Nacional de España. Retrieved from Ballet Nacional de España website: <http://balletnacional.mcu.es/>
- Hernandez, A. R., Perez, C. D.-F., Riojo, M. G., Morata, A. C., & Manero, B. (2016). Computers or tablets? Choosing devices for educational videogames. *Games and Virtual Worlds for Serious Applications (VS-Games)*, 2016 8th International Conference On, 1–8. IEEE.
- Hwang, G.-J., Wu, P.-H., & Chen, C.-C. (2012). An online game approach for improving students' learning performance in web-based problem-solving activities. *Computers & Education*, 59(4), 1246–1256. <https://doi.org/10.1016/j.compedu.2012.05.009>
- Iglesias, B. M., Fernández-Vara, C., & Fernández-Manjón, B. (2013). E-learning takes the stage: From La Dama Boba to a serious game. *IEEE Revista Iberoamericana de Tecnologías Del Aprendizaje*, 8(4), 197–204.
- Ministerio de Educación, C. y D. (2015). *ANUARIO DE ESTADÍSTICAS CULTURALES*. Retrieved from https://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadistica/s/cultura/mc/naec/2015/Anuario_de_Estadisticas_Culturales_2015.pdf
- Piscitelli, A. (2008). Nativos digitales/Digital Natives. *Contratexto*, (16), 43–56.
- SGAE. (2017). Anuario de la SGAE. Retrieved January 1, 2015, from SGAE website: <http://www.anuariosgae.com/anuario2017/home.html>
- Shapiro, B. R., & Hall, R. (2017). *Making Engagement Visible: The Use of Mondrian Transcripts in a Museum*. Philadelphia, PA: International Society of the Learning Sciences.
- Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51. <https://doi.org/10.1016/j.compedu.2012.11.019>
- Van Eck, R. (2007). Building Artificially Intelligent Learning Games. In *Games and Simulations in Online Learning: Research and Development Frameworks* (pp. 271–307). <https://doi.org/10.4018/978-1-59904-304-3.ch014>
- Wouters, P., & Van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1–16). Springer.

Prototyping of an Earthquake Evacuation Learning Game with VR Reproducing the Environment That Is Familiar to Learners

Raiya YAMAMOTO*, Asaka KITAOKA & Kei INOUE
Faculty of Engineering, Sanyo-Onoda City University, Japan
*ryamamoto@rs.socu.ac.jp

Abstract: Recently, big earthquakes happen worldwide, and UNDRR adopted the Sendai Framework for Disaster Risk Reduction. Learning about earthquakes is getting more important. When learning about disasters, it is important to treat them as it is their own problem. For this, Mitsuhashi proposed the GLI (Global, Local, and Individual) model of disaster education. To make learners learn as it is their problem, we need to develop an environment that they can perform the local and individual disaster learning. It is also important to keep learning about disasters. For this, we prototyped evacuation learning game with VR, and the game reproduces the environment that is familiar to learners with VR to achieve local and individual level disaster learning. In this study, we reproduced building No.5 in our university where students mainly used every day because the target students of the game are our university students. We also conducted a preliminary experimental evaluation of the learning game. The results showed that (1) the reproduced environment can encourage learners to imagine a situation that an earthquake happens at the place that is familiar to them, (2) the game factors may motivate learners to learn about disasters.

Keywords: Disaster Education, Game-based Learning, Virtual Reality, Learning Assist System

1. Introduction

Recently, big earthquakes happen worldwide, and UNDRR (2015) adopted the Sendai Framework for disaster risk reduction. Learning about earthquakes themselves and how to evacuate from earthquakes are imperative and developing a learning environment of an earthquake is an important issue.

It is also essential to keep learning about disasters because we cannot take appropriate decisions at disasters if we do not remember the knowledge about disasters at any interval. Besides, learners must participate in disaster education as it is their problem. For this, Mitsuhashi (2018) proposed the GLI (Global, Local, and Individual) model of disaster education, which each level is defined by being focused on what people need to learn. On the global level, disasters themselves are focused and people learn general knowledge. For example, if they learn earthquakes, they learn knowledge such as earthquake intensity and magnitude and how to keep safe, etc. Then, on the local level, a community that people belong to is focused and they learn how to apply the knowledge they learned on the global level to the community. Finally, on the individual level, people learn how to expect what will happen and how to behave as an individual person when a disaster occurs in the community.

For learning about disasters as their own problem, we considered that learning on the local and individual level is important. Therefore, we aim to construct a learning platform where learners can perform local and individual level disaster learning. In order to achieve local and individual level disaster learning, we reproduce an environment that is familiar to learners with VR. Freina and Ott (2015) reported that immersive VR can offer the following advantages; VR can allow learners to have partial immersive and can support providing a safe training environment avoiding real danger. In addition, Krokos et al. (2018) revealed that immersive VR can assist learners in having better memory ability compared to non-immersive VR environment. Thus, we reproduce an environment with VR.

Many disaster learning systems have been developed, but most of them are not for the local and individual level learning (for example, the reproduced environment is an ordinary building). Li et al. (2017) developed the learning system for earthquake safety training through the virtual drill. In this

system, learners undergo training in the reproduced room with VR. The learners attempt to detect potential danger and a way to protect themselves through an immersive training experience. Ruffino et al. (2018) developed the simulation and serious game for fire evacuation training. This game is designed to learn fire evacuation from a building in urban areas with immersive VR and serious game approach. However, the reproduced places on their learning system are general buildings so that learners cannot learn on the local and individual level. Hence, we reproduce an environment that is familiar to learners.

In addition to the above, we adopt the game-based learning approach to motivate learners. Even if we develop a good learning environment, it will not have enough learning effectiveness if learners do not learn repeatedly. Many research (like Perrotta et al. 2013) reported that game-based learning and gamification approach are effective to motivate learners. So, we adopt a game-based learning approach and will implement game factors to the learning platform. Therefore, in this research, we prototype the learning game that consists of an environment that is familiar to learners and game factors to motivate them to keep learning. We also conducted a preliminary evaluation of the learning game.

2. Consideration of Learning Content

2.1 Stage in the Learning Game

As mentioned in the introduction, we aim to achieve the situation that learners can learn earthquake evacuation on the environment that is familiar to learners. Therefore, the stage in the learning game is the building No.5 in our university reproduced with VR by Unity (Figure 37).

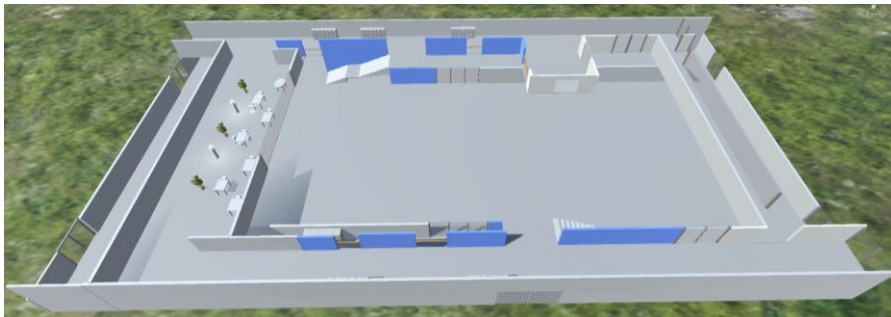


Figure 37. The Building No.5 reproduced with VR

2.2 Game Features in the learning game

2.2.1 Timer and Safety Percentage

The learning game has a timer so that learners can see the time that they spent to evacuate. The time is recorded and used by the evaluation function mentioned later. When we evacuate, not only fast but also safe evacuation is important. So, the learning game has the parameter that shows how safe learners are evacuating. This parameter is called “safety percentage.” safety percentage will decrease when they take inappropriate actions such as not hiding under the table when an earthquake occurs, going near fragile things like windows, and running when they evacuate. The game is over if the safety percentage becomes 0%. We expect that showing decreasing safety percentage facilitates learners to learn inappropriate action in earthquake evacuation.

2.2.2 Dangerous Action Count and Restart Function

If learners take a dangerous action that may lead to death, the learning game will restart, and the learners have to starting again from the start point. Also, when they take a dangerous action, the number that shows how many times they took dangerous actions (dangerous action count) will increase. Figure 38 shows the screen after restarting. The alert about the dangerous action that they took is shown. Then, they need to start again, being careful of dangerous actions. This time, we employed taking an elevator

as a dangerous action in the learning game. Thus, when learners get closer to an elevator, the dangerous action count will increase, and the learning game will restart. The game will be over when the count becomes three.



Figure 38. The ScreenShot after Restarting

2.2.3 Evaluation Function

The learning game has an evaluation function to get learners motivated to achieve better evacuation result. The function evaluates their evacuation judging from the Safety Percentage and clear time after they played the learning game. Also, the learning game records their best score.

2.2.4 Achievement List

The learning game has an achievement list to encourage learners to learn repeatedly. We expect that learners consider better behaviors on evacuation to clear assignment in the achievement list. Figure 39 shows the achievement list. When they achieve an assignment, Clear is shown on the right side of the list. The assignments in the list get more difficult from the first one to the fifth one. We expect that including the conditions (such as clear time and safety percentage) in the assignment statement facilitate learners to consider such as what better actions on the evacuation are, how to evacuate more safely.

Achievement List		Already Cleared?
Assignment 課題内容	課題一覧	達成済み
1. 避難に成功する(評価は問わない)	1. Succeeded in evacuation (with any Rank)	Clear
2. 安全度80%以上で避難に成功する	2. Succeeded in evacuation with over 80% Safety Percentage	Clear
3. 3分以内に避難に成功する	3. Succeeded in evacuation within 3 minutes	Clear
4. 評価Sで避難に成功する (条件:2分30秒以内かつ安全度90%以上で避難に成功する)	4. Succeeded in evacuation with Rank S (within 2:30 minutes and over 90% Safety Percentage)	Clear
5. 安全度100%でかつ評価Sで避難に成功する	5. Succeeded in evacuation with Rank S and 100% Safety Percentage	

Figure 39. The Achievement List

3. Learning Flow with the Game

In this chapter, we explain the learning flow with the learning game. This game is implemented with Unity for VR. Learners put Head Mount Display (HMD) and use a controller to play the learning game (HMD is Oculus Rift, and the controller is Oculus Touch).

When learners start the learning game, the screen shown in Figure 38 but without an alert will appear. The learning game starts at the students' lounge which is on the second floor of the building No.5 (Left side of Figure 37). First, learners hide under a table following the instruction until the tremors of the earthquake stopped. If they do not hide this time, the safety percentage will decrease. After stopping the tremors of the earthquake, learners start the evacuation. They are going to find doors to the outside, which are on the first floor of the building. If they can find one of them and can get out of the building, the evacuation is a success. As mentioned in chapter 2, the safety percentage will decrease if they run or go near windows, and the dangerous action count will increase, and the learning game will restart if they get closer to an elevator. The game is over if the safety percentage becomes 0% or if they do dangerous actions three times. If they succeeded in an evacuation, the result screen (Figure 40) will appear, and learners can see an evaluation. The best score is updated if the score of the learning game is higher than the current best score.



Figure 40. The Result Screen

If learners succeed in an evacuation, they select whether they play the learning game again or they go back to the title screen on the result screen. If they failed to evacuate, they select whether they play the learning game again or they go back to the title screen on the game over screen. After playing the game, they can check the achievement list (Figure 39) to achieve better evacuation on the next learning. Repeating this flow, learners can learn how to evacuate from the building No.5 repeatedly.

4. Preliminary Experimental Evaluation

4.1 Hypotheses

We have the following two hypotheses to verify the effectiveness of the Learning Game:

Hypothesis 1: Learning with a reproduced environment which is familiar to learners assist them in imagining what is going happen when an earthquake occurs.

Hypothesis 2: The game features can stimulate their learning.

4.2 Overview of the Evaluation

First, we show the manual of the learning game and explain what learners should do in this evaluation. Then, they put the HMD and practice usage of Oculus Touch with practice mode. When they think they practiced enough, they start the earthquake evacuation learning game. They play the learning game twice, regardless of succeeding to evacuate or not. After playing the learning game, learners answered

the questionnaire for the learning game on Google Form. The questionnaire is a four-grade evaluation; four is good, and one is poor. Twelve engineering students participated in this evaluation.

4.3 The Result and Considerations

4.3.1 The Result of the Questionnaire

Table 11 shows the results of the questionnaire. The numbers colored with red insist positive, and the numbers colored with blue insist negative. The totals may not become 100% due to rounding off.

Table 11
The Contents and Results of the Questionnaire

Questions	4pt	3pt	2pt	1pt
Q1: Could you imagine a situation that an earthquake happens at the place that is familiar to you by this experience?	0%	75%	25%	0%
Q2: Did you think that the things that you learned with this learning game are useful for real evacuation?	8.3%	41.7%	50%	0%
Q3: Did showing Safety Percentage and restart function assist you in learning inappropriate actions and dangerous actions?	33.3%	50.0%	0%	16.7%
Q4: Did you think you want to learn repeatedly by evaluation function and best score recording functions?	16.7%	66.7%	16.7%	0%
Q5: Did the achievement list assist you in learning with this game repeatedly?	8.3%	83.3%	8.3%	0%
Q6: Is the quality of the game (e.g., playability, user-interface, and so on) good?	25%	41.7%	25%	8.3%
Q7: Did you think the learning game made you think that you need to keep learning about disaster?	16.7%	66.7%	16.7%	0%

Moreover, some learners described the following opinions in the free describing field:

- It is better if you can make it clear where the exit is.
- I little wonder about going through the goal.
- I felt that the walking speed in the learning game is different from that in the real world. It should be a little faster.
- I wanted some hints like when we need to reset the learning game.
- It was not clear where the exit is, so it is better to use something like a maker to make it clear.
- It was not clear where the exit is. The walking speed is too slow so that I wanted to run while I know it is dangerous.
- It is better to make it clear where the exit is.

4.3.2 Consideration of hypothesis 1

Q1 and Q2 correspond to hypothesis 1. Q1 is a question to check the environment that is familiar to learners reproduced with VR encourages learners to imagine what happens to the familiar place if an earthquake occurs, and Q2 is to check that learners learned with this learning game are useful for actual earthquake evacuation. 75% of learners answered yes on Q1, while the ratio of yes and no is half and half. We consider why we got the answer on Q2; this is because the reproduced building No.5 is not perfectly reproduced and the learning game does not have the function to provide let the learners learn the features that people have to do in actual earthquake evacuation such as rescuing wounded people and appropriate actions when firing and/or blackout happens.

From above, the learning game can encourage learners to imagine what happens to the place that is familiar to them if a big earthquake occurs, but learners cannot learn actual behaviors that they have to do in earthquake evacuation. Thus, hypothesis 1 is partially supported.

4.3.3 Consideration of hypothesis 2

Q3, Q4, and Q5 correspond to hypothesis 2. Q3 is the question to check that showing Safety Points and restart function assisted learners in learning inappropriate action and dangerous action. 80% of learners answered that those features assisted their learning. Q4 is to check evaluation and best score recording functions encourage to learn repeatedly and Q5 is to check the achievement list encourages to learn repeatedly. Nearly 85% of learners had positive answers on Q4 and nearly 90% had positive answers on Q5. Thus, it is suggested that the game features stimulate learning. Thus, this result can support hypothesis 2.

4.3.4 Other Discussion from the Results on the Evaluation

From the result of Q6, only 60% of learners had positive answers. This is not as high percentage as other question except Q2. Besides, from the opinions in the free description area, they described the negative opinions about the quality of the learning game. This result means that the learning game should be refined to have better quality. Also, this results in the lower satisfaction on Q2. However, 75% of learners had positive answers on Q1 even the learning game does not have good quality. These results may lead that reproducing the place that is familiar to the learners and developing the environment that they can practice earthquake evacuation on their familiar place are effective to encourage learners to imagine a situation when a big earthquake occurs even the quality of the learning game is not necessarily good. We consider they could have achieved earthquake evacuation learning on the local and individual level.

Considering the result of Q7, we thought Q7 was too general so that the question does not evaluate any portions of the learning game. However, this question asks that the learning “game” can motivate learners. Thus, We consider that this result shows the effectiveness of Game-based Learning.

5. Conclusion

In this research, we prototyped an earthquake evacuation learning game with VR. The learning game aims to learn earthquake evacuation on the environment where learners are familiar. The learning game has game factors: safety percentage and dangerous action count which encourage learners to learn inappropriate actions on earthquake evacuation, the restart function, the evaluation function, the achievement list. As future work, we will add more situations that may happen in earthquakes such as fire, blackout, and rescuing wounded people. Besides, physical movement in the learning game is not perfect. We will handle both problems of learning contents and game quality.

References

- Freina, L., Ott, Michela. (2015). A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. *Proceedings of the 11th International Scientific Conference eLearning and Software for Education*, 133-141.
- Krokos, E., Plaisant, C., & Varshney, A. (2018). Virtual memory palaces: immersion aids recall. *Virtual Reality*, 1-15. doi: <https://doi.org/10.1007/s10055-018-0346-3>
- Li, C., Liang, W., Zhao, Y. & Yu, L.-F. (2017). Earthquake Safety Training through Virtual Drills. *IEEE Transactions on Visualization and Computer Graphics*, 23(4), 1388-1397.
- Mitsuhara, H. (2018). The Present and Future of ICT-Based Disaster Education Systems. *Transactions of Japanese Society for Information and Systems in Education*, 35(2), 66-80. (in Japanese)
- Perrotta, C., Featherstone, G., Aston, H. & Houghton, E. (2013). Game-based Learning: Latest Evidence and Future Directions (NFER Research Programme: Innovation in Education). Slough: NFER.
- UN Office for Disaster Risk Reduction (UNDRR). (2015, March 28). Sendai Framework for Disaster Risk Reduction. Retrieved from https://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf
- Ruffino, P. A., Permadi, D., Mahadzir, M.B., Osello, A., & Aris, A. B. (2018). Simulation and Serious Game for Fire Evacuation Training. *Proceedings of 17th International Conference on Computing in Civil and Building Engineering*.

Research on Application of Steam Teaching in Primary School Based on Gamification Task Orientation-Taking the Fifth Grade Campus Sandbox as An Example

Xue-fei DING^a, dong-ming QIAN^b & meng-ni FENG^a

^a Department of Education Information Technology, East China Normal University, China

^b National Education Macro Policy Research Institute, East China Normal University, China

*ding-xuefei@qq.com

Abstract: Steam education integrates multiple disciplines for education and teaching. It is an educational concept and measure for international education to explore talent training in the 21st century. This paper uses the literature analysis method to sort out the current research status and hotspots of the current Steam class design and teaching application. It is also proposed to combine the strengths of gamification teaching, and to drive the project-driven and task-oriented design of the Steam lesson to be closer to the game, whether or not it contains modern technology. Combined with relevant theories, this paper attempts to construct a design model of steam teaching activity based on gamification task orientation, and describes the design and practice process of the successful case of the fifth grade primary school campus sand table. After the course, the questionnaire and interview method will be used to understand the teaching effect of the steam class. The results show that the teaching mode is effective in primary school classroom teaching, which can enhance students' enthusiasm for learning and provide a better learning experience. It is expected that the case in this study will provide a reference for steam education researchers in practical exploration.

Keywords: Steam education; gamification teaching; K-12 field; instructional design

1. Introduction

The training goal of “disciplinary literacy” in China's current steam education has limited its development to some extent. Many international researchers have advocated that core literacy and subject literacy are put into the same position, and carried out some experiments. On the other hand, the gamification education in recent years has faced serious problems of “fragmentation” and “hollowing” due to the lack of physical curriculum. The teaching methods of cultivating the core literacy advocated by gamification teaching, such as fun practice, exploration and innovation, provide a reference for the improvement of steam education. According to the above situation, this paper proposes a steam education model based on gamification task orientation. It shifts its educational goal orientation from a disciplined literacy to a comprehensive literacy that combines discipline and practical collaboration, from general problem-solving ability to creative problem-solving ability.

2. Research Background

2.1 Overview of Research on Steam Teaching Practice at Home and Abroad

STEM education is a national development strategy proposed by the United States to cope with future social challenges [1]. Because it is in line with the direction of education reform, countries around the world have introduced stem or steam education to varying degrees. This paper focuses on the design of classroom examples and teaching practice, and uses the literature analysis method to sort out the current research status and hotspots. The literature is mainly collected through domestic and foreign literature databases, such as CNKI database, Taiwan Huayi database, EBSCO, ERIC, etc. The time is from 2009 to 2019. Keywords include “steam lesson”, “steam class”, “steam practice”, "steam application" and so

on. After screening, 120 articles with higher correlation were obtained, and the keyword co-occurrence network map was drawn, as shown in Fig. 1.

According to Figure 1, the co-word network of the entire keyword is not completely dense, and the secondary nodes are more prominent. The network is centered on steam and stem education, and it is surrounded by relatively obvious secondary node centers such as project learning, curriculum design, self-learning, empirical research, 3D printing lessons, scratch, flipping classrooms, and interdisciplinary integration. There are also three-level nodes based on science, biology, physics, geography, innovative thinking, and case teaching. To a certain extent, this shows that the research on steam education has a certain breadth and depth. The research hotspot is based on the project design research of science, physics, chemistry and other disciplines. On the other hand, it is the curriculum development to train high-level steam ability.

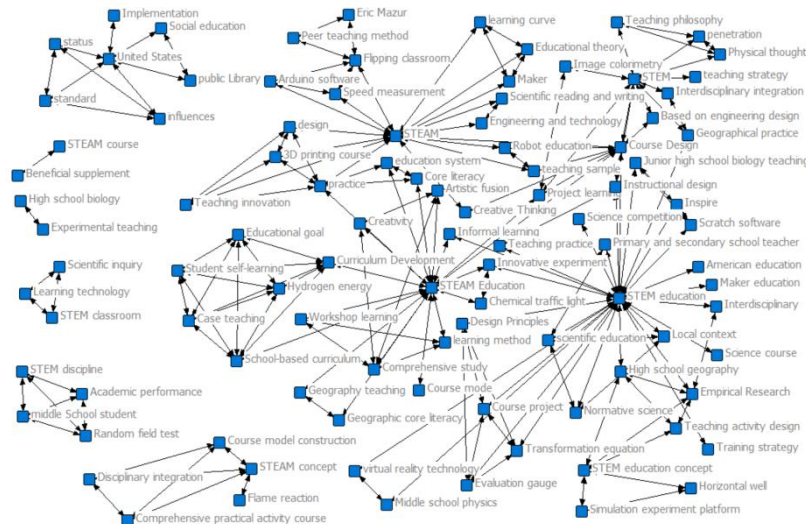


Figure 1. Keyword co-occurrence network diagram

2.2 The Origin of Gamification Teaching and the Integration with Steam Education

In ancient times, the relationship between education, life and games was close, but with the introduction of the class system in Comenius, education and games gradually became opposed [2]. Educational researchers are beginning to seek change. German educator Froebel suggests that early childhood education should be combined with games. Italian educator Montessori believes that games are a necessary stage for early childhood development, and American educator Dewey believes that games should be part of the school curriculum. To sort out the development of gamification education, we find that researchers all over the world agree that games are important in the development of children.

China's steam education is more dependent on a specific subject, and relatively more emphasis on the cultivation of subject quality, it is easy to fall into the trap of the surface steam education. Combined with the above literature analysis, we found that there is a broad application prospect is the task orientation of life and gamification, so the team proposed to design the task of the Steam lesson to be closer to the game, whether or not it contains modern technology. At the same time, combined with the theory of steam teaching design, build a steam teaching activity design pattern.

3. Steam Teaching Design Related Theory

There are three parts of the theoretical thoughts related to steam instructional design. One is the learning theory represented by the theory of generative learning and the theory of learning in practice. Generative learning theory means that learning is mainly influenced by existing knowledge and experience, rather than by external stimulus or reinforcement [3]. The active generation process is different from the simple storage information, which emphasizes the characteristics of the learning subject and cognitive strategies. The theory of learning in practice is the theory put forward by the educator Dewey on the basis of the ideas of “education is life” and “education is the transformation of experience” [4].

The second is the design theory of steam teaching activities based on project-based learning, group cooperative learning, and 6E teaching mode. Project-based learning provides students with the questions they need to solve in real situations, helping to understand science, technology, engineering, and mathematics [5]. Group-based collaborative learning is a mutual-assisted learning in which students have a clear division of labor in order to accomplish common tasks. The 6E learning model divides the STEAM teaching process into attraction, inquiry, interpretation, design, development and evaluation, focusing on integrating the engineering design process into classroom teaching [6]. The third is the idea of gamification teaching, which is very compatible with the characteristics of the mobile Internet, such as diversification, experience, fragmentation and so on. It can stimulate learning motivation, promote cognitive development, and develop high-level steam skills such as problem solving, collaborative learning, and creativity [7].

4. Construction of Steam Activity Design Pattern based on Gamification Task Orientation

The design model of steam teaching activity based on gamification task orientation is constructed by four aspects. The first aspect aims to cultivate a comprehensive quality that combines the discipline and practical cooperation. At present, the focus of the steam education reform in all countries falls on the core literacy. The early European Union proposed the eight literacy of lifelong learning, and the United States proposed the core skills of the 21st century, as shown in Figure 2 [8]. In 2016, China proposed six core literacy: responsibility, practical innovation, humanistic heritage, scientific spirit, learning to learn and healthy living. After the literature and field research, we decided to put the cultivation of basic subjects such as writing, reading and calculation into an important position in combination with the situation of local schools, and at the same time increase core literacy, such as induction, reflection and other learning literacy; classification, spatial imagination, Modeling and other abstract thinking; innovative creativity, etc. The second aspect is to design a reasonable fun game as a task-driven and analyze its rationality. For example, analyzing learners' learning foundations and requirements based on generative learning theory, analyzing specific subject knowledge and core literacy, determining learning objectives from the perspective of three-dimensional goals, and finally analyzing feasibility, such as whether learning space, resources, etc. are sufficient to support the completion of practice.

In the third aspect, the learning style is learned in the practice of group cooperation. First, design teaching context and specific module teaching knowledge in order to ensure each training goal, and then complete the learning bracket design, including designer activities, learning expectations and solutions. In the fourth aspect, the evaluation method is a dynamic evaluation of discussion sharing and creative problem solving orientation. Based on the above, a design pattern of steam teaching activity based on gamification task orientation is drawn, as shown in Figure 3.



Figure 2. 21st Century Core Skills

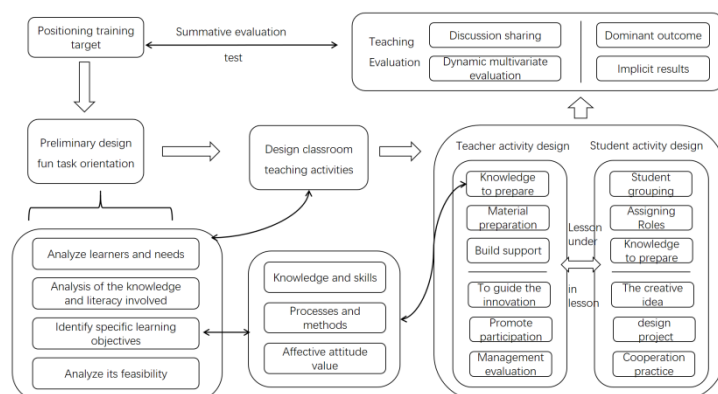


Figure 3. Pattern diagram

5. Steam Teaching Practice

5.1 Research Design and Teaching Process

Based on the established model, we designed four major themed steam classes—observing leaves, garbage sorting, origami art, and campus sandboxes. The course is about 2-3 class hours per week, and a theme class is completed in one month. The first round of practice took four months, targeting five fifth-grade classes in a city in Fujian Province, China. The second round of practice is in progress in Qingdao. The schedule is shown in Table 1. This article selects "Campus Sandbox" to elaborate.

The first lesson is observation and description. First of all, creating a game task situation leads to the theme of this section, inspiring interest and thinking about whether it can make a campus sand table for the alma mater, and then the teacher guides and emphasizes the importance of the project plan. Study the orientation description method through case study to test the learning effect. Finally, leave the after-school activities to draw a floor plan, and use the wrong floor plan to inspire and guide the importance of the ratio. The second lesson is measurement and calculation. First, use the case to explore new knowledge, and introduce the discussion of proportional calculation and application with maps. Then enter the group practice (steam course is based on the teacher's understanding of the classmates to divide the 5-6 heterogeneous study group), the teacher helps the students to divide the "work" of the teaching building, playground, etc., and measure the campus by means of step test. Then summarize and convert to model data.

The third lesson is a hands-on building. After the model data is compared, the construction starts. Each team member should be able to work with others to share work, exchange experience, help each other, evaluate, and solve various production and complex layout problems. Teachers can give advice and reference practices as appropriate. The fourth lesson is communication expansion. Each group displays the works of this group in turn, and evaluates each other whether the image is accurate and beautiful. And recommend students to share their own task list, reflect on defects, give advice. Finally, teachers use onion cells, planetary models, etc. to guide students to understand that from small to very large, modeling is an important idea and method.

Table 1 *schedule of steam experiment*

Phase	Time	Explore content
The first round	September to December 2018	Teaching design and implementation, teaching, activity design model feasibility verification.
The second round	March to June 2019	Model improvement and teaching case improvement.

The problem	Often	Sometimes	Never
1. Do you hope your group's campus sand table works can be exhibited in the class?			
2. In the process of sharing, do you want your works to be praised by your classmates and teachers?			
3. Do you want to be more dexterous?			
4. Do you think your ability to make works will improve if you practice hard?			
5. Do you try to make good work when other students are discouraged?			
6. Do you want to make your work more functional and creative because you want to be praised by the teacher?			
7. Do you work hard and want to be recognized?			
8. Do you want to get good comments in the production of campus sand table works?			
9. Do you want people to admire your work?			
10. Do you want to improve the level of the campus sand table works made by yourself?			
11. When you started making the campus sand table, did you think you could do it well?			
12. Can you overcome all kinds of difficulties in order to complete the work?			

Figure 4. Learning motivation questionnaire

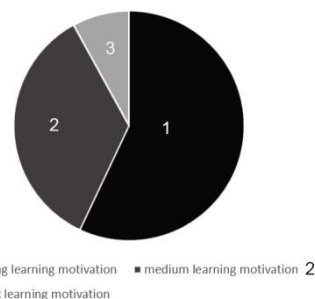


Figure 5. Learning motivation pie chart

5.2 Analysis of the Effect of Teaching Practice

5.2.1 Questionnaire Analysis of Learning Motivation

The Learning Motivation Scale was based on the revised learning motivation test conducted by Zhou BuCheng of East China Normal University in 1991. The modified subscale is shown in Figure 4. After completing the questionnaire online, 102 valid questionnaires were retrieved. The scale has good reliability and validity, and its half-correlation coefficient is 0.83-0.89, and the re-test correlation

coefficient is 0.79-0.86. Questionnaire options are never, sometimes, and often recorded as 1-3 points, so the total score is 12-36 points. The evaluation method below 21 points indicates that the learning motivation is weak, and the higher than 28 points indicates strong. Calculated to have a mean value of 28.1, that is, at a strong level of learning motivation, and draw a fan chart proportionally, as shown in Figure 5, it can be clearly seen that the gamification task drive basically makes the students have a more positive learning attitude, which is good for the steam course and the learning effect.

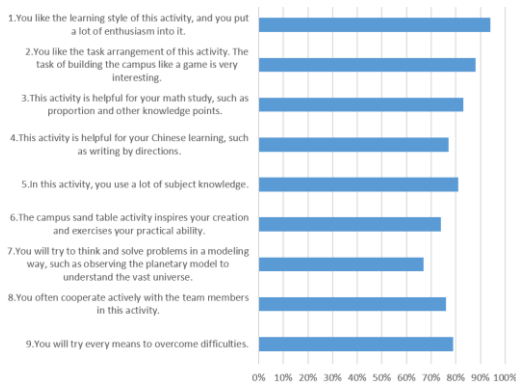


Figure 6. Learning situation summary graph

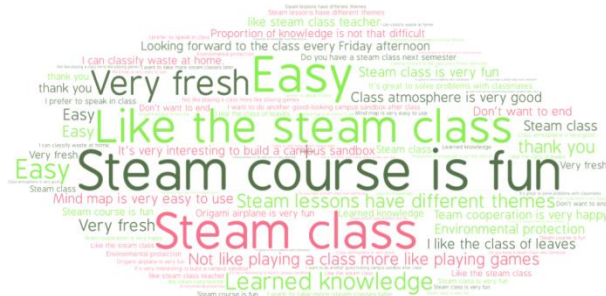


Figure 7. Keyword cloud

5.2.2 Questionnaire Analysis of Learning Situation

The team designed a questionnaire to investigate the student's steam course learning and results, using the Rickett scale to have a positive statement, from completely disagree to very agree with the five options, the score is 1-5 points. After testing, the reliability coefficient is $0.912 > 0.7$, and the validity coefficient KMO $0.902 > 0.6$, so the reliability and validity of the questionnaire is good. Count the percentage of the total number of people in each of the 121 valid questionnaires that reached 4 points and above, as shown in Figure 6. The proportion of consent for most topics reached 75%, indicating that students have a lot of experience and gains in the course, especially in terms of learning interest, participation, learning of subject knowledge, and ability to solve problems. The percentage of the two topics related to whether gamification task orientation can stimulate creative inspiration and develop modeling thinking is relatively low, but it is also close to 70%. It shows that this steam course has played a certain role in the above aspects, but because the cultivation of ability and literacy is a long-term process, it is difficult to show obvious effects in the case. In addition, we conducted interviews with students, and the results are shown in Figure 7, which is mutually validated with the results of the questionnaire.

Evaluators	N	Correlation coefficient	Sig
a&b	64	0.855	0

Figure 8. Correlation coefficient

group	The design and production have certain innovation	to meet certain needs, solve practical problems	the structure of the work is reasonable and complete	with certain aesthetics	clear ideas, complete content, clear expression	concise and smooth	Reasonable and clear division of labor	Clever design and smooth operation	score
A1	8	7	8	7	6	7	8	7	58
A2	6	8	8	6	8	6	8	6	56
A3	6	7	6	6	7	7	7	6	52
B1	8	9	8	9	8	8	8	7	65
B2	7	7	8	6	6	5	7	5	51
C	10	9	9	9	9	7	6	8	67
D	8	10	9	9	10	8	5	8	67
E	9	10	9	8	8	9	7	7	67
On average	7.75	8.375	8.125	7.5	7.75	7.125	7	6.75	60.375

Figure 9. "Campus Sand Table" works score

5.2.3 Course Work Evaluation

The team invited two graduate students who participated in the steam project as scorers, including the final eight works. Each aspect that can be evaluated is up to 10 points. A paired sample T test was performed on the two sets of data to obtain Fig. 8. The probability value was less than 0.001 and the correlation coefficient was positive, indicating that significant correlation and evaluation were

effective. The average score is taken and rounded off as the final score, as shown in Figure 9. Overall, the team that finished the work scored higher, with an average score of 60 or more (total score of 80), indicating that the completion was good. However, the gap between the groups is large. In combination with the specific teaching situation, most of the groups of 5-6 people complete the work, while the C, D, and E groups are all the whole class to complete a teacher's centralized guidance and gather the whole class, so the score is relatively higher, especially in terms of the integrity, aesthetics and presentation of the work. However, after communicating with the teachers, it was found that there were individuals in the three classes who contributed less to the group's works, which needs to be improved.

6. Reflection and Summary

Both the evaluation of the work and the interviews between teachers and students have verified the effectiveness of the gamification task orientation, but the student base is different, and some hands-on and collaboration skills are not high, so the design needs to be improved. For example, the preparation work should be designed more fully. In addition, regarding the group study, the students took turns to be the team leader, the spokesperson, the disciplinary officer, etc., and basically completed the task. However, it is necessary to pay attention to the targeted, disciplined and underachievers of the group activities. In general, this study focuses on the design and implementation of the steam class case, proposes the advantages of combining gamification teaching, and constructs the design pattern of teaching activities. The "Sandpaper Sandbox" is taken as an example to analyze the implementation process and effect. This is a scientific practice inquiry, including the stages of participation, exploration, sharing, expansion and evaluation, and forms a classroom mode in which groups and classes share, communicate and question. The research results show that the students' interest and experience are good, the classroom teaching effect is obvious, and the cultivation of disciplines and core literacy has achieved initial results. However, there are still shortcomings in the research. For example, only through questionnaires, there is a lack of discussion on deeper cognitive rules. The sample size is small, and long-term verification is needed to promote the improvement of core literacy. In the future, we will continue to explore steam classroom teaching based on gamification task orientation, and hope to contribute to the development of steam education.

Acknowledgements

This research was funded by the research project of the Department of Education of East China Normal University, "Application and Practice of STEAM Education in Weak Areas", and was completed in cooperation with team members. I would like to thank you.

References

- Jin Hui, Hu Yingqi. Leading the Future of Education with STEM Education Innovation——Interpretation and Enlightenment of the US STEM 2026: STEM Education Innovation Vision Report[J]. *Journal of Distance Education*, 2017, 35(1): 17-25.
- Shang Junjie, Yan Laisi. Reinventing learning style: the core educational value and application prospect of games [J]. *China Electro-education Education*, 2015 (05): 41-49.
- WITTROCK M C. Generative processes of comprehension [J]. *Educational psychologist*, 1989, 24(4): 345-376.
- Liu Guangli. Dewey's Teaching Theory of "From Middle School" and Its Enlightenment to China's Basic Education[J]. *Continuing Education Research*, 2008(3): 77-80.
- Robert. Project-based STEM learning [M]. Shanghai: Shanghai Science and Technology Education Press, 2016: 15-16.
- Burke B N. The ITEEA 6E Learning By Design™ Model: Maximizing Informed Design and Inquiry in the Integrative STEM Classroom[J]. *Technology and Engineering Teacher*, 2014, 73(6): 14-19.
- Liang Lin. Problems in sports gamification teaching and its countermeasures [J]. *Teaching and Management*, 2018 (21): 113-115.
- Yan Liyuan. Development of Mathematical Activities Based on STEAM Education Concept [D]. Shanghai Normal University, 2018.

InNervate AR: Mobile Augmented Reality for Studying Motor Nerve Deficits in Anatomy Education

Margaret COOK^{a*}, Austin PAYNE^a, Jinsil Hwaryoung SEO^a, Michelle PINE^b & Timothy MCLAUGHLIN^a

^a*Department of Visualization, Texas A&M University, United States*

^b*Department of Veterinary Integrative Biosciences, Texas A&M University, United States*

*atmgirl@email.tamu.edu

Abstract: Augmented reality applications for anatomy education have seen a large growth in their literature presence as an educational technology. However, the majority of these new anatomy applications limit their educational scope to the labelling of anatomical structures and layers, and simple identification interactions. There is a strong need for expansion of augmented reality applications, in order to give the user more dynamic control of the anatomy material within the application. To meet this need, the mobile augmented reality (AR) application, *InNervate AR*, was created. This application allows the user to scan a marker for two distinct learning modules; one for labelling and identification of anatomy structures, the other one for interacting with the radial nerve as it relates to the movement of the canine forelimb. A formal user study was run with this new application, which included the Crystal Slicing test for measuring visual spatial ability, the TOLT test to measure critical thinking ability, and both a pre- and post- anatomy knowledge assessment. Data analysis showed a positive qualitative user experience overall, and that the majority of the participants demonstrated an improvement in their anatomical knowledge after using *InNervate AR*. This implies that the application may prove to be educationally effective. In future, the scope of the application will be expanded, based on this study's analysis of user data and feedback, and educational modules for all of the motor nerves of the canine forelimb will be developed.

Keywords: Anatomy Education, Mobile Augmented Reality, Educational Technology

1. Introduction

Due to the increased accessibility of educational technologies, the higher education anatomy curriculum has seen rapid reformation (Biassuto et al., 2006). Traditionally, anatomy courses are primarily taught with the methods of didactic lectures and cadaver dissection. The anatomy classroom teaching materials are characterized by static and two-dimensional images. Laboratory involves dissection guides, animal cadavers, and aids such as plastinated anatomical models (Peterson, 2016). However, decreased laboratory funding and laboratory time, and increased technology development, have led to limiting animal use to only teaching procedures which are considered essential (King, 2004; Murgitroyd et al., 2015; Pujol et al., 2016). With the evolution of learning theories in the classroom, as well as the growth of 3D interactive technology, there is a need for those who work in the anatomy higher education field to re-examine the learning tools that are used in anatomy courses (Azer & Azer, 2016).

One of several new trends to emerge in anatomy education technology is mobile augmented reality (Mobile AR) applications. Augmented reality is defined as a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view. Mobile AR has a large potential to serve in education, as it can make the educational environment more engaging, productive, and enjoyable. Furthermore, it can provide a pathway for students to take control of their own learning and discovery process (Lee et al., 2012). However, the majority of anatomy AR applications focus primarily on labelling of anatomical structures and layers, or simple identification interactions (Jamali, 2015, Kamphuis, 2014, Ma, 2016). It is important that anatomy content in Mobile AR be expanded from simple identification questions, and labelled three-dimensional structures.

Mobile AR allows students to dynamically interact with digital content that is integrated with current print-based learning materials.

We developed a mobile AR application called *InNervate AR* for smart mobile devices, which explores the selected topic: deficits to canine muscle movement, in response to motor nerve damage. This is a concept that is often frustrating to undergraduate anatomy students because it requires mental visualization of the anatomical structures involved. *InNervate AR* also supports students' critical reasoning skills, while learning about motor innervation of the canine limb. This paper describes the development of *InNervate AR*, as well as the user study to test its effectiveness as an anatomy learning tool.

2. Application: InNervate AR

2.1 Design

InNervate AR includes two learning sections, which are incorporated in user study handouts. Canine cadavers are used in many undergraduate school anatomy courses, as well as the course that participants in this study have used in their coursework. Therefore, the anatomy of the canine was used in this anatomy mobile AR application.

- Section 1 involves labelling and identification of the structures of the canine thoracic limb. This was purposefully developed to make sure that *InNervate AR* offered the same baseline tools for user experience and learning as the existing anatomy applications that are available. Figure 1-1 shows the view of the participant during both learning modules.

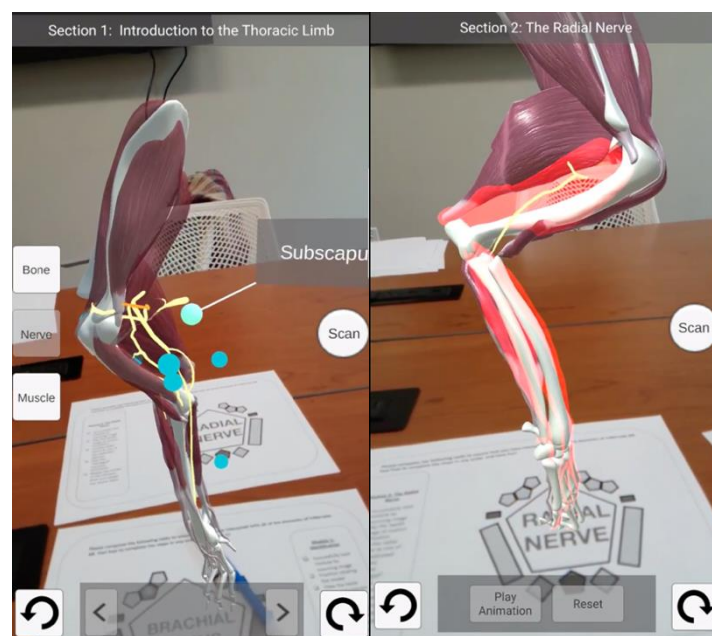


Figure 3-41 The view of *InNervate AR* participant during the labelling module (left) and the radial nerve animation module (Right).

- In Section 2, the user dynamically interacts with a musculoskeletal system of the canine thoracic limb, and plays animations of a healthy canine limb's range of movement. They can then visualize "damage" to different areas of the nerves of the limb, and be educated on what deficits exist. The "damage" is cuts to the nerve with the swipe of a finger on the device screen, and the resulting muscle action deficits are displayed with before and after animations of the muscles' ability (or inability) to move. Thus, the user can explore different combinations of affects upon the anatomy, and become more actively engaged in the educational process of the mobile AR application.

2.2 Development

For the radial nerve animation module, a total of five animation sequences were created. The first animation sequence was the entire healthy range of motion of the canine thoracic. The other four scenarios involved changes in movement capabilities of the limb, based on the motor innervation provided by the radial nerve. These four radial nerve scenarios represented different possibilities of damage that could have occurred to the radial nerve. Due to the infinite number of possible damage scenarios to an organic animal's nerves, the number of nerve damage scenarios was narrowed down to a more finite set of 4 ranges. These 4 ranges would produce the most visually distinctive results between each of the scenarios. This was done so that the scenario possibilities would not overwhelm the user.

InNervate AR was designed as a marker-based system with Google ARCore software, utilizing image recognition developments from Viro Media. This means that the camera of the mobile device detects a shape on a piece of paper, known as the marker, and then the application loads the programmed learning module that corresponds to that marker (see Figure 3-2).

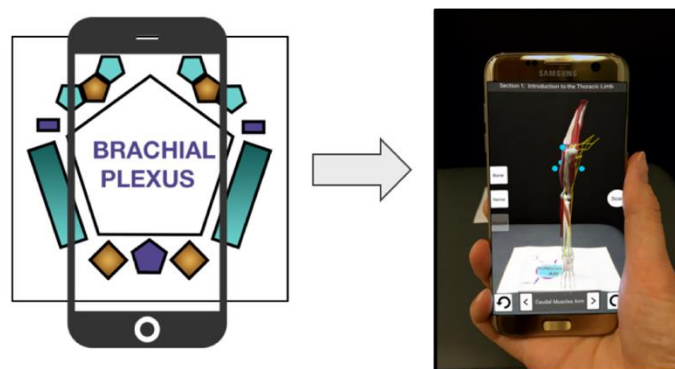


Figure 3-42 The process of image recognition to load the InNervate AR application

2.3 Learning Scenario

Anatomy students struggle with combining several layers of their knowledge together to make logical conclusions about motor nerves and their relationship to the muscles which they innervate. An example of this difficulty is when the students are asked to answer an exam question about which muscle movement deficits would exist based on the information provided about an injury to a specific section of the thoracic limb. When answering that question, the student has to complete several mental steps. First, *they must correctly mentally visualize the muscles and nerves of the thoracic limb*. Next, *they must recall which motor nerves are located in the injured section of the thoracic limb*. Afterwards, *they must recall which muscles are specifically innervated by the motor nerves in that area of the thoracic limb*. By processing that information, they can recall what the actions of those muscles are, and then describe which muscle movements will be impaired. The final consideration that they must make is if the nerves which were damaged continued further down the limb, because if so, then further deficits might exist distally due to the linear relationship between nerve signals and the muscles that they communicate with.

InNervate AR was designed to give students a learning platform for seeing a 3D representation of these clinical reasoning scenarios. The AR technology allows the students to view all of the anatomical structures together, and then actually see how they work together when healthy, or become impaired with damage.

3. User Study

3.1 Study Procedure

Students from the a Texas A&M physiology course were recruited for a user study with InNervate AR. This course is the class that Texas A&M students are required to take in their degree plan after completion of their required anatomy course. Each participant was given 90 minutes maximum to complete the activities of the user study. First, the participant was asked to complete a pre-activity

questionnaire. The participant was then asked to complete the timed Crystal Slicing Test (Ormand et al., 2013). They had 3 minutes to complete the test. The participant was next asked to complete the 38 minute Tobin and Capie 1981 TOLT (Test of Logical Thinking) test (Tobin & Capie, 1980).

After completion of the TOLT test, the participant was provided with a mobile device (SAMSUNG Galaxy) and a corresponding paper handout for how they were to proceed with interacting with InNervate AR. This handout asked them to perform specific tasks, in a defined sequence, in order to ensure that the user had interacted with all parts of the application. The handout had a place for them to check-off when they had completed a task within the application. This handout also had image markers that InNervate AR could scan, to bring up the different learning modules that are built into the application.

The participant's duration of use of the application was recorded. The participant was free to ask the user study facilitator questions about navigation of the application. While the participant was using the application, another mobile application on the same device was recording the screen of the device. The participant's interaction with the mobile AR application was also recorded on video with a camera. After completing their interaction with InNervate AR, the participant was asked to complete a post-activity questionnaire.

Within the quasi-experimental design of this study, the non-equivalent groups design was followed. This means that no randomized control group exists, but a pre- and post- test is given to groups of people that are as similar as possible, in order to determine if the study intervention is effective or not. The pre- test was written to include anatomy knowledge questions, a free response question, demographics questions, and Likert-Scale based questions about their anatomy education experience. The post- test was written with five knowledge-based questions, three of which mirrored the anatomy knowledge questions of the pre-test, with the same concept being asked in a different way. The post-test also included Likert-Scale based questions about their experience with the InNervate AR system, as well as place to write additional feedback. The objective of these questionnaires was to obtain quantitative data based on the anatomy knowledge questions, and qualitative data based on the Likert and free-response questions.

3.2 Data Analysis

The data from all of the user study participants was compiled and analyzed for patterns and trends. This involved grading the users' performance on the learning instruments used in this study, and identifying similarities and differences in the qualitative answers given during the pre and post questionnaires. We specifically wanted to identify how the critical thinking scores, and the visual spatial ability scores of the users effected their change in performance on the anatomical knowledge questions after using *InNervate AR*. Furthermore, the screen recordings and video recordings of the study were reviewed to analyze the overall user experience of the study participants.

4. Results & Discussion

4.1. Participant Demographics

There was a total of 22 participants in the user study for the *InNervate AR* application. All of the participants were Biomedical Sciences majors at Texas A&M University, and had taken the TAMU VIBS 305 Biomedical Anatomy course within the two previous academic years. Five of the participants were male, and 17 were female. When asked, 18% of these participants answered "Strongly Agree" and 59% of them answered "Agree" to the statement "I consider myself to have a high level of critical thinking ability."

4.2. Participant Crystal Slicing Test Results

The highest possible score that a participant could make on this 3-minute test was 15 points. Only 9.09% of participants scored a 10 or better on this test. The majority of the user study pool (54.55%) made a score in the point range of 7-9. The next most common point range (22.73%) was a score of 5 or 6. The remainder of the participants (13.64%) scored less than 5 points. This data demonstrates that the participants in this user study had average or low visual spatial ability in general.

4.3. Participant Test of Logical Thinking Results

With an allotted time of 38 minutes, the highest score that a participant could make on the TOLT was 10 points. A perfect score of 10 was made by 40.91% of the participants. A score of 9 was achieved by 31.82% of the participants. A score of 8 was made by 9.09% of the participants. Only one participant (9.09% scored a 7 on the test. The remaining participants (13.64%) scored a 6 or lower on the TOLT. This data showed that the participants trended toward having high critical thinking skills.

4.4. Participant Anatomical Knowledge Scores Results

In the pre-questionnaire, the participants had 3 anatomical knowledge test questions. In the post-questionnaire, the participant had 5 anatomical knowledge test questions, 3 of which were matched to the pre-questionnaire test questions. In other words, the same content was tested on in those 3 questions, but asked in a different way. The scores of the participants were analyzed, and 77.27% of the participants' scores improved on the 3 matched questions, after using the *InNervate AR* application. 18.18% of the participants made the exact same score on the matched anatomy questions, and 4.55% of the participants had a lower score in the post-questionnaire on the 3 matched questions. This data shows that the majority of the user study participants showed an improvement in their performance on the matched anatomy knowledge questions in the post-questionnaire.

4.5 User Experience with *InNervate AR*

In the post-study questionnaire, a series of Likert-Scale questions were asked in regards to the participants' perception of *InNervate AR*. The users' responses to these questions were all positive. Some of the categories they were asked to rate were: the usefulness of the application as a visual aid for the spatial relationships between the anatomical structures, the flow and user interface of the application, and the usefulness of the application to practice critical reasoning scenarios.

The participants were also given free response questions. The first question asked: "What did you like least about the *InNervate AR* application?" Common themes to how the participants answered this question included: no "zoom" feature, problems with how to cut the nerves, and problems with selecting the labelling spheres. The second question asked was "What did you like most about the *InNervate AR* application?" The most frequent responses to this question included: getting to visualize the actions of the muscles with the animations, the graphic aesthetic of the application, how easy the application was to use, and the accuracy of the anatomical content. The last free response question asked the participants if they had any further suggestions about the *InNervate AR* application. The responses included adding the ability to compare the healthy and damaged animation scenarios side-by-side, adding even more details about the muscles in the labelling module, and further customizing the visual user interface of the tool.

Finally, some of the verbal comments of the participants during their use of the *InNervate AR* application were:

"Oh man I wish I had had this when I was in anatomy lab...because it really connects it all together, especially with all of the bones articulating and everything being there. I remember having to draw so many layers. (ID: 1001)."

"This is super helpful, I just can't get over it, it's one thing to see the words on paper, but to see a cut branch! (ID: 1001)."

"Nice way to look at the anatomy from different angles... Most apps don't have what would happen if something is wrong, they just have the structures (ID:1009)."

5. Conclusions and Summary

The goal of this project was to create a mobile anatomy AR application, *InNervate AR*, which provides more dynamic interactions than other mobile AR applications that have been previously created. This mobile AR technology is innovative because rather than having another simple viewing interaction and

labelling interface, the user was able to take a more interactive roll in what information was being presented by the application.

The results of this user study showed an extremely positive response from the participants, both in their qualitative feedback data, as well as their anatomical knowledge improvement. The majority of the participants tested for a high critical thinking ability, and there was one student with an average or low visual spatial ability. Therefore it was difficult to investigate how the base critical thinking ability could impact on learning anatomy using mobile AR. In terms of spatial visualization, there was no significant difference between high spatial visualization students and low spatial visualization students. The qualitative feedback from the participants demonstrated areas where the *InNervate AR* application could use improvement, such as problems with how to cut the nerves, and problems with selecting the labelling spheres. However, the responses from participants were overwhelmingly positive in many categories. They enjoyed getting to visualize the actions of the muscles with the animations, the graphic aesthetic of the application, how easy the application was to use, and the accuracy of the anatomical content.

It is planned to use the data and feedback from this study as a guideline while further expanding *InNervate AR* to include all of the motor nerves of the limb as learning modules. Any future user studies will be completed in a classroom setting, so that a larger participant population can be guaranteed, and statistically significant results can be achieved. Furthermore, the limitations such as a low number of matched anatomy knowledge questions, and gender bias will be addressed. Future user study and application design will also be more error tolerant, so that user errors with the technology, or differences in user background will not have huge consequences when analyzing results (Rouse, 1990).

This study was a wonderful learning opportunity because it showed the great potential that *InNervate AR* has for anatomy higher education, and brought to light what weaknesses in the technology and research study design should be worked on in the future. It our hope that this initial push for expansion of anatomy content in mobile AR will hopefully encourage other researchers to add additional interactive content to their educational tools, and strengthen the presence of this technology in higher education anatomy curricula.

Acknowledgements

We would like to thank all of the contributors who have worked on the project. We also appreciate support from the Department of Visualization at Texas A&M University.

References

- Biassuto, S. N., Caussa, L. I., & Criado del R  o, L. E. (2006). Teaching anatomy: Cadavers vs. computers? *Annals of Anatomy*, 188(2), 187–190.
- Peterson, D. C., & Mlynarczyk, G. S. A. (2016). Analysis of traditional versus three-dimensional augmented curriculum on anatomical learning outcome measures. *Anatomical Sciences Education*, 9(6) 529-536.
- King, L. A. (2004). Ethics and welfare of animals used in education: An overview. *Animal Welfare*, 13(SUPPL.), 221–227.
- Murgitroyd, E., Madurska, M., Gonzalez, J., & Watson, A.:3D digital anatomy modelling - Practical or pretty? *The Surgeon*, 13(3), 177–180 (2015).
- Pujol, S., Baldwin, M., Nassiri, J., Kikinis, R., & Shaffer, K. (2016). Using 3D modeling techniques to enhance teaching of difficult anatomical concepts. *Acad Radiol* 23(4), 193–201.
<https://doi.org/10.1016/j.molmed.2014.11.008.Mitochondria>
- Azer, S. A., & Azer, S. (2016). 3D Anatomy Models and Impact on Learning: A Review of the Quality of the Literature. *Health Professions Education*, 2(2), 80–98.
- Lee, K.: Augmented Reality in Education and Training. *TechTrends*, 56(2), 13–2 (2012).
- Jamali, S. S., Shiratuddin, M. F., Wong, K. W., & Oskam, C. L. (2015). Utilising Mobile- Augmented Reality for Learning Human Anatomy. *Procedia - Social and Behavioral Sciences*, 197(February), 659–668.
- Kamphuis, C., Barsom, E., Schijven, M., & Christoph, N. (2014). Augmented reality in medical education? *Perspectives on Medical Education*, 3(4), 300–311.
- Ma, M., Fallavollita, P., Seelbach, I., Von Der Heide, A. M., Euler, E., Waschke, J., & Navab, N. (2016). Personalized augmented reality for anatomy education. *Clinical Anatomy*, 29(4), 446–453.
- Ormand, C. J., Shipley, T. F., Tikoff, B., Manduca, C. A., Dutrow, B., Goodwin, L., Hickson, T., Atit, K., Gagnier, K. M., & Resnick, I. (2013). Improving Spatial Reasoning Skills in the Undergraduate Geoscience Classroom Through Interventions Based on Cognitive Science Research. Talk presented at the AAPG Hedberg Conference on 3D Structural Geologic Interpretation.
- Tobin, K. G., & Capie, W. (1980). *The Test of Logical Thinking: Development and Applications*.
- Rouse W.B. (1990) *Designing for Human Error: Concepts for Error Tolerant Systems*. In: Booher H.R. (eds) Manprint. Springer, Dordrecht

Enhancing Computer Assembly Skill Using Virtual Computer Laboratory in Conjunction with Team Game Tournament Method

Acharaporn KHANTHUSAENG^a & Nadh DITCHAROEN^{b*}

^aMajor in Science Education (Computer), Faculty of Science, Ubon Ratchathani University

^bDepartment of Mathematics, Statistics and Computer, Faculty of Science,
Ubon Ratchathani University

*nadh.d@ubu.ac.th

Abstract: The aim of this research was to develop the computer assembly skill using virtual computer laboratory in conjunction with team game tournament (TGT) method. We exploited virtual reality technology in developing virtual computer laboratory which was designed and developed as an interactive media in form of 3D animation model displayed in 360 degree perspective with sound effect so that learners are able to get experience like being in a real environment. TGT was used in our learning management process which motivated learners to encourage, help each other to solve problem, enjoy to learn and challenge from the activities in class. Research instruments consisted of VR-based virtual computer laboratory, pretest-posttest, questionnaires and lesson plan. By purposive sampling, the participants in our preliminary study were 12 high school students who had low pretest score. The data were analyzed by mean, standard deviation, and t-test dependent. The results showed that the average posttest achievement score was statistically higher than that of pretest at significant level of .05. The average normalized gain of class was 35.83 percent which was at the medium level. In addition, the average of overall students' satisfaction was 4.29 (SD=0.60) which was at the high level.

Keywords: Virtual Laboratory, Virtual Reality, Computer Assembly, Team Game Tournament

1. Introduction

Computer assembly is a large part of a computer skills needed a logical, methodical manner when working with computer components, thus it is necessary to improve computer assembly skills dramatically with practice (Cisco, 2006).

Computer subject has been added, by the basic education core curriculum of Thailand, to be a parts of science area curriculum since 2017. The purpose of this subject is to develop students' knowledge and understanding about technology for everyday living (Ministry of Education, 2017). As, to know technology is not only knowing the way to use it, even so, to know the way to maintenance it is also important. In Thailand, some secondary schools who foresee this advantage also provided computer maintaining session for students along with general computer class.

In fact, the high school in Ubon Ratchathani, Thailand has placed computer maintenance subject in 12th grade curriculum in order to prepare first step of computer assembly knowledge and to arrange fundamental skill in computer maintaining for their students (Course description, 2018). With this kind of learning, the school certainly needs workshop class that allows student working on real computer devices. From on-site class experiences, regarding fresh trainees and pieces of computer component are highly electrical sensitive, the practical of students work on real equipment caused damages in almost all experimental accessories. Since the price of computer devices are expensive and the limited of school budget, a teacher needs to use the crashed components for students practicing and it led to the cause of students unable to learn computer assembly as good as it need.

From the problems mentioned above, the cutting-edge technology can help to simulate real event and offer ability to virtually engage activity. The demonstration method called virtual reality (VR) technology allows user to see and manage 3D images or objects by wearing head-mounted device or sensor glove in virtual environment (Malithong, 2000). Using this technology, the teacher does not

need to teach computer assembly subject by real hardware, on the other hand, they can use VR technology to enhance their classes without limitation of hardware damages.

Moreover, it also needs learning method for helping students to understand content and to engage in study. Since computer equipment is sophisticated electronic component, high school students who have no basis in this knowledge will be difficult to overcome subject matter which lead to boredom and undesirable learning. Team Game Tournament (TGT) is a learning method which provided fun learning activities that can be used for producing positive outcome on social, attitude, and academic performance dimension (DeVries, 1980). It manages students in classroom into small groups. Each group has member from all achievement level. TGT is a part of cooperative learning, which motivate learners to encourage and help each other master collaboration, critical thinking and communication skills required in the 21st century skills (Battelle for Kids, 2019).

Therefore, in this study, we integrated the using of VR technology with the advantage of TGT to enhance computer assembly skills of high school students. The VR-based virtual computer laboratory was developed to allow learners to explore inside components of a computer, and to compose each component to get the complete result of the virtual computer by learning activities with elaborating the collaborative from teamwork learning.

2. Methodology

2.1 Populations and Sample Group

30 students from the high school in Ubon Ratchathani, Thailand, who were studying in science classroom program on the first semester of the academic year 2019 were asked to take the pretest. The 12 students who had low pretest scores were selected to be our sample group.

2.2 Research Instruments

2.2.1 Virtual Computer Laboratory

In order to allow students to learn about functions of computer devices and practice their skills leading to enhancement of their computer assembly skills, we designed and developed the VR-based virtual computer laboratory by using the Oculus Rift and Oculus Touch (www.oculus.com) to display and interact with 3D models in virtual environment. In the virtual laboratory, it's consisted of 3 sections including learning center for learning of the devices (Figure 1 (a)), computer assembly practice (Figure 1 (b)), and exercises (Figure 1 (c)).

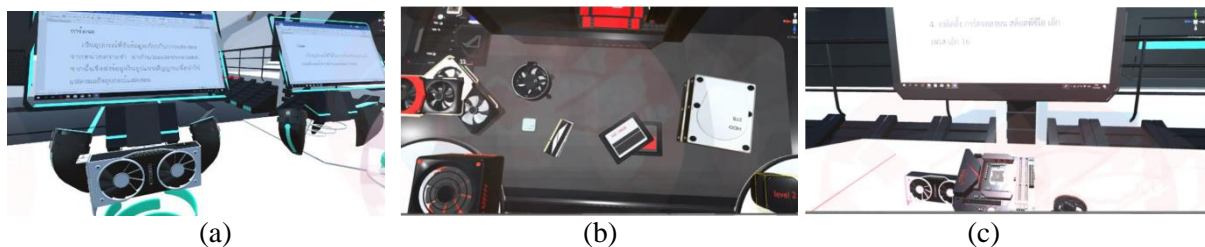


Figure 1. VR-based virtual computer laboratory (a) learning center (b) computer assembly practice (c) exercise

2.2.2 Teaching Plan

Teaching plan was developed based on TGT approach that consisted of 3 steps (Team, Game, and Tournament). In team step, students are divided into groups of 3-4 members. Each group has members from all levels of achievement (high, medium and low-achieving) which is categorized by their pretest score. In the game step, each group will choose the member to assign in tournament table according to member performance. The tournament step allows student do table activities to find the winner in each table and bring their score to be a group score. Our VR-based virtual computer laboratory will be used

during the game and tournament steps. Last step, praise the success, the teacher summarizes each team's score and gives the reward to the winner group.

2.2.3 Pretest-Posttest

Pre- and post- tests were designed based on concept of Bloom (1976) in terms of an objective test and multiple-choice items. There are 20 questions which were divided into 2 different sets (pre- and post-tests). Some examples are shown in Figure 2. They were developed using the Google Forms.

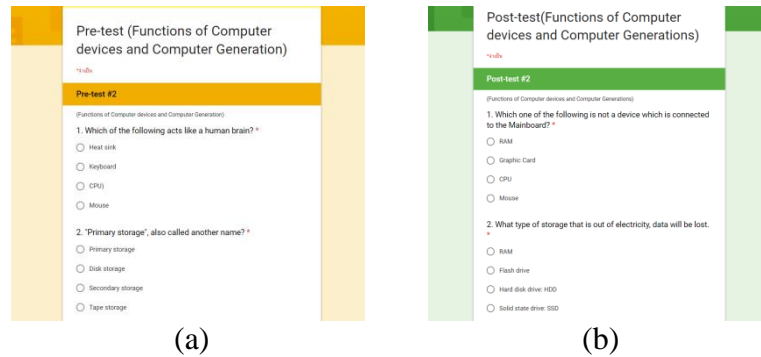


Figure 2. Examples of tests for evaluating students' understanding of functions of computer devices and computer generations (a) pre-test (b) post-test

2.2.4 Questionnaire

There were 11 items in questionnaires used for examine students' attitude toward the VR-based virtual laboratory with TGT. It was developed according to 5-point Likert scale including strongly disagree = 1, disagree = 2, neither agree nor disagree = 3, agree = 4, strongly agree = 5.

2.2.5 Computer Assembly Skill Assessment Forms

There are two skill assessment forms designed according of scoring rubric (Hart, 1994; Pickett & Dodge, 2001) including 5 scales: excellent – 5, good – 4, average – 3, poor – 2, and weak – 1 as illustrated in Figure 3. These forms were used to evaluate students' computer assembly skills.

Criteria	Scale				
	5 (Excellent)	4 (Good)	3 (Average)	2 (Poor)	1 (Weak)
Computer Assembly Level 1	Able to complete the computer assemble and completed on time.	Able to assemble more than 6 computer devices	Able to assemble more than 4 computer devices	Able to assemble more than 2 computer devices	Able to assemble less than 2 computer devices

Criteria	Scale				
	5 (Excellent)	4 (Good)	3 (Average)	2 (Poor)	1 (Weak)
Computer Assembly Level 2	Able to complete the computer assemble and completed on time.	Able to assemble more than 8 computer devices	Able to assemble more than 6 computer devices	Able to assemble more than 4 computer devices	Able to assemble more than 2 computer devices

Figure 3. Computer assembly skill assessment form (a) for level 1 (b) for level 2

2.3 Implementation

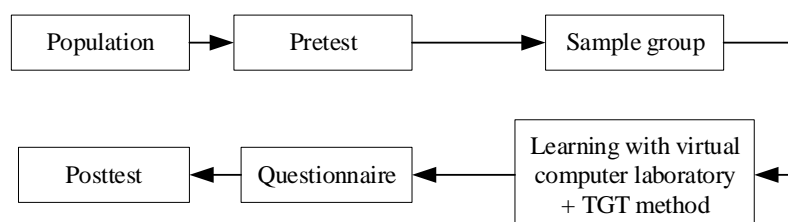


Figure 4. The implementation design and plan

The preliminary study was conducted to investigate the possibilities on the use of the developed virtual computer laboratory integrated with TGT approach in enhancing student understanding on the topic of computer assembly in a high school. Figure 4 illustrates the implementation design started with pre-test and chosen sample group from students, who had score at the last 12th in the score rank. After that, 12 students were categorized into 3 ability groups: high, medium and low achieving. Then the teacher assigned the students into 4 teams called home group; each home group consisted of 3 members with different abilities. In the team step shown in Figure 5 (a), students from home group were shared the duties among their team to study the computer generations and functions of computer devices through the learning center in the VR-based virtual laboratory; each group was given the limited time. In the tournament step shown in Figure 5 (b-c), the student was assigned to assemble the computer in the virtual laboratory in which there were three levels. The low, medium, and high achieving students were assigned the level of 1, 2, and 3, respectively. The score from each student was summed to be a group score. Finally, teacher and students summarize each team's score and students did the post-test. In addition, teacher asked the students to answer the questionnaire in order to investigate students' satisfaction toward virtual computer laboratory.

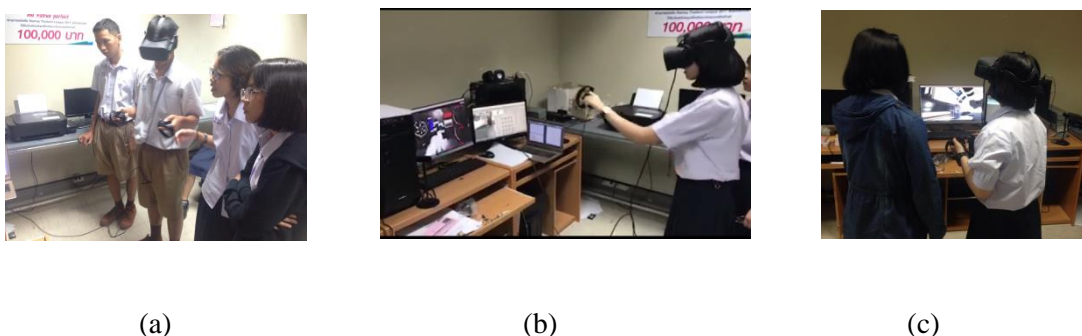


Figure 5. The preliminary study (a) learning with the virtual lab, (b) medium-achieving students assemble computers at level 2, (c) high-achieving students assemble computers at level 3

3. Results

3.1 Learning Achievement between Pretest - Posttest and Normalized Gain

Table 1

Learning achievement score

Students	Pre-test (20)	%Pre-test	Post-test (20)	%Post-test	Normalized gain
1	6	30	16	80	0.71
2	8	40	15	75	0.58
3	8	40	15	75	0.58
4	8	40	17	85	0.75
Low-achieving group		38		79	
5	9	45	16	80	0.64
6	9	45	16	80	0.64
7	9	45	16	80	0.64
8	10	50	16	80	0.60
Medium-achieving group		46		80	
9	11	55	17	85	0.67
10	11	55	17	85	0.67
11	11	55	18	90	0.78
12	11	55	18	90	0.78
High-achieving group		55		88	
Average	9	46	16	82	0.67

Table 1 shows the percentage of learning achievement test in which post-test was statistically higher than that of the pre-test. The average of student pre-test score is 46% while that of post-test score is 82%. Moreover, the class normalized gain is 0.67 which is at medium level.

3.2 The comparison of learning achievement between pretest - posttest and normalized gain by difference levels of achievement

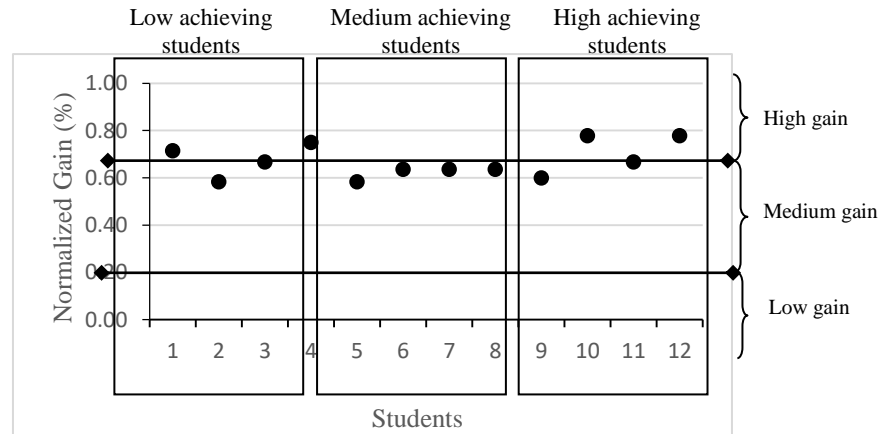


Figure 6. The normalized gain by difference levels of achievement

Figure 6 illustrates the percentage of normalized gain of students in 3 difference learning abilities. From the results, we found that two students in low group were able to develop their skills to be in the high gain. This could be concluded that our developed approach could help students gain more understanding of concept of computer assembly.

3.3 Students' Satisfaction toward Virtual Computer Laboratory

Table 2

The Results of Students' Satisfaction toward Virtual Computer Laboratory

Questions	Mean	Std. Dev. (SD)
The image is beautiful, clear, realistic.	4.25	0.52
The sound is clear.	4.17	0.70
Satisfaction with animation	4.33	0.53
Satisfaction with 3D models	4.42	0.53
Can be used easily	4.17	0.63
Get fun and enjoy	4.25	0.48
Get more knowledge about computer equipment	4.33	0.48
Get more knowledge of computer assembly	4.58	0.70
Able to recognize more computer devices	4.25	0.63
Understand how to build a computer	4.25	0.42
Total of satisfaction with virtual laboratories	4.17	0.67
Overall	4.29	0.60

Table 2 shows that the average of students' satisfaction was at the high level (Mean = 4.29, SD = 0.60). From the results, we found that the highest satisfaction was *get more knowledge of computer assembly* followed by the *satisfaction with 3D models*, and *satisfaction with animation*.

3.4 Students' Computer Assembly Skills

The score, collected by the assessment forms, of computer assembly skills was compared between groups. The results, illustrated in Figure 7, show that all of the team score is higher than 70 percent which is the pass criteria. Figure 7 shows the comparison between team score and individual score.

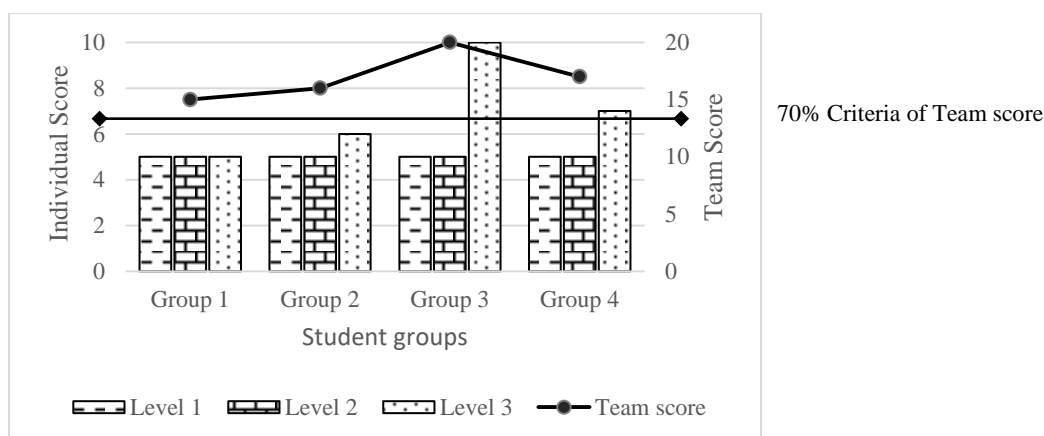


Figure 7. The comparison score between students' groups with criteria of 70 percent

4. Conclusion and Discussion

This paper is presented the using of virtual computer laboratory in conjunction with team game tournament learning method for 12th grade students. By using the advantages of virtual reality technology that can simulate computer laboratory and various computer equipment. The virtual reality technology allows the user to manage objects in virtual environment for training computer assembly skill. To promote learners to collaborate and to increase the challenge of learning, we integrate the team game tournament (TGT) learning approach to encourage students to have cooperative skill and helping each other to solve problems. The most important is that it leads to the development of learning achievement of learners better. From the result of this preliminary study, we found that all 3 achieving groups had a higher post-test score than pre-test, and the average of normalized gain was at the medium level. With virtual reality technology integrated with TGT approach, it enables learners who have different levels of learning skills to help each other and increasing the learning progress of low group to have a post-test score, equivalent to students in the medium and high groups.

Acknowledgements

The authors acknowledge the Project for the Promotion of Science and Mathematics Talented Teachers (PSMT), the Institute for the Promotion of Teaching Science and Technology (IPST), Ministry of Education, Thailand. They thank Mr.Ruengyos Soikeaw from Ubon Ratchathani University for assisting development of the virtual computer laboratory. The authors are also grateful to Faculty of Science, Ubon Ratchathani University for use of facilities.

References

- Battelle for Kids. (2019). *Framework for 21st century learning definitions*. Retrieve from <http://www.battelleforkids.org/networks/p21/frameworks-resources>
- Beal, V., (2010). *The five generations of computers*. Retrieved from https://www.webopedia.com/DidYouKnow/Hardware_Software/FiveGenerations.asp
- Buttach, N., (2018). *Course description: Computer maintenance*. Ubon Ratchathani: Lukhamhan warinchamrab school.
- Cisco. (2006). *IT Essentials: PC Hardware and Software v4.0*. Retrieved from http://www.uobabylon.edu.iq/eprints/publication_3_27197_1357.pdf
- DeVries, D. L. (1980). *Teams games tournament: The team learning approach*. New Jersey: Educational Technology publication.
- Malithong, K. 2000. *Educational Technology and Innovation*. Bangkok: Aroonprinting.
- Ministry of Education. (2017). *Learning Area of Science*. Bangkok: The Agricultural Cooperative Federation of Thailand.

Instructional Design of STEAM Education Based on Virtual Reality Technology —— Taking LEGO Bricks as An Example

Fangyuan Tong^{a*}, Jiahui Ruan^b, Lingling Meng^c

^{a,b,c}Department of Education Information Technology, East China Normal University, China
*296627577@qq.com

Abstract: STEAM education can integrate interdisciplinary knowledge through projects, develop learners' hands-on practical skills and promote innovation ability through project learning, which is conducive to the development of learners' comprehensive development. Meanwhile, with the rapid development of computer network technology, virtual reality (VR) plays an increasingly important role in education, and its combination with education has become much closer. This study combined the STEAM education with VR and developed an instructional design with LEGO Bricks as an example. The instructional design developed was proved not only to pass on learners' interdisciplinary knowledge imperceptibly, but also cultivate their hands-on practical skills and innovative ability during the whole project. Finally, VR technology is applied to realize the instructional design in a visualized way.

Keywords: STEAM education, virtual reality, instructional design

1. Introduction

Nowadays, countries are faced with the contradiction of lack of comprehensive talents while the task of cultivating innovative competence is becoming more and more crucial. However, problems remained in STEAM education can be attributed to three-fold. First, the results of some theoretical studies are too macroscopic to be landed in practical teaching and application. Second, there is a considerable gap in establishing a systematic and rigorous curriculum resource system. Third, lack of funds and education equipment has led to the inability of popularizing STEAM education to some extent.

The purpose of this study is to first, design a curriculum resource by accessing to literature and field research, and combine science, technology, engineering, arts, and mathematical into a project to help solving problems derived from reality. Second, apply virtual reality technology to create an immersive and simulated learning environment aiming at fostering learners' ability. LEGO Bricks were used as an example since this study is aimed at preschool children.

2. STEAM education

By integrating knowledge and skills in science, technology, engineering, art and mathematics, STEAM Education integrates disciplines that are traditionally unconnected and isolated, combining knowledge points in the discipline (Hetland & Winner, 2004; Liao, 2016; NAEA, 2016). In the process of project learning, it can effectively avoid the knowledge split caused by traditional teaching, which is not only conducive to cultivating students, but also help cultivate the comprehensive quality, hands-on practical skills, teamwork ability and innovation ability (Gates, 2017).

3. Virtual reality (VR) technology

With the development of technology, virtual reality technology has made great progress. At the same time, virtual reality has lots of advantages (Ke et al., 2015). Applying virtual reality technology to STEAM education can combine the advantages of both: providing a virtual immersive teaching environment to stimulate students' interest; adopting interdisciplinary project teaching methods to

improve students' comprehensive quality and innovation ability. Besides, the combine of these two can cost the least but create the more.

4. Instructional Design

4.1 Basic Ideas of Instructional Design

Project-based learning is a student-centered approach to teaching, and it emphasizes the tasks and problems of the real world, can be related to the reality of life. At the same time, project-based learning is usually carried out in small groups, they can communicate and collaborate with their partners, which is conducive to cultivating students' teamwork and communication skills. The project-based learning can be used in the STEAM education, and it also cultivates the students' comprehensive ability.

The "learning by doing" was proposed by the famous American educator Dewey. He believes that "all learning is a by-product of action, so teachers should 'do' to encourage students to think and learn". The instructional design of this study starts with the problems in real situation, which lets the students solve problems through hands-on practice to achieve real "learning by doing".

4.2 Instructional design

In the case of consulting relevant literature materials and field visits of existing STEAM training institutions, the content of the teaching is determined to "a big manor". It is necessary to use LEGO bricks to make houses, cars and parking lot gates. The project design is shown in Table 1.

Table 1

Project design

<p>Project name: A big manor</p> <p>Project goals:</p> <ul style="list-style-type: none"> To master basic scientific knowledge points and methods for building LEGO bricks. To use building blocks to build space thinking ability and hands-on practical skills. To enhance teamwork and communication skills through cooperation.
<p>Design Challenge:</p> <p>Students are expected to design a manor. The project' success criteria are aesthetics and practicality.</p>
<p>Mini project Challenge 1: <i>Lovely house</i></p> <ul style="list-style-type: none"> To know the common building blocks in LEGO bricks. To master the method of "stitching" the building blocks. <p>Mini project Challenge 2: <i>Fast car</i></p> <ul style="list-style-type: none"> To know the special blocks such as nine-hole beams and pulleys in LEGO bricks. To initially understand the existence of friction and simply determine the amount of friction. <p>Mini project Challenge 3: <i>Flexible parking lot gate</i></p> <ul style="list-style-type: none"> To know the special blocks such as beams, gears, handles and turbine boxes in LEGO bricks. To know the simple turbine and understand its self-locking function.

4.3 Virtual Reality Technology Implementation

This study uses Photoshop and 3ds Max to create classroom models and LEGO bricks' models, teacher role models and other models required by the classroom. And then import 3D models into Unity3d for integration and optimization, to achieve the basic functions of STEAM teaching. The software structure diagram for instructional design is shown in Figure 1.

The STEAM instructional design is mainly divided into three parts: "Recognition", "Learning" and "Doing".

"Recognizing": Introduce the LEGO bricks. Introduction and voice of the corresponding will appear if click on the Lego bricks. The mouse can control the zooming and rotation of the scene.

“Learning”: Through the “Project Introduction Video”, the tasks that need to be completed in this project are proposed. Other videos explain the skills of building LEGO bricks.

“Doing”: Apply the above-mentioned learning to the actual situation. The left mouse button controls the appearance of the LEGO bricks, the middle mouse button controls the zoom of the screen, the right mouse button controls the rotation of the screen, and the keyboard control the movement of the screen. As all shown in Figure 2.

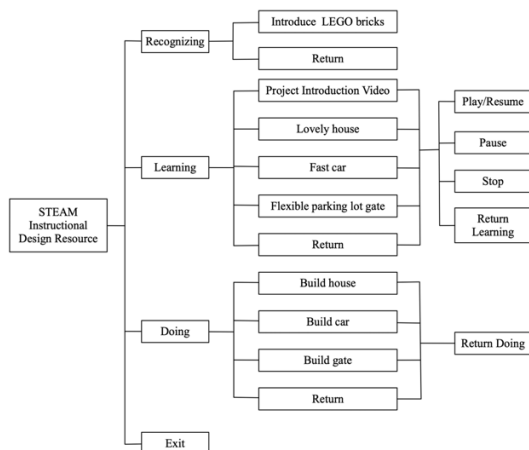


Figure 1. instructional design



Figure 2. schematic diagram

5. Results

After the instructional design content was completed by virtual reality, it was practiced for three months in the audition course of a STEAM institution, and the results are quite good.

Students can integrate the life content in the learning process and make it easy to understand. And the construction of LEGO bricks requires not only three-dimensional design, but also the observation and analysis of the geometric features of the LEGO bricks, which has great benefit to the development of students’ spatial thinking ability. Besides, through project-based learning, students can develop creativity and hands-on practice in the process of independent inquiry.

6. Conclusion

The development of STEAM education can be combined with the emerging technologies of the moment, and the combination with virtual reality is a good way. Virtual reality has many advantages such as stimulating learners’ motivation, creating a realistic teaching environment, feeling immersive learning experience, and integrating cross-border knowledge. Because of the commonality of STEAM education and virtual reality, their combination will create a huge advantage. However, how to better demonstrate the advantages of virtual technology in teaching and how to better implement and popularize STEAM education is worth further exploration.

References

Gates, A. E. (2017). Benefits of a STEAM collaboration in Newark, New Jersey: Volcano simulation through a glass-making experience. *Journal of Geoscience Education*, 65(1), 4-11.

Hetland, L., & Winner, E. (2004). Cognitive transfer from arts education to non-arts outcomes: Research evidence and policy implications. In E. W. Eisner, & M. D. Day (Eds.). *Handbook of research and policy in art education* (pp. 135-162). Routledge.

Ke, F., Im, T., Xue, X., Xu, X., Kim, N., & Lee, S. (2015). Experience of adult facilitators in a virtual-reality-based social interaction program for children with autism. *The Journal of Special Education*, 48(4), 290-300.

Liao, C. (2016). From interdisciplinary to transdisciplinary: An arts-integrated approach to STEAM education. *Art Education*, 69(6), 44-49.

National Art Education Association [NAEA] (2016). Using art education to build a stronger workforce.

CodAR: An Augmented Reality Based Game to Teach Programming

Vandit SHARMA^a, Jeevankur TALUKDAR^a & Kaushal Kumar BHAGAT^{a*}

^a*Indian Institute of Technology Kharagpur, India*
**kkntnu@hotmail.com*

Abstract: In recent years, technology has revolutionized all spheres of life. Since programming is the heart of software technology, it is thus imperative that the demand for programmers is also increasing day by day. With advancements in the field of augmented reality (AR) and Computer Vision (CV), we can now develop applications for unique experiences in the field of education. This study aims to develop a game for elementary school students to learn programming skills. Students are provided with cards which act as markers for our Game. Each marker acts a distinct programming block in AR which causes our game character to perform a certain action. The student needs to place these blocks in the right manner to accomplish a given task. It thus enables students to learn some basic programming skills in a way appealing to them.

Keywords: Augmented reality (AR); elementary education; programming; gamification; interactive learning environment; skill development

1. Introduction

Programming is involved behind every new technology developed around us. Therefore, learning how to program is becoming more and more common among individuals day by day. Research has shown that programming promotes logical thinking, reasoning skills, and general problem-solving skills in students of all age groups (Liao & Bright, 2011). Recent developments in the field of AR have created new opportunities for unique experiences in healthcare, education, and entertainment. This study uses AR to develop coding skills of elementary school students. The research involves the integration of computer vision techniques with AR to develop an AR-based game: CodAR. The developed game enables students to learn the importance of algorithms, sequencing and loops in an interactive and visually appealing way. The game has been developed in Unity3D software with support of Vuforia software development kit. We believe that the introduction of this game will captivate students' interest in the field of programming and give them an intuitive understanding of the same.

In the following sections, we discussed the related work (Section 2), elaborated the design and development of the game (Section 3) and followed by conclusion and future directions (Section 4).

2. Related work

Gamification in education refers to the inclusion of game design and principles in education to provide unique and engaging experiences. Research has shown that it can influence users in mainly three aspects - cognition, emotion and social (Detering, Khaled, Nacke & Dixon, 2011). Gamification broadens the player's perception about learning, besides teaching users the process of thinking (Detering, Khaled, Nacke & Dixon, 2011). Many research works have been done in this field, but we did not find such an AR-based game to learn programming skills. Guenaga et al. (2014) developed a game to learn programming skills but it was not totally AR-based.

Lee and Hammer (2011) proposed the use of game-like rules system, player experiences and cultural roles to shape learners' behavior. Researchers have also used meta-analytic techniques to show that Serious games influence learning in two ways, by changing cognitive processes and by affecting

motivation (Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek, 2013). Games invoke a variety of emotional experiences - from frustration to joy. Students learn to see failure as an opportunity and in turn try to find out the cause of the error instead of feeling helpless about it. Our research, therefore, not only helps in cognitive development but also motivates the next generation of innovative programmers.

3. Proposed system

The proposed system is developed using a smartphone/computer and a set of markers which are provided to the user as cards. Figure 1 shows the markers used in the proposed system. There are five types of markers used in this game. The detailed purpose of each marker is described below:

- Task - It generates a virtual game area on the screen comprises of our game character, generated level and the task of the level.
- Direction - It moves the game character one block forward in the direction it is facing
- Rotation - It rotates the game character by 90° in the clockwise/anti-clockwise direction
- For Loop - It repeats the marker placed immediately before a given number of times, which is generated automatically by the level
- While Loop - It repeats the set of markers placed before it infinitely or till level completion.

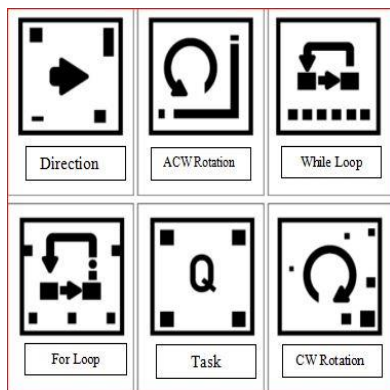


Figure 1. Markers

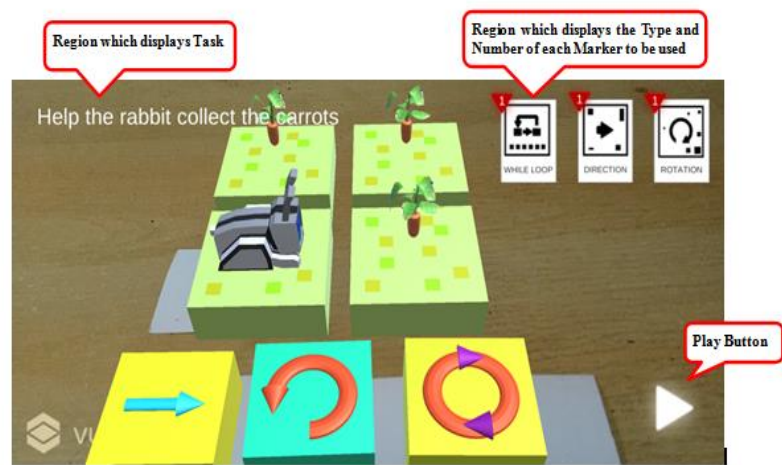


Figure 2. Screen of the proposed system

Figure 2 shows the screen of the proposed system. On the screen, a virtual block corresponding to the recognized marker is displayed. The user moves the marker which moves the corresponding virtual block on the screen. The virtual blocks are brought in the scene one by one and placed next to each other to generate a sequence of commands. The set of these virtual blocks corresponding to each marker is shown in Figure 3.

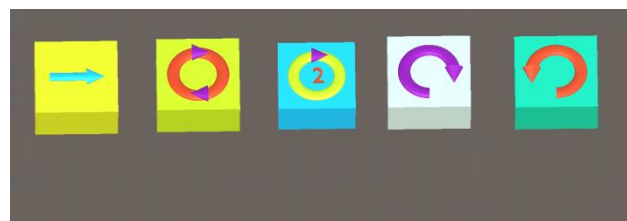


Figure 3. Virtual Blocks corresponding to the markers

After the user is satisfied with his/her sequence of markers, the play button should be pressed. Then, our game character performs actions corresponding to the set of instructions received, if feasible. The type of markers and the number of each marker required on each level are displayed in the upper right corner

of the screen. The task is displayed on the upper left corner of the screen. The user has to complete the task using the specified type and number of markers. As the user clears easier levels, the game automatically starts incorporating the use of Loops and tougher Sequences. In Figure 4(a) the user has to complete the level just by using simpler blocks (Direction and Rotation). There can be various solutions to the same level also. Such a solution to the same level is shown in Figure 4(b) where the user is trying to clear the level by using For Loops. This analogy between conventional programming and virtual programming helps the user to understand the intuition of using Loops better.

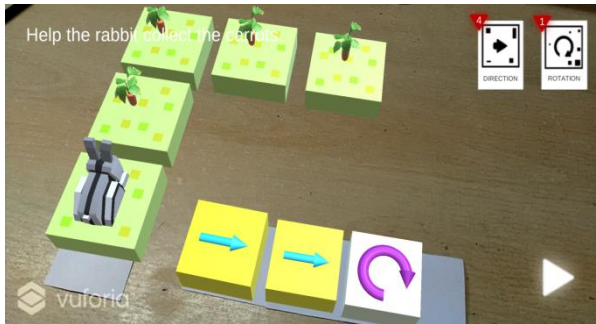


Figure 4(a). Solution 1 to Game level

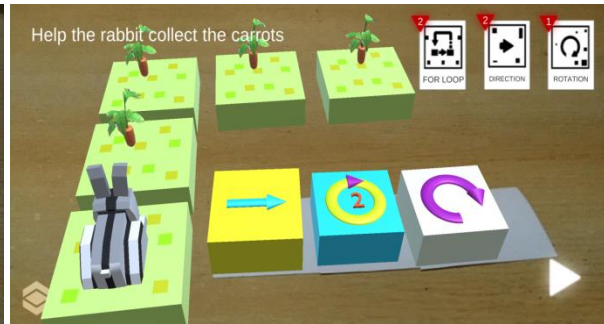


Figure 4(b). Solution 2 to same Game level

4. Conclusion and future directions

Conventional techniques to teach programming generally require a good understanding of various mathematical concepts and are difficult to visualize. Therefore, various visual programming techniques are being developed to teach concepts of programming to younger students. In this paper, we presented CodAR which takes this idea to the next level in terms of user understanding and engagement concepts by practical implementation and visualization. Our approach is unique mainly because of three reasons. First, there is a simple relationship between abstract mathematical concepts of programming viz. Sequence, For Loop and While Loop and the Virtual Blocks. Second, the game is developed using AR, which makes it quite engaging and interactive. Third, with recent smartphone penetration, such devices are now easily accessible. This makes our approach quite portable and easily scalable to a wider audience.

Our work certainly has a limitation. Currently, we have not conducted an empirical evaluation of our developed game. We are currently working on it. Some future improvements which can be made to the game include the addition of more types of markers corresponding to Conditionals, Boolean, Switches, Data Structures, Debugging etc. along with more challenging levels.

Reference

- Detering, S., Khaled, R., Nacke, L. E., & Dixon, D. (2011). Gamification: Toward a Definition. In *Proceeding of the CHI 2011 Gamification Workshop, Vancouver, Canada*
- Guenaga, M., Menchaca, I., de Guinea, A. O., Dziabenko, O., García-Zubía, J., & Salazar, M. (2013). Serious Games, Remote Laboratories and Augmented Reality to Develop and Assess Programming Skills. In *International Simulation and Gaming Association Conference* (pp. 29-36). Springer, Cham.
- Lee, J. J., & Hammer, J. (2011). Gamification in education: What, how, why bother? *Academic exchange quarterly*, 15(2), 146
- Liao, Y.-K. C., & Bright, G. W. (1991). Effects of computer programming on cognitive outcomes: A meta-analysis. *Journal of Educational Computing Research*, 7(3), 251-268
- Sharma, V., Musarrat, R., Chimalakonda, S., Reddy, Y. R. (2017). Muse : A Musically Inspired Game to Teach Arrays and Linked Lists. In Yang, J. C. et al. (Eds.) *Proceedings of the 25th International Conference on Computers in Education (ICCE 2017)* (pp.802-807). Christchurch, New Zealand: Asia-Pacific Society for Computers in Education.
- Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of educational psychology*, 105(2), 249.

English Language Learners' Perceptions and Experiences of an English Mobile Urban Game

Yu-Feng (Diana) YANG ^{a*}

^a*Department of Foreign Languages and Literature, National Sun-Yat Sen University, Taiwan*

**dyang@faculty.nsysu.edu.tw*

Abstract: This study aims to explore English language learners' (ELLs) perceptions and experiences of a mobile urban game. How ELLs perceive its value for English learning and exploit the affordances of the mobile game are emphasized. Through the data analysis of 15 ELLs' recorded gaming processes, written reflections, and interviews, this study reported that these ELLs perceive the mobile urban game to be useful in creating meaningful and authentic English language learning opportunities. In addition, they exploit multiple mobile applications (e.g., search engines, online chatting tools, online dictionaries, google maps, built-in cameras) and interact with social members in their surroundings (e.g., peers, the instructor, vendors, foreign visitors) for the game quests. Future research can examine larger study population's experiences of similar mobile urban games and explore how social members can facilitate ELLs' language learning processes in the game.

Keywords: mobile urban game, mobile location based game, mobile game, gamification, English language learning

1. Introduction

In recent years, mobile games have received increasing attention in the field of second and foreign language (hereafter L2) education. Featured with portability, social interactivity, context sensitivity, connectivity, and individuality (Squire & Klopfer, 2007), mobile games can offer ubiquitous, interactive, contextualized, authentic, and personalized learning environments for L2 learners (Holden & Sykes, 2011). Harnessing these features, a few scholars have started to investigate how to integrate mobile location-based game in L2 classrooms (Holden & Sykes, 2013; Perry, 2015). Different from immersive digital games that involve players solely in virtual spaces, mobile location-based games integrate the physical and the virtual in a gameplay. Players use a physical public space, such as a museum or a town, as the game board, and move around the assigned geographical locations in reality, to gather real-world materials, items and clues, while solving problems and completing missions in the virtual gameplay (Pitura & Terlecka-Pacut, 2018). Encouragingly, research shows that some L2 learners exploit artifacts, materials and social members distributed in their surroundings to accomplish game quests during the gameplay (Thorne & Hellermann, 2017). Studies in this line can continue to contribute to our understanding of L2 learning in mobile gaming contexts.

The purpose of this study is to explore English language learners' (hereafter ELLs) perceptions and experiences of an English mobile urban game, a type of mobile location-based games that integrates virtual-physical spaces and exploits the portability, the social interactivity, and the context sensitivity features of portable devices. How ELLs perceive its value for English learning and exploit game-mediated affordances for game quests/challenges are emphasized.

2. Research Methodology

2.1 *The Mobile Urban Game: Scavenger Hunt in YanCheng*

The English mobile urban game of this study, *Scavenger Hunt in YanCheng*, aims to familiarize students with cultural concepts and artifacts, and their associated English expressions that can be used

to introduce local communities and histories in town. On the basis of experiential learning and situated problem solving, the researcher designed 7 challenges/quests in relation to the local areas/shops for the players to experience English meaning making *in situ* (Thorne & Hellermann, 2017).

Game challenges, structured in the forms of riddles, crosswords, multiple-choice questions, and short-answer questions, were developed to elicit players' motivation of searching for clues and information in physical and virtual spaces. Game quests, delivered through error correction, speaking and writing tasks, were devised to encourage players to reflect, speak and write in English, and to make use of the English words, expressions and concepts they have just experienced during the game challenges.

Players used Google Form to go through the game, reading challenge/quest information and submitting answers and tasks for the game by using their mobile phones. QR codes were embedded in the Google Form for players to access context-aware information about the game

2.2 Participants and Data Collections

This small-scale research was conducted in an undergraduate course designed for ELLs in a university in southern Taiwan. 15 students, from Taiwan, Indonesia, Korea, France and Mainland China, were divided into 6 different groups to play the game in this study. Their English proficiency levels ranged from intermediate to advanced level based on the English standardized tests taken prior to the study.

This study was qualitative in nature. Data collections consisted of audio/video recordings of the gaming processes, students' reflections of the game, as well as stimulated recalls and interviews.

2.3 Data Analysis

Because this study is interested in learning about ELLs' perceptions and experiences of the English mobile urban game, their emic perspectives and gaming behaviors are taken as the primary focus in the data analysis. The researcher applied Miles, Huberman, and Saldana's (2014) guidelines for initial coding and pattern coding, allowing emerging codes and reoccurring patterns to be derived from the data (e.g., participant evaluations: authenticity; contextualization, meaningfulness; gaming behaviors: online inquiry; expert inquiry; peer discussion; vendor support; foreigner conversation, etc.). These codes and patterns were further examined for possible relationships and analyzed for explanations for their roles and importance. Clustering, matrices, and analytical memoing were used as analytical strategies to examine and document these relationships and explanations.

3. Findings and Discussion

3.1 Student Perceptions of Game Value for English Learning

The findings of this study suggest that the study participants hold positive opinions about the mobile English urban game. While many of them reported that they enjoy participating in the scavenger hunt situated in the game, they also appreciate the opportunities to explore the culture of the local YanCheng district.

Notably, the mobile English urban game provides the study participants great opportunities to experience English meaningfully and authentically. While the game contextualizes the participants in a meaningful English storyline, it also requires their close attention to the background information and the clues offered in the physical spaces for quest solving. Such designs encourage students to "be attentive to the historical introduction provided by the store on the site...and repeatedly engage in trial and error for problem-solving (Student C)." Meanwhile, being at the physical place where particular English expressions are used, students experience and embody the authenticity of the English language, revealing the importance of meaning making *in situ* (Thorne & Hellermann, 2017). As student D revealed in the written reflection,

...during the time that we have spent in the coffee shop, I have learned lots of stuffs that related in coffee [in English], including the syphon coffee maker, coffee pot, and each specific step I need to do before I can get a cup of coffee.

3.2 Student Exploitation of Game-Mediated Affordances

In terms of affordances enhanced by the mobile urban game, participants in this study take advantages of mobile applications, such as search engines, online chatting tools, google maps, online dictionaries, and built-in cameras to assist them with their English game quests. Clearly, study participants exploit the portability, social interactivity, context sensitivity, connectivity, and individuality features of their mobile devices. However, peer scaffolding and instructor support remain salient during the process of quest solving. With peers in the physical space, and the instructor in the virtual space, students exploit collective (Donato, 1994) and expert scaffolding (Wood, Bruner, & Ross, 1976) for the problems they are incapable of solving independently. Interestingly, foreigner visitors and local vendors also play major roles in ELLs' quest solving. Some students reported that it was the cultural talks with the owner in the coffee shop enhanced their capability of solving the English quest in the mobile urban game.

4. Conclusions and Future Research

Although the number of the participants in this study is limited, the findings of this study can provide initial understanding about ELLs' perceptions and experiences of a mobile location-based game. Future research can apply similar designs to study larger research population and to examine ELLs' perceptions and difficulties during the gameplays. It is also worth exploring how social members (e.g., peers, experts, vendors, foreigners, etc.) can contribute to ELLs' language learning during the mobile location-based gameplay.

Acknowledgements

This research was supported by the Ministry of Science and Technology of Taiwan (MOST 108-2410-H-110 -023). I would like to thank all the participants who joined this research project.

References

- Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33-56). Norwood, NJ: Ablex.
- Holden, C. L., & Sykes, J. M. (2011). Leveraging mobile games for place-based language learning. *International Journal of Game-Based Learning (IJGBL)*, 1(2), 1-18. doi:10.4018/ijgbl.2011040101
- Holden, C. L., & Sykes, J. M. (2013). Complex L2 pragmatic feedback via place-based mobile games. In N. Taguchi & J. M. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching* (pp. 155-183). Philadelphia: John Benjamins.
- Miles, M., Huberman, M., & Saldana, J. (2014). *Qualitative Data Analysis: A Method Sourcebook (3rd Edition)*. Thousand Oaks, CA: Sage.
- Perry, B. (2015). Gamifying French language learning: A case study examining a quest-based, augmented reality mobile learning-tool. *Procedia - Social and Behavioral Sciences*, 174, 2308-2315. <https://doi.org/10.1016/j.sbspro.2015.01.892>
- Pitura, J., & Terlecka-Pacut, E. (2018). Action research on the application of technology assisted urban gaming in language education in a Polish upper-secondary school. *Computer Assisted Language Learning*, 31(7), 734-763. doi:10.1080/09588221.2018.1447490
- Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences*, 16(3), 371-413. doi:10.1080/10508400701413435
- Thorne, S. L., & Hellermann, J. (2017). Mobile augmented reality: Hyper contextualization and situated language usage events. In J. Colpaert, A. Aerts, R. Kern, & M. Kaiser (Eds.), *Proceedings of the XVIII International CALL Conference: CALL in Context* (pp. 721-730). Berkeley, CA: University of California.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. doi:10.1111/j.1469-7610.1976.tb00381.x

MAIN CONFERENCE C6

TECHNOLOGY ENHANCED LANGUAGE LEARNING (TELL)

- FULL PAPER -

- A REVIEW OF 3D VIRTUAL ENVIRONMENTS FOR LANGUAGE LEARNING: NEW TEACHING PRACTICE AND RESEARCH TREND 608**
CHUN-PING ZHENG, LING-YU XU, MENG-YA GAO, QIAN-QIAN CHENG, ZI-TONG YANG & LI-LI WANG
- ENHANCING LOW ACHIEVERS' EFL LEARNING WITH INTERACTIVE DIGITAL TECHNOLOGIES 617**
*SHU-YUAN LIN, MING-PUU CHEN, LI-CHUN WANG, YU-TING KAO, DI ZOU, HAORAN XIE
- DESIGN AND DEVELOPMENT OF A CONJUNCTIVE WORD-LEARNING SUPPORT SYSTEM FOR CONJUNCTIVE EXPRESSIONS WITH DIFFERENT-MEANING COMMUTATIVITY AND ITS EXPERIMENTAL USE 624**
KENGO IWAI, FONTELES FURTADO PEDRO GABRIEL, YU OGATA, YUSUKE HAYASHI, TSUKASA HIRASHIMA
- A TURKISH EFL TEACHER'S CHANGE PROCESSES THROUGH AN OPD PROGRAM: A CASE STUDY 634**
BEHICE CEYDA SONGÜL, ÖMER DELIALİOĞLU, ÇAĞRI ÖZKÖSE BIYIK

- SHORT PAPER -

- VALIDATING AN INSTRUMENT FOR EFL LEARNERS' ENGLISH NEW MEDIA LITERACY AND THE RELATION TO ENGLISH LANGUAGE SELF-EFFICACY 644**
LIN LUAN, YAN DONG, SHASHA WANG, YIN LI
- INVESTIGATING CHINESE UNIVERSITY EFL LEARNERS' SELF-EFFICACY IN A BLENDED LEARNING ENVIRONMENT 650**
JINGJING LIAN
- MULTITASK LEARNING FOR CHINESE GRAMMATICAL ERROR DETECTION 656**
YU-JIE ZHOU, YONG ZHOU
- IMPROVING SUMMARY WRITING PERFORMANCE VIA THEORY-BASED LEARNING SYSTEM 662**
CHIOU SHENG CHEW, WEN-CHI VIVIAN WU & NORISMA IDRIS & ER FU LOH
- ENHANCING EFL COLLEGE STUDENTS' LANGUAGE PERFORMANCE VIA EBOOK SUPPORTED LEARNING 667**
FANG-YING RIVA LO, WEN-CHI VIVIAN WU, CHENG-HUAN CHEN & JUN SCOTT CHEN HSIEH
- AUGMENTED REALITY AND 3D MODEL FOR CHILDREN CHINESE CHARACTER RECOGNITION - HONG KONG PRIMARY SCHOOL EDUCATION 673**
KA-YAN FUNG, KWONG-CHIU FUNG & DR. SALLY, WAI-YAN WAN
- DIAGNOSTIC LANGUAGE ASSESSMENT: LESSONS LEARNED FROM RAPID PROTOTYPING 679**
ALLAN NICHOLAS & JOHN BLAKE
- EFFECTS OF LEARNING ACTIVITIES BASED ON AUGMENTED REALITY ON STUDENTS' UNDERSTANDING AND EXPRESSION IN AN ENGLISH CLASS 685**
YANG YANG, SHUMIN WU, DANTONG WANG, YAOHUA HUANG & SU CAI
- DEVELOPING AN INTEGRATED SYSTEM OF ROBOTS AND TOYS WITH INTERNET OF THINGS FOR CHILDREN'S LANGUAGE DEVELOPMENT 691**
YA-WEN CHENG, YUPING WANG, MING-RONG LIANG, NIAN-SHING CHEN & YA-FEI YANG
- AUTOMATIC VOCABULARY STUDY MAP GENERATION BY SEMANTIC CONTEXT AND LEARNING MATERIAL ANALYSIS 698**
BRENDAN FLANAGAN, MEI-RONG ALICE CHEN, LOUIS LECAILLIEZ, RWITAJIT MAJUMDAR, GÖKHAN AKÇAPINAR, PATRICK OCHEJA & HIROAKI OGATA

- POSTER PAPER -

BUILDING A CONFUSED CHARACTER SET FOR CHINESE SPELL CHECKING 703

LUNG-HAO LEE, WUN-SYUAN WU, JIAN-HONG LI, YU-CHI LIN & YUEN-HSIEN TSENG

READING ASSISTANCE FOR EFL READERS WITH KIT-BUILD CONCEPT MAP WITH SOURCE-CONNECTION 706

BANNI SATRIA ANDOKO, YUSUKE HAYASHI, TSUKASA HIRASHIMA & ATIQAH NURUL ASRI

A Review of 3D Virtual Environments for Language Learning: New Teaching Practice and Research Trend

Chun-Ping ZHENG*, Ling-Yu XU, Meng-Ya GAO, Qian-Qian CHENG, Zi-Tong YANG & Li-li WANG

School of Humanities, Beijing University of Posts and Telecommunications, China
*zhengchunping@bupt.edu.cn

Abstract: Three dimensional virtual environments (3DVEs) have been noted as effective learning space for promoting learners' language acquisition and academic performance. This paper reviewed a total of 33 empirical studies published in the *Language Learning & Technology*, *ReCALL*, *Computer-assisted Language Learning*, *System*, *CALICO Journal* and *Computers & Education* from 2010 to 2017. At first, the current study explored the yearly publication trend of the journals and the research productivity of different countries. Then, based on systematic content analyses, the present paper shows that a majority of the studies employed the mixed research method and only a few articles adopted pure quantitative or qualitative method. The potential affordances of the 3DVEs include enhancing language learners' overall language proficiency, language performance or target language output as well as non-linguistic knowledge and skills. Drawing upon the synthesized research findings, we claim that the 3DVE has become a burgeoning area of research with great potential for innovating language education. Pedagogical implications were summarized and suggestions for future pedagogical innovations were also discussed at the end.

Keywords: Content analysis; three dimensional virtual environments (3DVEs); language education; virtual reality; augmented reality

1. Introduction

With the unprecedented momentum of technological development, scholars are calling for a better understanding of applying emerging technologies for educational purposes. Three dimensional virtual environments (3DVEs) or virtual world (VW) have been noted as positive learning space which promotes learners' knowledge construction and meaningful learning. The newest generation of virtual environments was featured by multiple interaction channels, such as voice or video-based synchronous online communication, simulated avatars, and virtual reality systems (Shih, 2014). Compared with conventional text-based or voice-based CMC technology, these innovative technologies provided more possibilities for transformative language learning and teaching, facilitating the transmission of symbols to convey users' meaning (e.g., Shih, 2014; Stevens, 2006). These innovations also represent major improvements upon traditional classroom settings which usually cannot create authentic or required conditions for specific language learning contexts (Chen et al., 2011; Clarke, Dede, & Dieterle, 2008).

2. Research Methodology

The 3DVEs refer to "immersive, three-dimensional (3D), multimedia, and multi-person simulation environments, where each participant adopts an alter ego and interacts with the world in real time" (Wagner and Ip, 2009, p. 250). In the field of applied linguistics, the application of 3DVEs is still an emerging research topic. Peterson (2012) made a preliminary comparison between the 2D virtual worlds and the 3DVEs and revealed the superior characteristics of the latter learning environment: (1) 3D interfaces with high representational fidelity may improve learner engagement; (2) multiple communication channels (synchronous and asynchronous textual, aural or visual interactions) could

promote target language production; (3) virtual representation of language learners (usually in the form of avatars) improve learners' social presence, alleviate their foreign language anxiety and further motivate their TL learning. 3DVEs were perceived as "stimulating environment for learners to undertake a range of beneficial forms of social interaction involving collaborative dialogue in the Target Language (TL)" (Peterson, p. 37). The use of 3DVEs in CALL has becoming an increasingly interesting topic and researchers are calling for more further research on how this dynamic and interactive context could enhance learners' target language development (Peterson, 2012). Since there are still few systematic review studies on the use of virtual environments for foreign language learning and teacher, this study aims to reveal the general publication trend of the studies published in key CALL journals from 2010 to 2017. Moreover, we are going to analysis the affordances of 3DVEs for linguistic and non-linguistic knowledge and skills.

3. Methodology

3.1 The process of identifying journals

This research adopted the systemic content analysis as the major method for reviewing studies. We selected six representative refereed journals in the field of technology-enhanced education, in particular, language learning as follows: Language Learning & Technology (LLT), The Journal of EUROCALL (ReCALL), Computer Assisted Language Learning (CALL), The CALICO Journal (CALICO), System (System), Computers & Education (C&E).

Drawing upon the selection criteria by Hsu, Hung and Tsai (2013), we select the above six journals with following features:

- All the journal articles are fully refereed;
- All the journals have a strong devotion to the application of technology for educational purposes, and the first four journals have a particular focus on the application of technology to language teaching and learning;
- Five of the journals are indexed in Social Science Citation Index (SSCI), and the CALICO journal is also chosen for its professional reputation in the field of computer-assisted language learning;
- All the journal articles are published in English;
- All the journals have international readership and authorship.

3.2 The process of data coding and data analyses

In the first stage, we formed a panel of five researchers, one senior researcher and four assistant researchers in the field of CALL. The researchers manually screened publications (from 2010 to 2017) in the above six journals by referring to the titles and the abstracts. When extracting articles, we only selected regular full-length research articles, excluding book reviews, technical reports or commentaries. Moreover, we only screened articles involve two sets of key words: (1) 3D virtual/interactive learning environments, virtual reality (VR), virtual worlds, virtual characters or Second Life; (2) target/foreign language (such as English, French, etc.), (first/second/foreign) language teaching (teachers), (first/second/foreign) language learning (learners), (second/foreign) language education or computer-assisted language learning (CALL). Since the application of virtual reality technology to language education is still a burgeoning research area, only 37 journal articles were identified after the first stage.

In the second stage, we followed the coding framework by Chai, Koh and Tsai (2013), integrating with the analytical framework summarized by Macaro, Handley and Walter (2012), and postulated the preliminary framework for content analysis. Then, we divide the four assistant researchers into two groups (two people in one individual group), and asked them to use the preliminary framework for initial article analysis. Researchers in each group were asked to code the articles independently, without any discussion before the final coding was completed. One article from each of the above journals was selected, and in total six articles were piled for the initial analysis. After the researchers completed the preliminary coding, disagreement and revising suggestions for the analytical framework were proposed for the whole panel discussion. The disagreements were resolved and revision upon the preliminary

analytical framework was amended after the panel discussion. The following table shows the finalized coding framework for in-depth content analysis.

In the third stage, the research panel followed the coding process in the preliminary analysis stage and completed the coding based on the finalized coding scheme. The disagreements were discussed and further resolved in the end. To identify three major types of language play and their subcategories, I, the author, independently coded individual students' data with the help of one trained research assistant. We then discussed data turn by turn by looking at their previous and subsequent discourse.

The Social Science (SPSS 22.0) and Nvivo 11.0 were used to conduct the content analysis in this study. The SPSS was used to do the descriptive data analysis for the coded items.

4. Results

4.1 Number of empirical studies published from 2010 to 2017

According to Table 1, there are in total of 33 empirical studies related to the applications of 3D virtual environments for language learning published in the six journals during 2010 to 2017. In the past eight years, Computers & Education (C&E) and Computer Assisted Language Learning published the largest number of studies (eight studies each). As further indicated by Figure 1, all the six journals have papers published concerning the applications of virtual learning environments to the language education. The solid lines show the yearly publications in the six journals while the dotted line indicates the total number of publications in the six journals. It's quite obvious that the total number of articles published concerning the integration of 3DVEs to language learning is still quite small, largely due to the virtual reality as an emerging technology and its well-recognized complexity of being integrated to the language classrooms (e.g., Kramsch & Steffensen, 2008; Mroz, 2015). Moreover, the year of 2016 witnessed the largest number of publications, in which, seven studies were published in five journals.

Table 1

Numbers of empirical studies published by six journals (2010 to 2017)

	2010	2011	2012	2013	2014	2015	2016	2017	Total
<i>Computers & Education</i>	1	2	0	1	1	0	1	2	8
<i>Computer Assisted Language Learning</i>	1	2	2	1	0	1	1	0	8
<i>Language Learning & Technology</i>	0	1	0	0	1	3	2	0	7
<i>ReCALL</i>	1	0	2	2	0	0	2	0	7
<i>System</i>	0	0	0	0	1	0	0	0	1
<i>The CALICO Journal</i>	0	0	0	0	0	1	1	0	2
Total	3	5	4	4	3	5	7	2	33

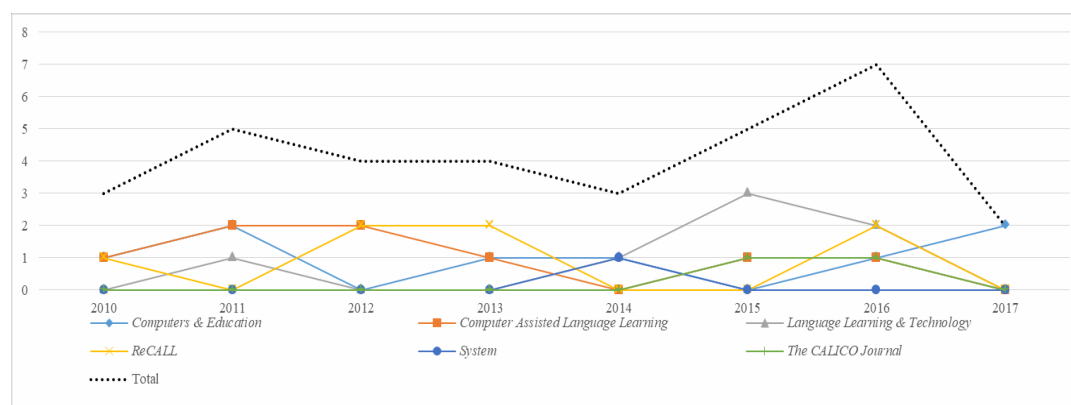


Figure 1. Numbers of empirical studies published by six journals (2010 to 2017).

4.2 Number of empirical studies published from 2010 to 2017

Figure 2 presents the countries or regions of the first author for the 33 empirical studies. It is quite interesting to see that thirteen studies were contributed by scholars from Taiwan, which accounts for over one third of the research in the previous seven years. Researchers from the US (four studies) and Japan (three studies) were also quite active in exploring the application of 3DVEs in language learning. No related investigations were contributed by authors from developing countries. In general, use of styles rather than manual formatting is preferable to enable us to give the proceedings a uniform appearance.

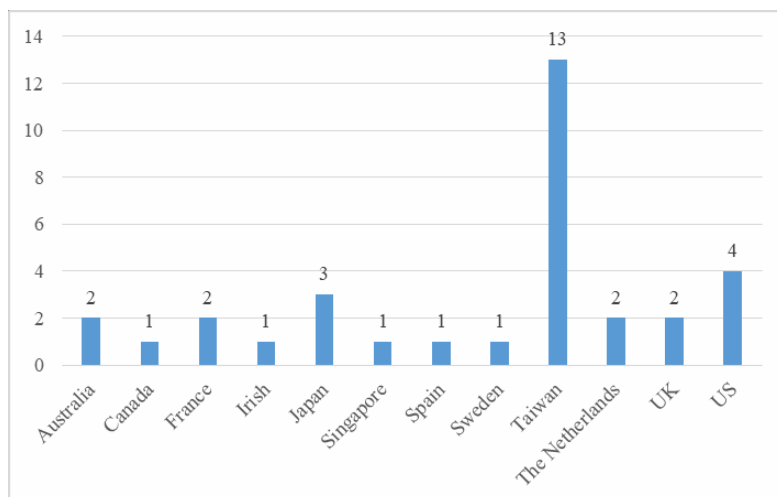


Figure 2. The country or region of the first author

4.3 Research sites of the empirical studies from 2010 to 2017

Quite similar to the Figure 2, about one third of the empirical studies were carried out in Taiwan (thirteen studies). Four studies were conducted in the US, three studies in Japan, and two in the Netherlands and Australia respectively. Morton and Jack (2010) conducted the only cross-country study and evaluated the application of a 3D virtual environment among French language learners in Scotland and EFL learners in mainland China. The environment combines “virtual worlds and virtual agents with automatic speech recognition technology” (Morton & Jack, 2010, p.296), and provides language learners with a unique platform for both verbal and non-verbal communications. Their study further investigated learners’ attitudes and motivations towards the environment and the nature of learner discourses when interacting with others within the virtual setting.

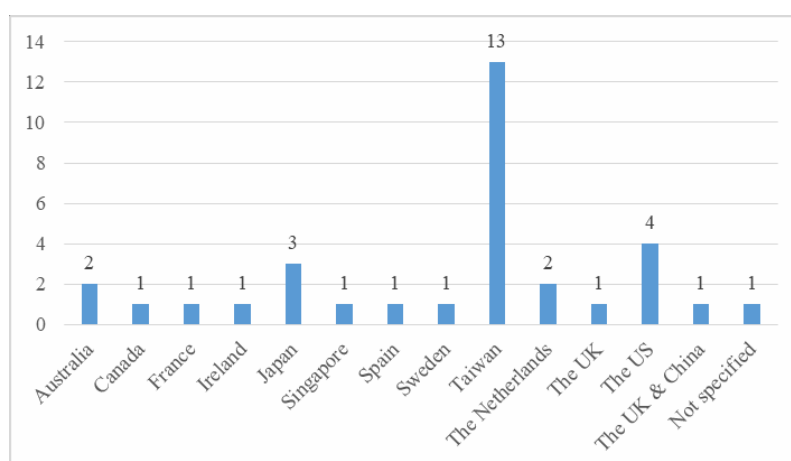


Figure 3. The research sites of the empirical studies

4.4 Target languages of the empirical studies from 2010 to 2017

As suggested in Figure 4, among the 33 empirical studies, it is quite obvious that English is the dominant target language (16 studies). English has already become the lingua franca of the world and its popularity in

the academic, economic and technological field has been well accepted (e.g., Lan, 2015; Lan, Sung, & Chang, 2007; Su, 2006). Therefore, how to learn English well with emerging technologies will still remain as the central topic in the field of CALL.

Worthy of note is that a sizable amount of research (five studies) has explored how to learn Spanish with VR technologies. Three studies explored learning Chinese in 3DVEs and since Chinese is gaining more popularity and we expect the related research may increase in the next few years. Five studies addressed the language education among speakers of different languages. For instance, Levak and Son (2017) investigated the effectiveness of using Second Life and Skype for improving Croatian and English language learners' listening comprehension. Jauregi, Graaff, and Canto (2011) reported learners' experience in a two-year cross-cultural language exchange project, in which sixty interactional tasks were designed for learners of Dutch, Portuguese, Russian and Spanish. In their study, positive negotiations of social and cultural meanings were revealed through the cross-cultural interactions between language learners and native student teachers.

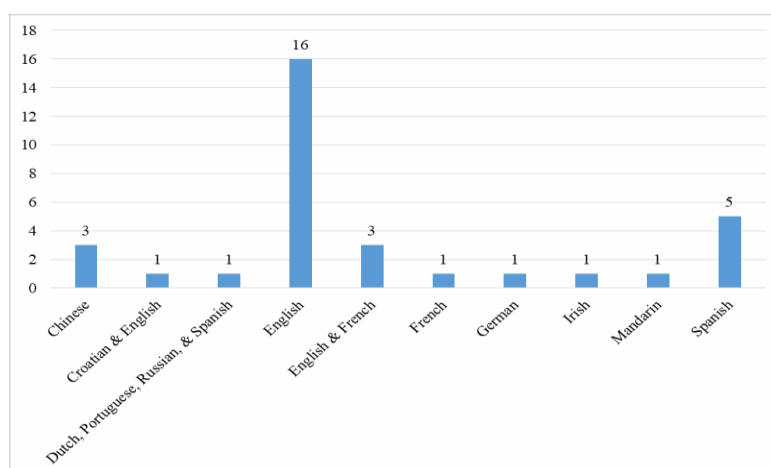


Figure 4. The target language of the empirical studies

4.5 Research settings of the empirical studies from 2010 to 2017

More than half of the articles we reviewed were conducted in high education settings, indicating the popularity of applying the 3DVEs to the language education among adult students. It is easy to understand since adult learners may be more familiar with 3DVEs or be more competent with new technological innovations. An interesting research trend is that researchers have started to realize the potential effectiveness of applying VR or more recently, AR technologies to promoting young learners' performance in learning language (either their mother tongues or a second or foreign language).

As indicated in Figure 5, five studies were conducted in the primary education setting while three studies were carried out in secondary education setting. As Lan (2015) summarized, although plenty of research has focused on the possibility of applying VR technologies, particularly game-based role playing for enhancing language learners' performance, few studies focus on supporting students in elementary schools to improve their self-directed learning. She further proposed three key design principles ("Individuality," "adaptability," and "scaffolding") for improving elementary students' performance in a VR-supported contextual learning environment, providing learners with more individualized and flexible learning experiences. Dalton and Devitt (2016) adopted an action research approach and further elaborated children's learning experience of and attitudes towards 3DVEs. Their findings further indicate that child language learners may prefer a more game-like 3DVE with clearly defined learning objectives and tasks (Dalton & Devitt, 2016). Cheng and Tsai (2014) explored children and parents' reading of picture books equipped with augmented reality (AR) technology, and revealed four interesting behavioral patterns as well as their corresponding cognitive attainment. As they suggested, more studies should be conducted to investigate factors that may affect children's learning behaviors when traditional learning materials integrated with VR or AR technologies (Cheng & Tsai, 2014). Hsu (2017) investigated learners variables, such as cognitive load, foreign language anxiety and effectiveness of learning among third-graders in primary school and offered experimental data for a better understanding of children learners' learning styles when engaged in AR educational games.

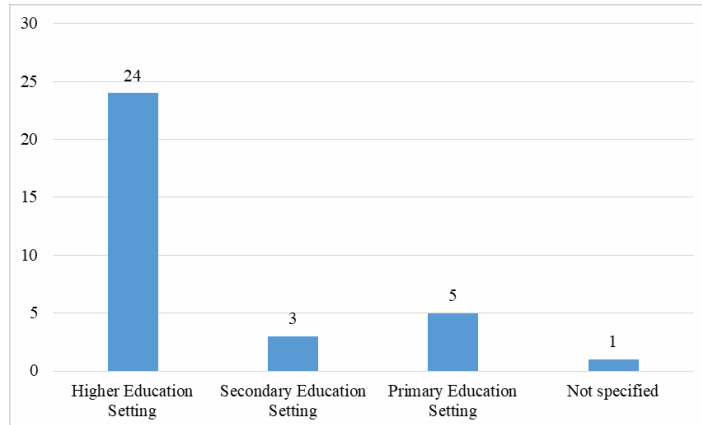


Figure 5. The research settings of the empirical studies

4.6 Affordances of 3D Virtual Environments for Language Learning

4.6.1 Affordance of 3D Virtual Environments for Linguistic Knowledge and Skills

Among the reviewed articles, we found nineteen studies focusing on the affordance of the 3DVEs for improving language learners' linguistic knowledge and skills. As we can see from Figure 6, over half of these studies (thirteen articles) explored the potential benefits of the 3DVEs for enhancing language learners' overall language proficiency, language performance or target language output. Others dealt with specific linguistic skills achieved through effective learning tasks in the 3DVEs or AR materials, such as listening (Levak & Son, 2017), speaking (Lan, 2014; Lan et al., 2016; Morton & Jack, 2010), writing (Collentine, 2011; Wang, 2017) and vocabulary retention (Franciosi et al., 2016).

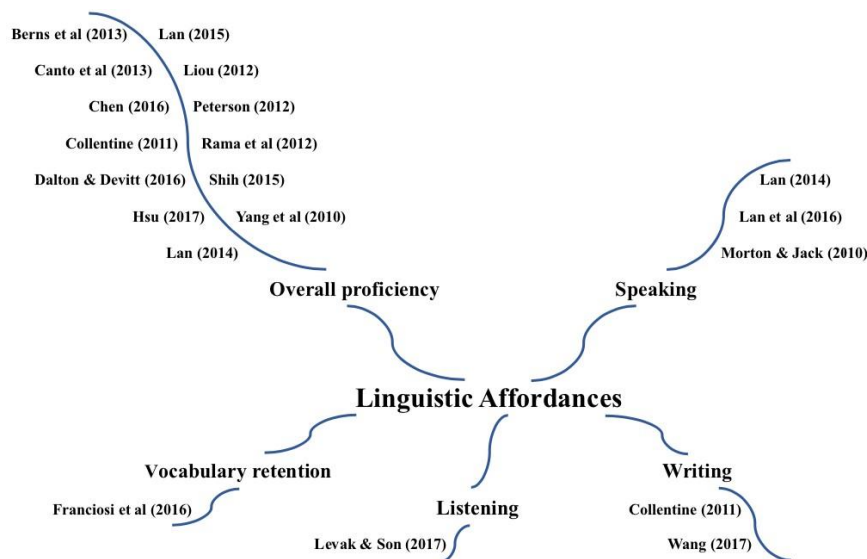


Figure 6. Distribution on the topic of affordances of 3DVE for linguistic skills

4.6.2 Affordance of 3D Virtual Environments for non- Linguistic Knowledge and Skills

Fourteen studies were identified as to explore the affordance of virtual reality technology for improving learners' non-linguistic knowledge and skills (Figure 7). A variety of aspects concerning their

non-linguistic knowledge and skills were discussed, among which, a majority of studies (eleven articles) centered on fostering language learners' communicative skills, strategies or their social cultural interactions through the applications of 3DVEs. The benefits of 3DVEs for other non-linguistic knowledge and skills also include boosting learners' critical thinking (Mroz, 2015), their cognitive attainment when using AR materials or playing AR games (Cheng & Tsai, 2014; Hsu, 2017), cultural learning (Shih, 2015), and collaborative learning skills (Ho et al., 2011; Kozlova & Priven, 2015). Be more specific, (Shih, 2015) discussed how learners' learning experience through the virtual environment may impact their cultural knowledge acquisition, positive attitudes toward the target culture and their cross-cultural adaption or adjustment. (Ho et al., 2011) explored learners' collaborative skill in the 3DVE while 27 proved teachers' collaborative learning through a 3DVE.

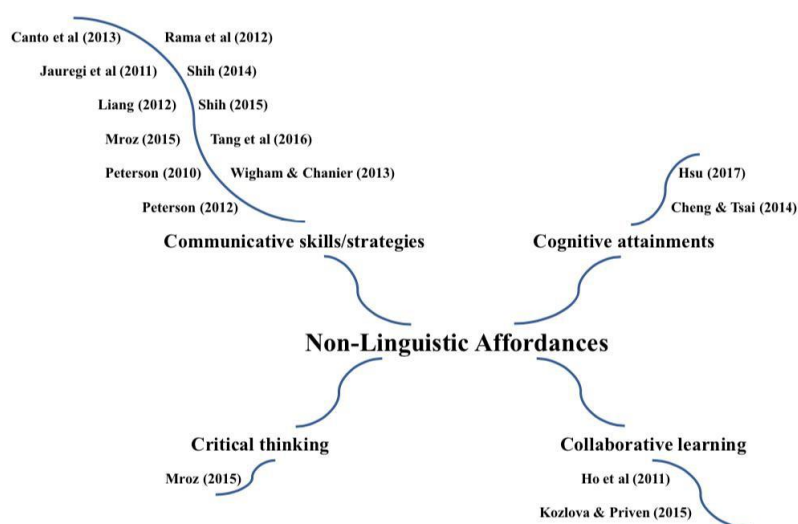


Figure 7. Distribution on the topic of affordances of 3DVE for non-linguistic skills

5. Conclusion

The present research is a systematic review of 33 empirical studies selected from six high-impact academic journals concerning the application of 3D virtual environments in language learning. Content analysis method was employed and a coding scheme was constructed from aspects of publication trend, location distribution of the first author, research sites, target language, research settings, and affordances of 3DVEs for language learning. Drawing upon the findings, we conclude that two essential issues on the application of 3DVEs in language learnings are “why to use it” and “how to use it”. The key to the application of 3DVEs in language teaching lies in the design of teaching activities. It requires the cooperation of front-line teachers, SLA scholars as well as technical specialists to fully realize the potential of 3DVEs in fostering better language teaching effect.

Acknowledgements

This research is funded by Fok Ying Tung Education Foundation (161093) and Postgraduate Teaching Reform Project of Beijing University of Posts and Telecommunications (2019Y002) awarded to Dr. Chunping Zheng.

References

- Berns, A., Gonzalez-Pardo, A., & Camacho, D. (2013). Game-like language learning in 3-D virtual environments. *Computers & Education, 60*(1), 210-220.
- Canto, S., Jauregi, K., & van den Bergh, H. (2013). Integrating cross-cultural interaction through video-communication and virtual worlds in foreign language teaching programs: Is there an added value?. *ReCALL, 25*(1), 105-121.

- Chai, C. S., Hwee Ling Koh, J., & Tsai, C. C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society, 16*(2).
- Chen, J. C. (2016). The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life. *Computers & Education, 102*, 152-171. doi:10.1016/j.compedu.2016.08.004
- Chen, J. F., Warden, C. A., Tai, D. W.-S., Chen, F.-S., & Chao, C.-Y. (2011). Level of abstraction and feelings of presence in virtual space: Business English negotiation in Open Wonderland. *Computers & Education, 57*(3), 2126-2134. doi:10.1016/j.compedu.2011.05.017.
- Cheng, K. H., & Tsai, C. C. (2014). Children and parents' reading of an augmented reality picture book: Analyses of behavioral patterns and cognitive attainment. *Computers & Education, 72*, 302-312.
- Clarke, J., Dede, C., & Dieterle, E. (2008). *Emerging technologies for collaborative, mediated, immersive learning*. In International handbook of information technology in primary and secondary education (pp. 901-909). Springer, Boston, MA.
- Collentine, K. (2011). Learner autonomy in a task-based 3d world and production. *Language Learning & Technology, 15*(3), 50-67.
- Dalton, G., & Devitt, A. (2016). Irish in a 3d world: engaging primary school children. *Language Learning & Technology, 20*(1), 21-33.
- Franciosi, S. J., Yagi, J., Tomoshige, Y., & Ye, S. (2016). The effect of a simple simulation game on long-term vocabulary retention. *Calico Journal, 33*(3).
- Ho, C. M., Nelson, M. E., & Müller-Wittig, W. (2011). Design and implementation of a student-generated virtual museum in a language curriculum to enhance collaborative multimodal meaning-making. *Computers & Education, 57*(1), 1083-1097.
- Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? *Computers & Education, 106*, 137-149.
- Jauregi, K., Canto, S., De Graaff, R., Koenraad, T., & Moonen, M. (2011). Verbal interaction in Second Life: towards a pedagogic framework for task design. *Computer Assisted Language Learning, 24*(1), 77-101.
- Kozlova, I., & Priven, D. (2015). ESL teacher training in 3d virtual worlds. *Language Learning & Technology, 19*(1), 83-101.
- Kramsch, C., & Steffensen, S. V. (2008). *Ecological perspectives on second language acquisition and socialization*. In Encyclopedia of Language and Education (pp. 2595-2606). Springer, Boston, MA.
- Lan, Y. J., Sung, Y. T., & Chang, K. E. (2007). A mobile-device-supported peer-assisted learning system for collaborative early EFL reading. *Language Learning & Technology, 11*(3), 130-151.
- Lan, Y. J. (2014). Does second life improve mandarin learning by overseas Chinese students?. *Language Learning & Technology, 18*(2), 36-56.
- Lan, Y. J. (2015). Contextual efl learning in a 3d virtual environment. *Language Learning & Technology, 19*(192), 16-31.
- Lan, Y. J., Kan, Y. H., Sung, Y. T., & Chang, K. E. (2016). Oral-Performance Language Tasks for CSL Beginners in Second Life. *Language Learning & Technology, 20*(3), 60-79.
- Levak, N., & Son, J. B. (2017). Facilitating second language learners' listening comprehension with Second Life and Skype. *ReCALL, 29*(2), 200-218.
- Liou, H. C. (2012). The roles of Second Life in a college computer-assisted language learning (CALL) course in Taiwan, ROC. *Computer Assisted Language Learning, 25*(4), 365-382.
- Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching, 45*(1), 1-43.
- Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: a cross-cultural evaluation. *Computer Assisted Language Learning, 23*(4), 295-319.
- Mroz, A. (2015). The development of second language critical thinking in a virtual language learning environment: a process-oriented mixed-method study. *Calico Journal, 32*(3), 528-553.
- Peterson, M. (2010). Learner participation patterns and strategy use in Second Life: an exploratory case study. *ReCALL, 22*(3), 273-292.
- Peterson, M. (2012). EFL learner collaborative interaction in Second Life. *ReCALL, 24*(1), 20-39.
- Rama, P. S., Black, R. W., van Es, E., & Warschauer, M. (2012). Affordances for second language learning in World of Warcraft. *ReCALL, 24*(3), 322-338.
- Shih, Y. C. (2014). Communication strategies in a multimodal virtual communication context. *System, 42*, 34-47.
- Shih, Y. C. (2015). A virtual walk through london: culture learning through a cultural immersion experience. *Computer Assisted Language Learning, 28*(5), 407-428.
- Stevens, V. (2006). Second Life in education and language learning. *TESL-EJ, 10*(3), 1-4.
- Stockwell, G. (2007). A review of technology choice for teaching language skills and areas in the call literature. *Recall, 19*(2), 105-120.
- Su, Y. C. (2006). EFL teachers' perceptions of English language policy at the elementary level in Taiwan. *Educational Studies, 32*(3), 265-283.

- Wagner, C., & Ip, R. K. (2009). Action learning with second life--A pilot study. *Journal of Information Systems Education, 20*(2), 249.
- Wang, Y. H. (2017). Exploring the effectiveness of integrating augmented reality-based materials to support writing activities. *Computers & Education, 113*, 162-176.
- Wigham, C. R., & Chanier, T. (2013). A study of verbal and nonverbal communication in Second Life--the ARCHI21 experience. *ReCALL, 25*(1), 63-84.
- Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education, 55*(3), 1346-1356.

Enhancing Low Achievers' EFL Learning with Interactive Digital Technologies

*Shu-Yuan LIN^a, Ming-Puu CHEN^a, Li-Chun WANG^a, Yu-Ting KAO^a, Di ZOU^b, Haoran XIE^c
a Graduate Institute of Information & Computer Education, National Taiwan Normal University, Taipei, Taiwan
b Department of English Language Education, The Education University of Hong Kong, Hong Kong
c Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
*etchere@gmail.com

Abstract: Interactive digital technology has great potentiality in providing life-like learning contexts and in-time interactions to facilitate language learning, especially for the low achievers. Contexts are essential for effective language learning. In this study, two types of digital interactive technologies, including interactive digital map (iMap) and augmented reality (AR), were employed to deliver and enhance contextualized learning experiences with gamified learning tasks for the low achievers'. The purpose of this study was to investigate the effects of type of technology-enhanced learning, including the iMap-enhanced learning and the AR-enhanced learning, on low achievers' learning performance and attitude while learning from the gamified technology-enhanced contextualized EFL learning. A preliminary experiment showed that (a) the AR-enhanced learning group outperformed the iMap-enhanced learning group on learning performance, (b) all participants' post attitudes toward the received technology-enhanced learning were significantly increased in all attitude aspects of confidence, preference, anxiety, attention, and learning strategy, and (c) both technology-enhanced learning group revealed similar positive in most attitude aspects, excepted that the AR group showed a higher degree of attention than the iMap group. It was concluded that technology-enhanced contextualized learning is effective in promoting learning attitudes and helping EFL learners achieve acceptable learning performance.

Keywords: gamification, augmented reality, language learning, English as a foreign language

1. Introduction

Contexts are essential to meaningful information processing and knowledge acquisition for language learning. The learners need to consciously attend to the language input so that the perceived information can be further processed in the working memory, but novice learners usually lack the capability in attending to the key information and result in poor comprehension efficiency. Furthermore, digital interactive technology has great potential in providing life-like learning contexts and in-time interactions to facilitate language learning. Therefore, technology-enhanced contextualization is conducive to effective language learning as it emphasizes the need of language learning in authentic and interactive situations. Thus, this study employed two types of digital interactive technologies, including interactive digital map (iMap) and augmented reality (AR), to deliver and enhance contextualized learning experiences with gamified learning tasks for the learners and expected that the low achievers can benefit from the context-rich language learning activities and bring about effective meaningful learning.

2. Related work

Learning English as a Second or Foreign language (ESL or EFL) is a global imperative as English plays an essential role in the international society (Smith et al., 2013). With the development of innovative technologies, diverse approaches to technology-enhanced language learning have emerged in recent

years, providing learners with personalized, contextualized, and socialized learning experiences. Among various technologies, augmented reality (AR), virtual reality (VR), and context-aware ubiquitous learning systems have been frequently noted in the literature as facilitative for language learning, one common feature of which is that they are effective in contextualizing the learning content.

The use of AR has been increasingly popular in educational settings (Azuma et al., 2001). The literature indicates various advantages of AR, among which, the most significant one is its ability to create immersive hybrid learning environments where learning situations are contextualized through combining digital and physical objects (Akçayır & Akçayır, 2017; Dunleavy, Dede, & Mitchell, 2009). This contextualization is realized by visualizing concepts and events through displaying virtual elements over real objects (Wu, Lee, Chang, & Liang, 2013). It enables students to immerse themselves in the learning process, raises the level of engagement, and enhances educational outcomes.

Similarly, context-aware ubiquitous learning delivered by hand-held devices and smartphones has been widely acknowledged as being conducive to effective learning. With wireless communication and GPS facilities, context-aware ubiquitous learning provides learners with on-site interactions with physical objects and digital multimedia, contextualizes learning content in concrete and appropriate situations, and emphasizes the need of language learning in authentic and interactive situations. This new approach extends the applications of location-based learning and situated learning, enabling learners to interact with real-world learning targets without constraints of time or place in real situations (Ogata & Yano, 2004). Thus, it has great potentiality in promoting EFL learners' motivation and learning effectiveness.

The nature of how context-aware ubiquitous learning and AR-enhanced learning promote learning is to some extent similar in facilitating learners' sense of contextualization. They both contextualize learning contents, integrate virtual and real-world learning environments, and support real-time interactions. Despite a large number of studies on AR-enhanced learning and context-aware ubiquitous learning respectively, little research has been conducted to compare these two, and thereby the present study aimed to examine the similarity and difference between the AR-enhanced learning and the context-aware ubiquitous learning. Furthermore, as learners' attitudes towards technology-enhanced contextualized learning may vary according to the types of innovative technologies, and there has never been any attempt that compared the use of AR and interactive digital map in contextualized language learning, although these two have been investigated by several studies before, it is necessary to investigate and compare the effects of AR and interactive digital map in promoting contextualized language learning.

3. Implementations

3.1 Research design

An experimental design was implemented to investigate the effects of type of technology-enhanced learning on low English achievers of junior high school students' learning performance and attitude while learning from the technology-enhanced contextualized learning. Two types of technology-enhanced learning – the interactive map (iMap) learning and the augmented-reality (AR) learning – were implemented in the study. The participants, the experiment, data collection, and the learning activities are introduced as follows.

Sixty-five junior high students whose English achievement was under the medium score in Comprehensive Assessment Program for Junior High School Students in Taiwan participated in the experimental activities voluntarily. All participants had taken 4-class-per-week EFL courses since the seventh grade and were expected to acquire the basic grammar and the most frequently used 2000 English words through three years of study. The participants were invited to participate in the experimental technology-enhanced learning activities as a comprehensive practice for their EFL learning using tablet-PC individually. The experiment began with an orientation to the experiment project, followed by a 40-minute gamified contextualized learning activity, the iMap-enhanced contextualized learning and the AR-enhanced contextualized learning for the iMap group and the AR group, respectively. An achievement test and an attitude questionnaire survey were administered immediately after the treatment.

Analysis of Covariance (ANOVA) was conducted to examine participants' learning performance between groups with prior content knowledge as covariant to eliminate prior knowledge effect on learning achievement, paired t-tests were employed to examine participants' attitude changes after receiving the treatment, and Multivariate Analysis of Variance (MANOVA) was implemented to examine the differences in participants' attitude changes between groups. A significance Alpha level of .05 was used for the statistical analysis.

3.2 The treatment learning activity

Two types of technology-enhanced theme-based contextualized learning were implemented to serve as the treatments, including the iMap-enhanced learning and the AR-enhanced learning, to enhance participants' sense of contextualization of the story theme. The participants conducted their received learning activity using tablet-PCs individually. Two types of technology, iMap and AR, were employed to enhance the contextualization of the theme-based learning by navigating among learning units and triggering the target learning activity, respectively.

The two versions of technology-enhanced theme-based learning employed the same 3-learning-stages and gamified features, including (1) Learning from video dialogue, (2) Learning from pictorial vocabulary, and (3) Practicing what you learned. At Stage 1, participants conducted the learning tasks given on the associated learning worksheet and learn from the theme-based video dialogues by recording specific keywords and the meanings of specific dialogues. At Stage 2, the tasks were focused on learning from the given pictorial English vocabulary, phrases and sample sentences. The content knowledge included 12 target words and the application of these words in purposeful conditions. Finally, at Stage 3, a practice activity was conducted for the learners to apply what they have learned in previous tasks to some given contexts in order to gain mastery.

Moreover, a gamified framework was employed to encourage the participants to engage in the learning tasks by giving them specific numbers of awarding stars according to the levels of their task performance at each learning stage. Participants' final cumulated performance was presented along with the illustrations of their received stars.

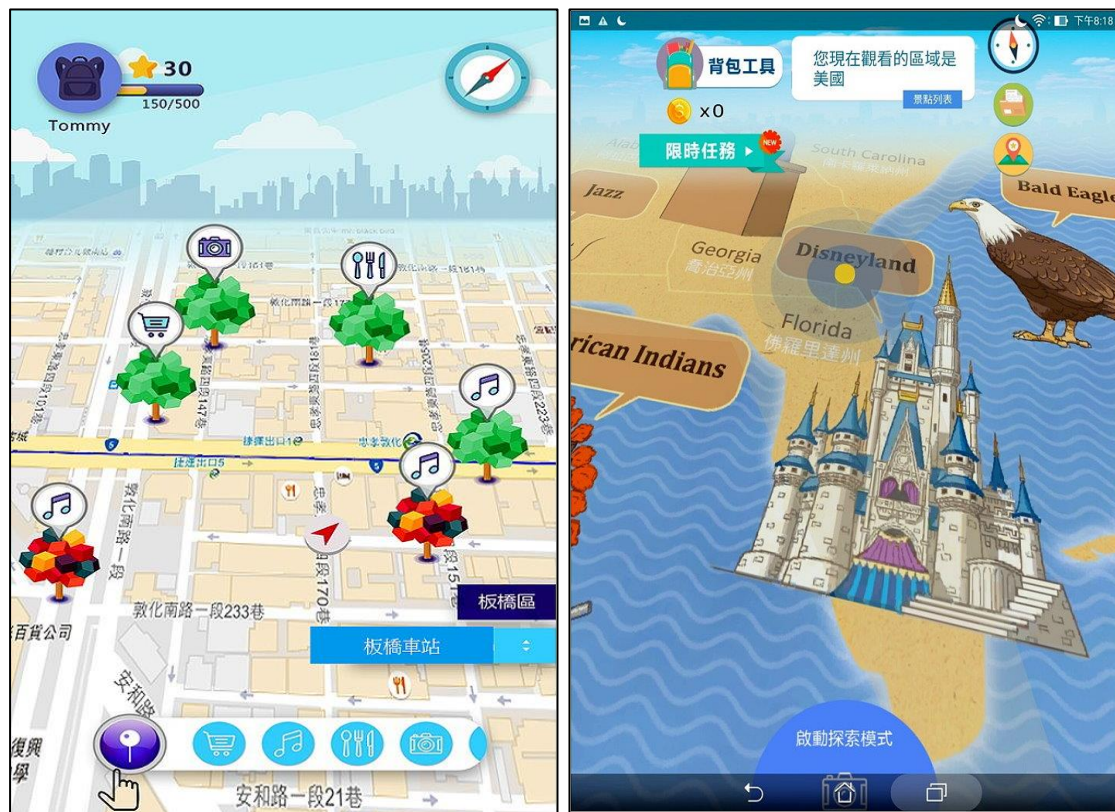


Figure 1. The interactive digital map (iMap, on the left) and the interactive AR (on the right) assist learners to navigate and learn the theme-based content.

As shown on the left in Fig. 1, the iMap-enhanced theme-based learning utilizes the navigation function and the GPS positioning of a digital map and allows learners to explore and learn the geographic-arranged theme-based content through intuitive navigation or real-time GPS positioning. Furthermore, as shown on the right in Fig. 1, the AR-enhanced theme-based learning utilizes the AR recognition function and allows learners to explore and learn the geographic-arranged theme-based content through AR interactions.

An English content knowledge test was administered prior to and immediately after the treatment to evaluate participants' previous acquired prior knowledge and learning achievement from the treatment, respectively. Similarly, a 5-aspect 25-item attitude questionnaire was conducted prior to and after the treatment to survey participants' learning attitudes toward previous EFL learning and their received versions of technology-enhanced contextualized EFL learning, respectively.

4. Results

4.1 Analysis on learning performance

The group means of participants' learning performance in English content knowledge are shown in Table 1. The overall mean scores reached acceptable levels of 49.49% and 56.51% of correct-answer-rate for the iMap group and the AR group, respectively. The AR group had a higher mean score and lower deviation than the iMap group. Analysis of Covariance (ANCOVA) was conducted with prior content knowledge as covariant to examine whether the difference of participants' learning performance between groups is significant. As shown in Table 2, the ANCOVA summary indicates that prior content knowledge was significant on learning achievement. That is to say, employing prior content knowledge as covariant in the analysis can eliminate the impact of prior content knowledge on participants' learning achievement. Moreover, the type of technology-enhanced learning is also significant ($F(1, 62)=13.07, p<.01$), indicating that the AR group outperformed the iMap group while learning from the technology-enhanced learning. In other words, the embodied AR-enhanced learning seemed to bring about better learning achievement than the virtual navigation of the digital map did.

Table 1
Group Means and Standard Deviations of Learning Performance

GRP	Mean	S.D.	n
iMap	49.49	8.98	32
AR	56.51	7.12	33
Total	53.25	8.71	56

Table 2
ANCOVA Summary on Learning Performance with Prior Content Knowledge as Covariant

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Prior Knowledge	279.33	1	279.33	4.62	.04	.08
Group	789.62	1	789.62	13.07	.00	.20
Error	3202.57	2	60.43			

4.2 Analysis of learning attitude

As shown in Table 3, the group means and standard deviations of pre-attitude and post-attitude indicate that participants seemed to possess higher confidence, preference, attention, and learning strategy, but lower anxiety. The summary of paired t-tests, as shown in Table 4, indicates that participants' attitude changes before and after the treatment are significant in all attitude aspects. That is to say, the implemented technology-enhanced learning brought about a higher degree of confidence, preference, attention, and learning strategy, but lower anxiety for the participants. Whether different types of technology-enhanced learning has a similar or different impact on attitude is further analyzed as follows.

As shown in Table 5, the group means and standard deviations of attitude changes the iMap group and the AR group suggest that the AR group seemed to obtain higher attitude changes in all

aspects of confidence, preference, anxiety, attention, and learning strategy. Multivariate Analysis of Variance (MANOVA) was implemented to examine the differences in participants' attitude changes between the experimental groups. The results suggest that both the AR group and the iMap group obtained a similar degree of attitude increases in confidence, preference, attention, and learning strategy, but the AR group possessed a higher attention increase than the iMap learners. The results suggest that the embodied AR-enhanced learning seemed to draw more attention of the learners and engage them in the learning activity than the virtual navigation on the digital map of the iMap learners did. It is reasonable to infer that the embodied AR interactions can engage the learners in the information processing efficiently and, therefore, result in meaningful and effective learning.

Table 3
Group Means and Standard Deviations of Pre-Attitude and Post-Attitude

	Attitude Aspect	Mean	Std. Deviation	n
Pair 1	-Confidence	2.59	1.50	65
	†t-Confidence	3.04	.91	65
Pair 2	-Preference	2.52	1.48	65
	†t-Preference	3.02	.93	65
Pair 3	-Anxiety	3.18	1.34	65
	†t-Anxiety	2.54	.60	65
Pair 4	-Attention	2.45	1.29	65
	†t-Attention	3.05	.57	65
Pair 5	-Strategy	2.61	1.45	65
	†t-Strategy	3.25	.80	65

Table 4
Summary of T-Tests on Attitude Differences

	Attitude Aspect	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Pair 1	†t-Confidence – Pre-Confidence	.46	1.27	2.04	64	.05
Pair 2	†t-Preference – Pre-Preference	.50	1.22	2.32	64	.03
Pair 3	†t-Anxiety – Pre-Anxiety	-.64	1.15	3.15	64	.00
Pair 4	†t-Attention – Pre-Attention	.60	1.11	3.05	64	.01
Pair 5	†t-Strategy – Pre-Strategy	.65	1.25	2.93	64	.01

Table 5
Group Means and Standard Deviations of Attitude Difference for the iMap and AR Groups

Attitude Difference	GRP	Mean	Std. Deviation	n
Confidence	iMap	.21	1.27	32
	AR	.46	1.07	33
	Total	.33	1.17	65
Preference	iMap	.24	1.22	32
	AR	.50	.97	33
	Total	.37	1.10	65
Anxiety	iMap	.32	1.15	32
	AR	.64	1.04	33
	Total	.47	1.10	65
Attention	iMap	.06	1.11	32
	AR	.60	1.02	33
	Total	.33	1.09	65
Strategy	iMap	.22	1.25	32
	AR	.65	1.16	33
	Total	.43	1.22	65

Table 6
MANOVA Summary of Type of Technology on Attitude Differences

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
GRP	Confidence	1.02	1	1.02	.74	.39	.01
	Preference	1.08	1	1.08	.89	.35	.01
	Anxiety	1.69	1	1.69	1.41	.24	.02
	Attention	4.75	1	4.75	4.17	.05	.06
	Learning Strategy	2.99	1	2.99	2.05	.16	.03
Error	Confidence	86.78	63	1.38			
	Preference	76.51	63	1.21			
	Anxiety	75.40	63	1.20			
	Attention	71.86	63	1.14			
	Learning Strategy	91.95	63	1.46			

5. Conclusions

This study employed a gamified framework to implement two types of technology-enhanced contextualized learning to promote learners to learn theme-based EFL learning content by means of 3-stage of learning activities of Learning from video dialogues, Learning from pictorial vocabulary, and Practicing what you learned. The employed gamified framework aimed to encourage the participants to engage in the learning tasks by giving them awarding stars in accordance with their task performance. The AR-enhanced learning and the iMap-enhanced learning were employed to provide life-like theme-based contexts and in-time embodied and virtual interactions to facilitate EFL learning. The preliminary analysis confirmed the research hypotheses as follows. Firstly, both types of technology-enhanced learning brought about acceptable learning effectiveness, and the embodied AR interactions can engage the learners in the information processing efficiently and, therefore, result in meaningful and effective learning. Secondly, all participants showed positive attitude changes toward technology-enhanced learning in confidence, preference, attention, and learning strategy, but lower anxiety, and it suggested that technology-enhanced contextualized learning is effective in promoting learning attitudes and helping EFL learners achieve acceptable learning performance. Thirdly, the embodied AR-enhanced learning draw more attention of the learners and engaged them in the theme-based learning contexts than the digital map virtual navigation of the iMap did, and therefore, it resulted in meaningful and effective learning. Finally, future studies are suggested to examine the superiority of AR interactions in facilitating learners' sense of contextualization and engagement in learning tasks in detail.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under the grant MOST 106-2511-S-003 -018 -MY3.

References

- Akçayır, M. & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, 1-11.
- Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6), 34-47.
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18(1), 7-22.
- Huang, T. C., Chen, C. C., & Chou, Y. W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, 96, 72-82.

- Huang, Y. Y., Liu, C. C., Wang, Y., Tsai, C. C., & Lin, H. M. (2017). Student engagement in long-term collaborative EFL storytelling activities: An analysis of learners with English proficiency differences. *Journal of Educational Technology & Society*, 20(3), 95-109.
- Ibáñez, M. B. & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, 123, 109-123.
- Liu, S. H. J. (2017). Text-based negotiated interaction of NNS-NNS and NNS-NS dyads on Facebook. *ReCALL*, 29(3), 294-312.
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge university press.
- Ogata, H. & Yano, Y. (2004). Knowledge awareness for a computer-assisted language learning using handhelds. *International Journal of Continuing Engineering Education and Life Long Learning*, 14(4-5), 435-449.
- Smith, G. G., Li, M., Drobisz, J., Park, H. R., Kim, D., & Smith, S. D. (2013). Play games or study? Computer games in eBooks to learn English vocabulary. *Computers & Education*, 69, 274-286.
- Sommerauer, P. & Müller, O. (2014). Augmented reality in informal learning environments: a field experiment in a mathematics exhibition. *Computers & Education*, 79, 59-68.
- Tsai, S. C. (2017). Effectiveness of ESL students' performance by computational assessment and role of reading strategies in courseware-implemented business translation tasks. *Computer Assisted Language Learning*, 30(6), 474-487.
- Wang, H. Y., Liu, G. Z., & Hwang, G. J. (2017). Integrating socio-cultural contexts and location-based systems for ubiquitous language learning in museums: A state of the art review of 2009–2014. *British Journal of Educational Technology*, 48(2), 653-671.
- Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41-49.
- Yang, J. C. & Quadir, B. (2018). Effects of prior knowledge on learning performance and anxiety in an English learning online role-playing game. *Journal of Educational Technology & Society*, 21(3), 174-185.
- Yang, M. T. & Liao, W. C. (2014). Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Transactions on Learning Technologies*, 7(2), 107-117.

Design and Development of a Conjunctive Word-Learning Support System for Conjunctive Expressions with Different-Meaning Commutativity and its Experimental Use

Kengo IWAI^a, Fonteles Furtado PEDRO GABRIEL^b, Yu OGATA^c, Yusuke HAYASHI^d,
Tsukasa HIRASHIMA^e

^{a,b,c,d,e}*Graduate School of Engineering, Hiroshima University, Japan*

*iwai@lel.hiroshima-u.ac.jp

Abstract: Conjunctive words connect two sentences logically. This property is divided into same-meaning and different-meaning commutativity. The former emphasizes a meaning that is explicitly represented in the sentences that already exist, but the latter adds a meaning beyond what is explicitly represented in them. Doing exercises to learn how to use both the former and the latter is important because conjunctive words play an important role in reading and writing effective sentences. Exercises to learn the former are easy, as there is only one correct answer to a fill-in-the-blank problem. On the other hand, to demonstrate mastery of the latter, the exercises must be difficult because various correct answers with different meanings exist. In this study, we designed and developed a conjunctive word learning support system for performing conjunctive word exercises to master different-meaning commutativity. In this system, the learner is provided with an exercise that entails assembling a sentence structure composed of two sentences and a conjunctive word and a conjunctive logical structure composed of two explicit situations and an implicit situation. The system also automatically diagnoses the learner's performance on the exercises and gives feedback based on the diagnosis results. Furthermore, we carried out an experiment to verify the validity of this system, and the results suggest that the system is effective for helping students to master this aspect of language.

Keywords: Conjunctive word, different-meaning commutativity, automatic diagnosis

1. Introduction

Connective words play an important role in both reading and writing (Aidinlou & Mehr, 2012; Manan & Raslee, 2018), and they are sometimes called “discourse markers” or “linking words” (Fox Tree, 2010; Yale Center, 2012). Connective words are especially important in Japanese-language learning because they play an important role in reading and writing effective sentences and are one of the most important elements of the language (Ichikawa, 1978; Baba, 2005; Yamamuro, 2008). Therefore, we focus on the role of connective words in Japanese-language learning. The role of conjunctive words is to logically connect the contexts that exist before and after them. Furthermore, if the context is understood semantically, the conjunctive word is not always required. In this case, the conjunctive word does not add a new meaning to the existing context. In fact, a conventional exercise on conjunctive words is to choose a conjunctive word that represents a connection/relationship. Whether the relationship is appropriate for the context is determined only by the sentences before and after the conjunctive word.

On the other hand, many studies have pointed out that an important role of conjunctive words is to give the reader a hint to make the future context easier to discern during the process of reading comprehension (Ujiie, 1973; Itou & Abe, 1991; Ishiguro, 2008). These words also play a role in establishing the context beforehand. In other words, conjunctive words indicate meaning in the process of reading comprehension and description, although the reason for the choice of conjunctive words is

obvious after all contexts become clear. At times, the meaning of the sentences is changed by the choice of conjunctive words. In this study, we define this property of conjunctive words as “different-meaning commutativity.” This definition is described in Chapter 2.

In this study, we designed and developed a conjunctive word-learning support system for mastering different-meaning commutativity. The implementation of such an exercise was not easy because it needed to target the implication that was changed by switching conjunctive words. In this study, the information structure of a conjunctive expression is understood as consisting of a sentence structure and a conjunctive logical structure. Furthermore, the sentence structure consists of three elements of two sentences and a conjunctive word that connects the sentences. The conjunctive logical structure consists of three elements of two explicit situations that the sentences that comprise the sentence structure explicitly express and an implied situation suggested by the conjunctive words. In correspondence with these two structures, we articulate a framework in which a conjunctive expression with different-meaning commutativity and its meanings constitute a mutual dependence between a conjunctive word and an implied situation model. As an exercise based on this model, we designed an activity that entailed assembling and manipulating the sentence and conjunctive logical structures and developed an environment in which the system could diagnose students’ performance and give feedback on the results of the activity (Ogata et al., 2015). Furthermore, to verify the validity of this system, an experiment using it and an evaluation of the results were performed.

2. Different-Meaning Commutativity

A conjunctive word provides a logical connection between the sentences before and after it. This is called a “conjunctive expression.” A conjunctive expression consists of two sentences and a conjunctive word that connects them. In addition, this expression holds even if the conjunctive words in the conjunctive expression change. This property is often called “commutativity” (Itou & Abe, 1991). Due to this property, the meaning does not change or the meaning changes as the conjunctive expression changes. We call commutativity that does not change the meaning “same-meaning commutativity” and commutativity that changes the meaning “different-meaning commutativity.” In this study, we focus on conjunctive expressions that reflect different-meaning commutativity. The following is an example of this type of commutativity (Ishiguro, 2008):

(Conjunctive expression 1) “I practiced hard every day and participated in the tournament, so I came in fourth out of twenty people.”

(Conjunctive expression 2) “I practiced hard every day and participated in the tournament, but I came in fourth out of twenty people.”

In these two conjunctive expressions, the two clauses before and after the conjunctive word are the same. Therefore, the meanings explicitly expressed are common to both. However, the meanings implied by the conjunctive expressions are different because of the difference in the conjunctive word. In conjunctive expression 1, the expression implies that coming in fourth out of twenty people is a reasonable result of practicing hard every day. On the other hand, in conjunctive expression 2, the expression implies that coming in fourth out of twenty people is a disappointing result after practicing hard every day. Thus, different-meaning commutativity entails a change in meaning when the conjunctive word is changed.

The importance of different-meaning commutativity has been pointed out in many studies. According to one interpretation, the conjunctive word is not meant to represent pure or objective logic, and conjunctive expressions include the subjective intentions of a writer as their premise (Ujiie, 1973). Through them, the individual logic of the writer is expressed. The importance of the role of conjunctive words has also been pointed out by scholars. Such words aid in the recognition of speakers and impart the subjective situational understanding of writers (Kawabata, 2009). In addition, another function of conjunctive words is to help the reader understand what a writer is implying (Ishiguro, 2008). This is regarded as a creative aspect of conjunctive words, and the implication appears logical if the subjectivity of writers is accepted by readers. On the other hand, it has been noted that exercises to examine conjunctive expressions with different-meaning commutativity have not been conducted

before (Ishiguro, 2008). Therefore, it is important to provide support for mastering the use of conjunctive expressions that involve commutativity.

3. Modeling a Conjunctive Expression with Different-Meaning Commutativity

3.1 Mutual Dependence between Conjunctive Words and the Implied Situation Model

The mutual dependence between conjunctive words and the implied situation model is a framework for representing the different-meaning commutativity model. The model represents the mutual dependence between the conjunctive word and the implied situation. A diagram of this model is shown in Fig. 1. The conjunctive expression consists of two sentences, with a conjunctive word placed between them. In addition, such sentences are composed of a front sentence and a back sentence. This is called “sentence structure.” In addition, the meaning of the sentence structure consists of an explicit situation and an implicit situation. The explicit situation represents an obvious connection between the front and back sentences. The implicit situation represents a situation implied by connecting the sentences with the conjunctive word. This is called a “conjunctive logical structure.”

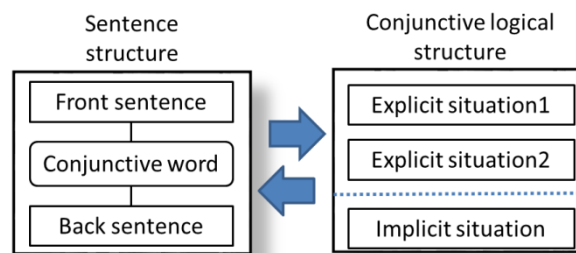


Figure 1. Mutual dependence between conjunctive words and the implicit situation model

3.2 Different-Meaning Commutativity Model

The different-meaning commutativity model is a principle that explains the essential content of the conjunctive word exercise in this study. Based on the diagram of Fig. 1, Fig. 2 shows a conjunctive expression with different-meaning commutativity. Its distinguishing characteristic is that two conjunctive expressions are given, and they respectively imply different meanings. The two expressions consist of two identical sentences with a different conjunctive word connecting them. The two sentences and the two conjunctive words are the components used to assemble these two conjunctive expressions, since the two conjunctive expressions are composed of the same two sentences and a different conjunctive word. These are called “sentence expression elements.” In Fig. 2, the meaning of each sentence is comprised of its explicit and implicit meanings. The explicit meaning is indicated explicitly in a sentence, and the implicit meaning is implied by the conjunctive word. Furthermore, the explicit meaning is represented by the same two sentences as the sentence expression elements, and the implicit meaning is represented by a different sentence. The meaning of the sentence is comprised of the two sentences that represent the explicit meaning and the sentence that represents the implicit meaning. Furthermore, the meaning of two-sentence structures in relation to different-meaning commutativity is inferred from two sentences that respectively represent the same explicit meaning and two sentences that represent a different implicit meaning for each conjunctive word. These are called “conjunctive logical expression elements.” Fig. 3 shows an example of two conjunctive expressions with different-meaning commutativity. This example is the same as the example in the introduction and is represented by this model expression. In this study, we designed a conjunctive word exercise by assembling a pair of sentence and conjunctive logical structures from the sentence expression and conjunctive logical expression elements. This design of the exercises is described in the next chapter.

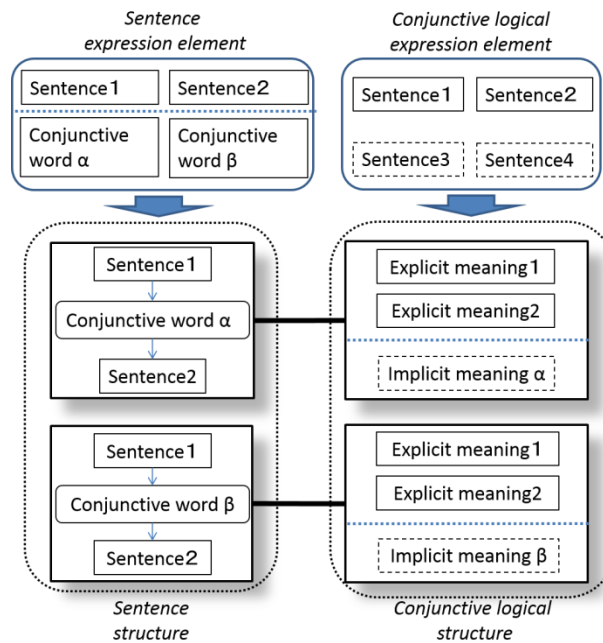


Figure 2. Definition of a different-meaning commutativity model

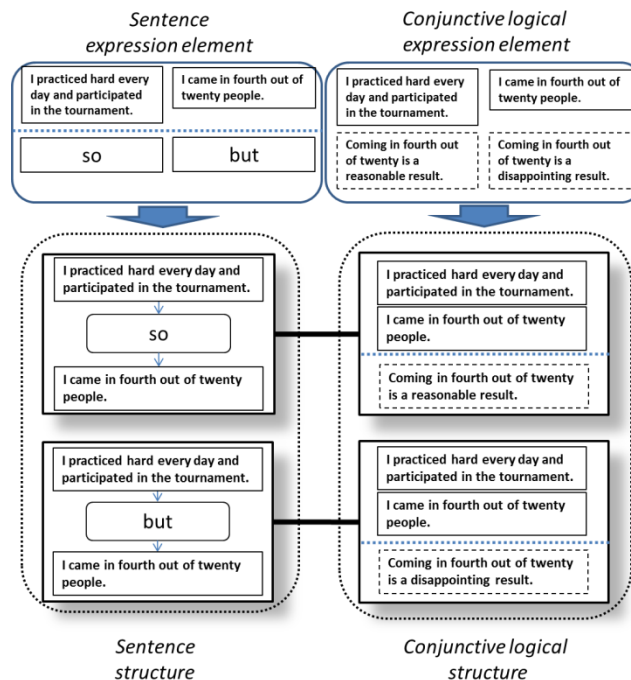


Figure 3. Example of a different-meaning commutativity model

4. Design and Development of a Conjunctive Word-Learning Support System

4.1 Learning Assignment Setting

In this study, this conjunctive word-learning support system includes providing the sentence expression elements and the conjunctive logical expression elements to learners. This system allows the learners to assemble the information structure of the conjunctive expression with different-meaning commutativity as a conjunctive word-learning assignment. By using the pairs in Figs. 2 and 3, a sentence expression

element list and a conjunctive logical expression element list, as shown in Fig. 4, can be prepared. By applying this system to the diagram of the sentence and conjunctive logical structures in Fig. 1, reconstructing Figs. 2 and 3 can be made into an assignment. In this type of assignment, although the relationship between the sentence and conjunctive logical structures is a many-to-many relationship, designing such an information structure and using it as an assembly activity can form the basis of a system that allows learners' level of mastery to be diagnosed automatically. This approach is called the "open information structure approach" (Hirashima & Hayashi, 2018). Furthermore, it has been confirmed that a system designed using this approach promotes the learning of information structures such as arithmetic word problems and concept maps (Hirashima, 2015, 2016). Therefore, we designed and developed a conjunctive word-learning support system based on this approach.

In order to diagnose students' level of mastery automatically with this system, we prepared sentences that included an explicit meaning (explicit sentences) and conjunctive words, and sentences that included an implicit meaning (implicit sentences) that could be indicated by their combination. By asking learners to combine the elements, this system provided them with assignments that helped them recognize conjunctive expressions that indicated different-meaning commutativity. In each assignment, two explicit sentences that could become the front and back sentences in the sentence structure were presented, as were two conjunctive words that could be inserted between them to construct sentences that made sense. Depending on the elements, if the conjunctive word was changed, the implicit sentence could change and vice versa. In this study, one of the goals of this conjunctive word exercise was to help learners understand different-meaning commutativity.

Furthermore, the variations in combinations increased by changing the type of target conjunctive word. The conjunctive words included in the assignment created using this system included resultative and adversative types to represent logical connections (Inoue & Hirashima, 2011). In the case of resultative and adversative conjunctive words, the following exercises, in which six sentences and two conjunctive words are presented, can be used to assess learners' mastery.

- (1) Prepare a pair of explicit sentences that can be used as the front and back sentences.
(Examples: "I wanted to see my friend" and "I went to the station")
- (2) Select two conjunctive words that can connect explicit sentences.
(Examples: "so" and "but")
- (3) Consider concrete examples of implicit sentences when the explicit sentences are connected by the conjunctive word and create a conjunctive logical structure.
(Examples: "My friend is at the station" and "My friend isn't at the station")
- (4) Prepare sentences that are the opposite of the front and back sentences.
(Examples: "I didn't want to see my friend" and "I didn't go to the station")

The elements created based on this procedure are both elements of the sentence and conjunctive logical structures. From the examples above, it can be seen that the learner's answer should correspond to a sentence and conjunctive logical structure that can be formed by the combining six sentences and two conjunctive words. In this procedure, multiple sentences can be composed. Fig. 5 shows the combinations that can be given as answers when the assignment is designed this way. The combinations of answers can be organized in a tree structure. The sentence structure includes 60 ways ($6 \text{ sentences} \times 2 \text{ conjunctive words} \times 5 \text{ sentences} = 60 \text{ ways}$), the conjunctive logical structure includes 120 ways ($6 \text{ sentences} \times 5 \text{ sentences} \times 4 \text{ sentences}$); thus, there are 7,200 total ways to answer (60×120). Therefore, the flexibility of the system is maintained because the number of potential combinations of answers is huge.

The types of exercises in this system are classified into (1) free-assembly exercises, (2) transition-assembly exercises, and (3) partial-assembly exercises. The free-assembly exercise requires learners to assemble the entire sentence and conjunctive logical structures. The transition-assembly exercise requires learners to change part of the sentence and conjunctive logical structures from a state in which all the components have been assembled in advance and then reassemble the other parts accordingly. In the partial-assembly exercise, some of the components of the sentence and conjunctive logical structures are decided in advance, and the remaining parts are assembled by learners. This exercise is divided into 5 components because there are multiple patterns of predetermined parts. Therefore, the parts assembled in each exercise are different.

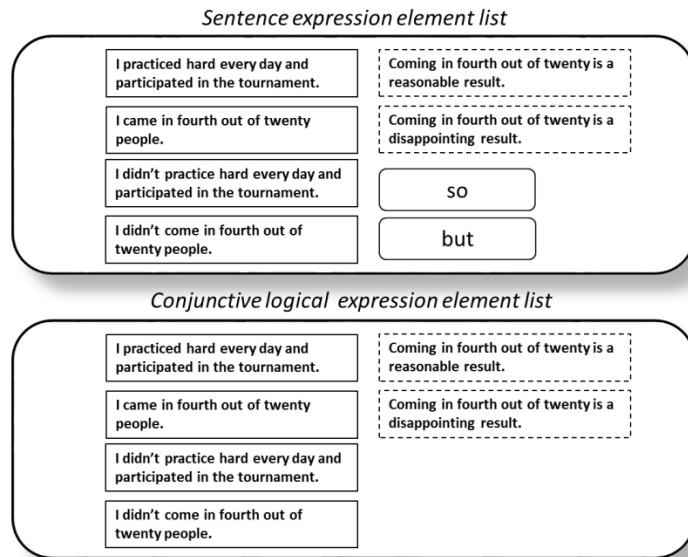


Figure 4. Sentence expression and conjunctive logical expression element lists

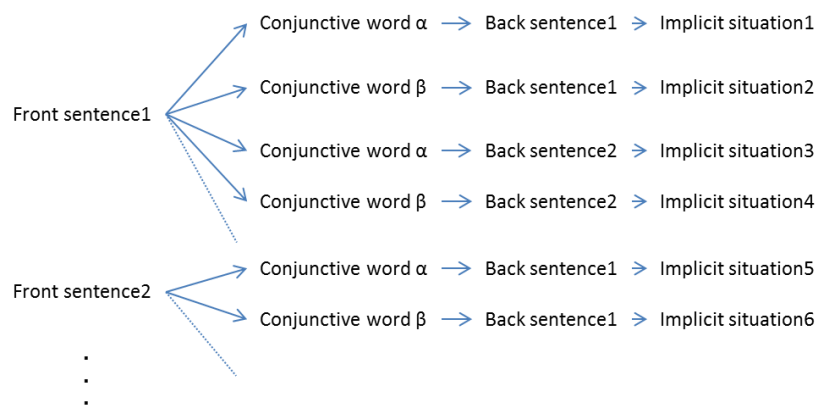


Figure 5. Instance of mutual dependence between a conjunctive word and an implicit situation

4.2 Immediate Feedback

Another important element in formulating conjunctive word exercises to address conjunctive expressions with different-meaning commutativity is immediate feedback. There are several types of feedback in learning (Hirashima, 2017), and we utilize true-false feedback. A correct pattern is a combination that does not cause a contradiction in the combination of the prepared sentence expression and conjunctive logical expression element lists. There are 48 correct combinations (6 front sentences × 2 conjunctive words × 4 back sentences = 48) among these combinations. Even though there are 5 back sentences, we count them as 4. The reason is that the combination of a positive and a negative sentence is excluded because it is an unnatural construction (example: “I wanted to see my friend, so I didn’t want to see my friend”). The correct pattern can be created from the 48 combinations above, excluding cases of unnatural sentences and those in which there is no implicit sentence in the expression element list. An unnatural sentence like “I want to see my friends, so they are at the station” would be excluded. In the assignment corresponding to this example, 32 patterns are correct. However, there are cases in which implicit statements are not uniquely determined. In such cases, the implicit sentence “cannot be decided” was judged as the correct answer (example: “My friend is at the station, but I didn’t go to the station”).

4.3 Interface of the Conjunctive Word-Learning Support System

An interface from this system is shown in Fig. 6 (translated from Japanese to English). The upper-left is a blank space where the sentence structure is assembled, the upper-right is a blank space where the conjunctive logical structure is assembled, and the lower half shows the options available to the learner. At this stage, the learner assembles the sentence and conjunctive logical structures at the top using the options at the bottom of the screen.

The sentence structure is constructed by moving the front sentence, the conjunctive word, and the back sentence from the lower-left card group to the upper-left blank. The conjunctive logical structure is constructed by moving explicit and implicit sentences from the bottom-right card group to the upper-right blank. After that, the result is shown by pressing the answer button.

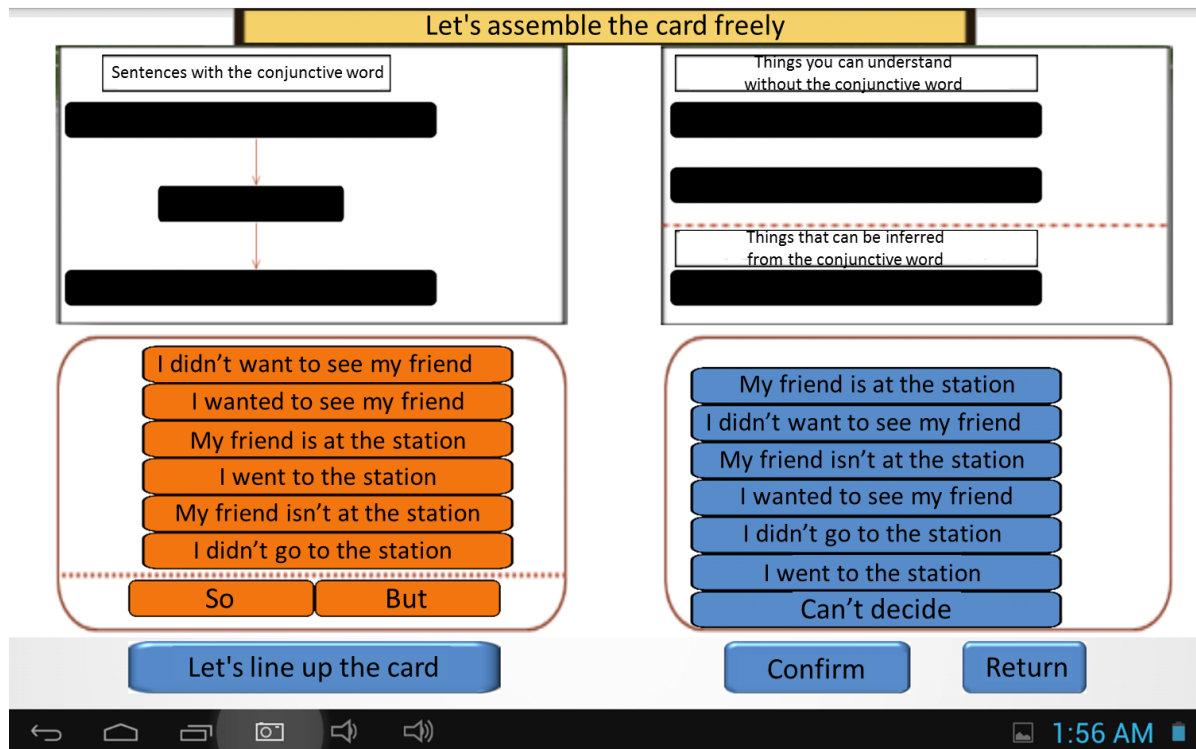


Figure 6. Interface of the conjunctive word-learning support system

5. Experimental Use of the Learning Support System

5.1 Procedure for Practical Use of the Learning Support System

The purpose of this experiment is to verify the following two items:

- 1) This system can be used by children in a primary school setting.
- 2) Use of this system positively affects primary school students' mastery of the use of conjunctive words.

The flow of the experiment is as follows:

- (1) Pre-test
- (2) Class: A teacher has test subjects recognize the implicit situation in a textbook with a description of different-meaning commutativity.
- (3) Description of system
- (4) Exercise 1 (16 minutes): Types 2 and 3 (5 questions for each assignment)
- (5) Exercise 2 (10 minutes): Type 1
- (6) Post-test and questionnaire

The test subjects were 39 fourth-grade students at an elementary school. Exercise1 and exercise2 were based on the words “friends” and “station.”

5.2 Log Data and Questionnaire

In order to verify that this system could be used by primary school students in practice, we analyzed the log data of this system and the questionnaire. The results of the questionnaire suggested that these types of assignments were difficult for some of the children because the proportion of negative responses to questions about their difficulty was high. On the other hand, the other questions yielded mostly positive opinions, and the exercises and system in this study were generally well-accepted. Therefore, the results suggest that the children understood the significance of the exercises and engaged in them with interest even though they recognized that they were difficult. Moreover, 32 out of 39 subjects filled out the comments section in the free-writing portion of the questionnaire, and many of their responses were “I want to use this system again” or “I enjoyed this system.”

An analysis of the results of the system log shows that the children answered an average of 1.95 times per minute during the first 5 minutes and 1.58 times per minute during the last 5 minutes of the 26-minute exercise. Therefore, we believe that students were continuously engaging in the learning activity because the system was used constantly from the beginning to the end. Moreover, we believe that the learners answered while thinking about the validity of the combinations of a conjunctive word and an implied situation within this framework. This is because the rate of correct answers given by the learners was higher than the rate of correct answers among all the combinations that the learners could choose in this exercise. Based on these observations, we believe that the children actively engaged with these exercises and gave them careful consideration when choosing their answers.

5.3 Pre-Test and Post-Test Comparison

In the pre-test and post-test that were conducted before and after the class, the same three questions were given in the form shown in Fig.7 (translated from Japanese to English). These tests were different from the assignment given using the system that we described, and they were given to investigate whether the children could find more combinations of conjunctive words and implied situations. During these tests, the test subjects were given adequate time to allow them to record all possible combinations.

Fig. 8 shows the results of the pre- and post-tests. The results are analyzed by dividing assignments into one category for the resultative and adversative and another for other conjunctive words. The recall rate (the number of correct answers in the learner’s responses divided by the number of all possible correct answers) and the precision rate (the number of correct answers given by the learner divided by the total of all learners’ responses) were determined for both assignments. The analysis results for the resultative and adversative conjunctive words show that the average recall rate increased from 64.7% to 73.7%. Furthermore, when an analysis of the variance was performed, a significant difference was confirmed ($p = 0.00661$), and the effect size was large ($d = 0.81$). The average of the precision rate decreased from 94.6% to 93.6%; thus, no significant difference was confirmed. On the other hand, the analysis results of the other conjunctive word assignments show that the average recall rate increased from 47.3% to 51.9%. Furthermore, when an analysis of the variance was performed, no significant difference was confirmed. The average precision rate decreased from 44.6% to 38.3%, indicating a significant difference ($p = 2.91E-05$).

The analysis of the assignment for the resultative and adversative conjunctive words suggests that more correct combinations were chosen in this exercise. The reason is that the recall rate significantly increased without changing the precision rate, although a few mistakes were still included. Therefore, this result suggests a meaningful learning effect on students’ mastery of conjunctive words of the resultative and adversative types. The analysis results of the other conjunctive word assignments did not suggest that there was a significant learning effect in the numerical sense. This is because the recall rate increased but was not significant, and the precision rate decreased significantly. However, the increase in the recall rate and the decrease in the precision rate suggest a tendency to use conjunctive words in various ways. Therefore, it is believed that this exercise affected the learners’ awareness and understanding of conjunctive words in general. At the same time, the results also suggest that it would be necessary to do similar exercises for each conjunctive word.

Pre-Test

student number _____

Please insert a conjunctive word in the parentheses and then connect the conjunctive word to the matching implied sentence with a line.

The puppy brought the ball, () the puppy was scolded by my mother.

.
 .

so

but

although

and

by the way

or

.
 .

My mother loves the puppy.

The puppy brought my mother's ball.

The ball the puppy brought was not my mother's.

The puppy always plays ball with my mother

.
 .

My mother didn't want the puppy to bring the ball.

Figure 7. Format of the pre-test and post-test

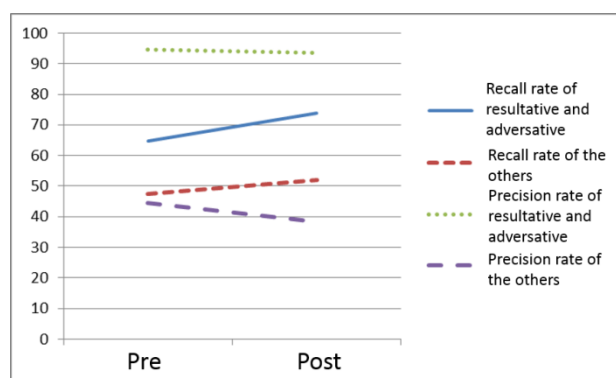


Figure 8. Results of the pre-test and post-test

6. Concluding Remarks

In this study, we discussed the significance of different-meaning commutativity and its modeling based on previous research. Based on the model, we proposed a conjunctive word exercise focused on different-meaning commutativity and developed a system for administering it. Furthermore, we performed an evaluative experiment to verify the validity of this system. From these results, one can see that the positive learning effect on students' mastery of different-meaning commutativity for conjunctive words of the resultative and adversative types was verified. In the future, we would like to create guidelines for assignments focusing on other types of conjunctive words, discuss more advanced methods of gathering feedback, and verify the usefulness of learning activities through a detailed analysis of learners' responses.

References

- Aidinlou, N. A., & Mehr, H. S. (2012). The effect of discourse markers instruction on EFL learners' writing. *World Journal of Education*, 2(2), 10-16.
- Baba, T. (2005). A research document on teaching and acquiring conjunctions in Japanese-language education and its outline, *Sapporo Japanese language study*, 11, 1-23 (In Japanese).
- Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, 4(5), 269-28.
- Hirashima, T., Hayashi, Y., Yamamoto, S., & Maeda, K. (2015). Bridging model between problem and solution representations in arithmetic/mathematics word problems, *Proc. of ICCE2015* (pp. 9-18).
- Hirashima, T., Hayashi, Y. (2016). Scaffolding of thinking about structure with kit-building tasks, *ICCE2016 Workshop Proceedings* (pp. 379-382).

- Hirashima, T. (2017). Externalization of computational thinking and process evidence: An open information structure approach. *Annual Conference of the Japan Society for Artificial Intelligence* (In Japanese).
- Hirashima, T., & Hayashi, Y. (2018) Design of meta-problems with an open information structure approach. *Workshop Proceedings of the 26th International Conference on Computers in Education*. Philippines: Asia-Pacific Society for Computers in Education.
- Ichikawa, T., (1978). *Introduction to national language education* (In Japanese).
- Ishiguro, K. (2008). *Sentences determined by conjunctions*, Koubunsha Shinsho (In Japanese).
- Inoue, H., & Hirashima, T. (2011). Visualization of the logical structures implied by conjunctions and the design of a learning support system based on them. *Proceedings of the Annual Conference of the Chugoku branch of Japanese Society for Information and Systems in Education*, 11 (1), 13-16, (In Japanese).
- Itou, S., & Abe., J. (1991). The function and necessity of conjunctions. *Psychological Research*, 59(4), 316-323 (In Japanese).
- Kawabata, M., (2009). A study of the function of Japanese connectives indicating logical direction. *Aichi Institute of Technology Research Report*, 44, 17-26 (In Japanese).
- Manan, N. A. A., & Raslee, N. N. (2018). Explicit discourse marker instruction to improve coherence and cohesion in academic writing. *International Journal of Academic Research in Business and Social Sciences*, 8(1), 465-483.
- Ogata, Y., Hayashi, Y., & Hirashima, T. (2015). Design and development of interactive assembly exercises between sentence expression with conjunctions and situation expressions with implication. *IEICE Technical Report*, 114(513), 163-168 (In Japanese).
- Ujiie, Y. (1973). "Conjunctions" as "relational expressions." *Bulletin of the Institute of Language Education of Waseda University*, 12, 1-16 (In Japanese).
- Yamamuro, K. (2008). *A study of syntactic content-handling in grammar education*, KEISUI Company (In Japanese).
- Yale Center, (2012). *Linking words (conjunctions and connectors)*, Graduate Writing Center, Yale Center for Teaching and Learning.

A Turkish EFL teacher's change processes through an OPD program: a case study

Behice Ceyda SONGÜ L, Ömer DELİALİOĞLU, Çağrı ÖZKÖSE BIYIK

^a*Dept of Foreign Languages and Cultures, Zonguldak Bülent Ecevit University, Turkey*

^b*Dept of Computer Education and Instructional Technology, Middle East Technical, University, Turkey*

^c*Dept of English Language and Literature, Yaşar University, Turkey*

*behiceceydacengiz@gmail.com

Abstract: In this paper, we present an analysis of the change processes of a Turkish EFL teacher who participated in an Online Professional Development (OPD) program that included online lesson study procedure and webinars. The aim of the study was to uncover the short term and long term changes of this teacher through the use of Clarke and Hollingsworth's (2002) Interconnected Model of Teacher Professional Growth (IMTPG). As part of the OPD program, this teacher attended six webinars about language teaching methodology and the use of technology in language teaching. In a period of 13 weeks, she also became part of online lesson study procedure with three other teachers. To this end, they met online once a week to do lesson planning for a research lesson and to later discuss about the effectiveness of that lesson by viewing its video recording. In this case study, the data collection tools comprised a background questionnaire, interviews, group meeting extracts, pre-and post-observations, lesson plans prepared by the group, video recordings of the research lessons and group meetings. The findings showed that various short term changes were initiated by different domains including personal domain, the lesson study discussions and webinars in the external domain with various pathways of change. More long-lasting changes, on the other hand, came about when she integrated her learning from the webinars into her classroom teaching and reflected on the student outcomes of these new practices, which promoted changes in her earlier beliefs and attitudes towards some aspects of teaching. It was also revealed that the changes in her cognition and teaching behavior were related to her increased knowledge and skills about incorporating web 2.0 tools into language classes.

Keywords: Online professional development, webinar, online lesson study, teacher change, Interconnected Model of Teacher Professional Growth

1. Introduction

Conventional approaches to PD which include one-time training efforts with few examples of effective teaching and few opportunities for teacher collaboration and reflection have been maligned for their inefficacy in leading to teacher change (Bickerstaff & Cormier, 2015). More constructivist forms of PD which are situated in teachers' daily teaching practice and view teachers as 'active and reflective practitioners' have been shown to contribute to teacher development (Wideman, 2010, p.4). With a common goal, collegial support and activities that focus on student learning, teachers are found to be better positioned to alter their teaching practices (Prenger, Poortman & Handelzalts, 2017). The observations from the extant body of research in PD also point out that teachers feel the need to 'experiment with new instructional strategies and materials' in socially bound professional learning communities (Schipper, Goei, de Vries & van Keen, 2017).

In the last decades, the online PD programs have been shown to meet many of these teachers' professional needs (Dede, Ketelhut, Whitehouse, Breit & McCloskey, 2009). When OPD appears in the form of online professional learning communities, it presents a wealth of advantages. First and foremost, the teachers have a chance to get connected with other teachers having similar problems (Lieberman & Pointer Mace, 2010), to engage in joint enterprise for improving their practice (Lantz-Andersson, Lundin, & Selwyn, 2018) and to perpetuate these activities in an ongoing manner by transcending time and place limitations (Powell & Bodur, 2019). They can also have access to various

resources including expert knowledge thereby meeting their learning needs as professionals (Blitz, 2013). In the current study, an OPD program was designed based on an online professional learning community model with a consideration of the key features of effective professional development. To this end, online lesson study procedure was implemented with a group of Turkish EFL teachers along with a series of webinars offered to these teachers (Yursa & Silverman, 2015). Wideman (2010) argues that despite the abundance of studies on OPD, very few studies had a direct assessment of the changes in teachers' cognition and classroom practices after their participation in an OPD program. To address this research gap, this study sets out to examine how an OPD program which integrates online lesson study with webinars contributes to a language teachers' development.

2. Literature Review

2.1 Lesson Study

Lesson study originated in Japan early in the 1900s (Cheung & Wong, 2015; Fernandez & Yoshida, 2004) and since then became widespread in many countries all over the world. (Clivaz & Ni Shuilleabhain, 2017) In LS, generally, small groups of teachers (usually four to six teachers) (Fernandez & Yoshida, 2004) work collaboratively to set goals for student learning and co-plan research lessons that target these goals. One of the teachers teach the lesson in his/her own classroom and other teachers act as observers and collect data on student learning. In post-lesson discussion, they reflect on the data to improve the lesson and if possible and teach the revised lesson in one or more classroom (Lewis 2002; 2009). LS started to be adopted worldwide as a viable means for supporting teachers to improve their professional knowledge and practices. For instance, the World Association of Lesson study founded in Hong Kong in 2005, has 'members from over 60 countries' now (Xu & Pedder, 2015). Some countries took initiatives to conduct lesson study nation-wide as a school improvement policy or fostering reform-minded teaching (e.g. Hadfield & Jopling, 2016). Extensive studies conducted in different parts of the world such as US (Fernandez, Cannon, & Chokshi, 2003), UK (e.g. Cajkler, Wood, Norton, & Pedder, 2014), Australia (e.g. Hollingsworth & Oliver, 2005), China (e.g. Yang, 2009) Indonesia and Malaysia (White & Lim, 2008) among many others (Doig & Groves, 2011) also reflect the growing interest for LS all over the world. In their review of studies on LS, Xu and Pedder (2015) report many studies which indicate improvements in teachers' subject content knowledge, pedagogic knowledge, pedagogic content knowledge and knowledge about pupils as a result of their learning from lesson study procedure (Xu and Pedder, 2015). Despite the abundance of research on LS, an online version of LS has received only scant attention in literature. Only two studies mentioned how they implemented online lesson study with a group of teachers (Sharma & Pang, 2015; Yursa & Silverman, 2011). However, as a limitation, only self-report instruments were used and teacher change was not within the scope of these studies.

2.2 Online Professional Learning Communities

Online community is defined as 'a group of people' who have 'common interests and shared goals' and utilize online technology for communication in lieu of 'face-to-face interactions' (Hew, 2009). The idea of online community, which bolsters the sociocultural approaches to learning (Mackey & Evans, 2011) arises as a relatively new professional development model (Barab, Kling & Gray, 2004). Learning communities are characterized by 'explicit learning goals' and are aimed at achieving certain pre-defined outcomes (Whitehouse, McCloskey, & Ketelhut, 2010).

Wideman (2010) indicates that although OPLCs are around at least for a decade, the number of studies that examine their effectiveness and sustainability are quite limited. Most of the research centres on a description of teachers' experiences through the use of self-reports (Blitz, 2013) whereas the impact of collaborative teamwork on teachers' processes of learning are quite unstudied. There are also very few studies that looked at changes in teachers' classroom practices as an outcome of their participation in OPLC or the impact of community activities on student learning (Wideman, 2010). This study, which focuses on teachers' development of new knowledge and skills through an OPD program also aims to fill in the gap in the literature related to the effect of OPLCs on teachers' knowledge development and pedagogical practices.

2.3 Interconnected Model of Teacher Professional Growth

In IMTPG, Clarke and Hollingsworth (2002) proposes four different domains where change can occur: the personal domain (PD), the domain of practice (DP), the domain of consequences (DC) and the external domain (ED) (See Figure 1). The personal domain consists of teachers' knowledge, beliefs and attitudes. When new knowledge, beliefs and attitudes are acquired, there is a change taking place in this domain (Voogt et al. 2011). The domain of practice refers to teachers' professional experimentation. This experimentation occurs when teacher try out new practices. The domain of consequences is related to the inferences teachers draw from their practices about themselves and their students as affected by their value system' (Witterholt, Goedhart, Suhre, & van Streun, 2012). In order for a change to happen in this domain, the consequences need to be salient to the teacher. For example, when a teacher observes improvement in student outcomes consequent to the use of a new teaching practice, this can be perceived as salient by the teacher and change this domain. Finally, the external domain refers to the 'outside of the teacher's personal world' (Witterholt et al. 2012) and is composed of any external source of information that yields change such as meeting with colleagues, attending professional seminars, etc. There are two mediating processes that enable a change between the four domains: enactment and reflection. Enactment is putting new ideas into action by trying out these new ideas or practices as professional experiments (Wang, Kim, Lee, & Kim, 2014). Reflection, on the other hand, refers to teachers' evaluation of their' students' learning outcomes' and their self-analysis of their 'teaching beliefs, attitudes and knowledge' as a lens for examining the effectiveness of the enacted pedagogical practices (Hung & Yeh, 2013, p.154). The processes of enactment and reflection have the primary function of enabling multiple pathways between the domains.

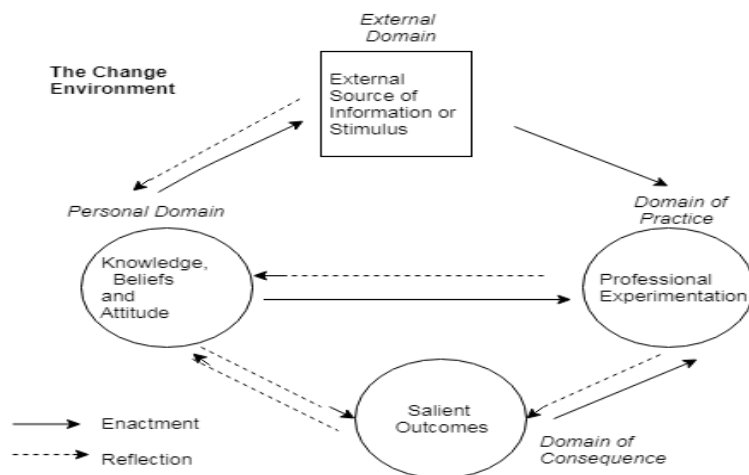


Figure 1. The interconnected model of teacher professional growth (Clarke & Hollingsworth, 2002)

According to Clarke and Hollingsworth (2002), these pathways can lead to two types of teacher change, which are either temporary changes or more lasting changes. Called as change sequence, the temporary changes represent the learning processes, the effect of which continues for a short time on teachers' mentality and practices. When the change in any of the three domains (domain of practice, personal domain, domain of consequence) is long-lasting, it is seen as an indicator of professional growth and the change sequence is turned into a growth network. Growth network is a more meaningful outcome of professional development initiative compared to short-lasting effect of a change sequence (Clarke & Hollingsworth, 2002).

3. Methodology

3.1 Research Design

As the research design, case study is used in the current study. Case study refers to “the in-depth study of instances of a phenomenon in its natural context and from the perspective of the participants involved in the phenomenon” (Gall, Gall & Borg, 2003, p. 436). It is one of the most frequently used qualitative research methods (Borg, 2003) which provides a detailed exploration of a case through the use of multiple data sources (e.g. interviews, observations, documents, etc.) (Creswell, 2007). In this study, the case consisted of a Turkish EFL teacher who participated in an OPD program. To identify the unique pathways of change experienced by this teacher, the following research questions guided the direction of this study:

1. What sequences of change, mediated by reflection or enactment processes are observed for the Turkish EFL teacher participating in the OPD program?
2. Which growth networks are identified for this teacher?

3.2 Case participant

As revealed by a background questionnaire administered at the beginning of the study. Sally had her undergraduate degree in Translation and Interpreting and held MA degree in Curriculum and Instruction. She had 15 years of teaching experience apart from working in Research and Development Division at Ministry of Education. Sally took part in some PD activities in Turkey and she travelled abroad for some Erasmus projects. At the beginning of the OPD program, she expressed her desire to improve her teaching skills since she was not a graduate of English Language Teaching Department and felt the need for professional development. She followed some teacher, project or graduate research groups in Facebook.

3.3 Overview of the Online Professional Development Program

This OPD program which lasted for 13 weeks between November 2017 and January 2018 consisted of two main activities for the participating teachers, which were the webinars and online lesson study procedure. As part of a Marie Curie Project that targets language teachers’ professional development in Europe, six webinars on different topics were prepared by the researcher and her academic advisor and piloted with the teachers in the online lesson study group. They focused not only on information delivery but also on some hands-on practice. The topics of the webinars included reflective practice for language teachers, innovative techniques in teaching English I, II, Using New Technologies in Language Teaching I, II and III. In these webinars, the focus was on language teaching methodologies and the web 2.0 tools that can be used in language classes. WizIQ was used as the webinar platform.

The online lesson study component of the OPD program, which was implemented once a week throughout the study was prepared with some adaptations made to Dudley’s (2015) version of lesson study. In online meetings held in WizIQ, the teachers exchanged ideas about lesson planning filling in the lesson study proposal and the research lesson planning, observation and discussion sheet collaboratively with the use of Google Docs. The focus of the research lessons was three case pupils and the teachers did the lesson planning with these three students in mind. The co-planned lessons were taught by one of the teachers and some small-data were collected from the students including interviews or student work samples. Later the video recordings of these lessons were shared with the other teachers with the use of Google Drive. Following the implementation of the research lesson, the teachers had a post lesson discussion in which they referred to their observations about the efficacy of the lesson for student learning. Based on the teachers’ joint contributions, the research lesson was revised and taught by another teacher in the group. After the second implementation, the group held another post-lesson discussion in order to evaluate the impact of the lesson on the students. This LS procedure was completed three times in the online lesson study group.

3.4 Data Collection Tools

Since it was a case study, various data collection tools were utilized in order to provide an in-depth understanding of the case (Creswell, 2007). Firstly, a background questionnaire was applied to the teacher before the OPD program to get some information about her professional background and about the activities she does for her professional development. She was also interviewed three times in total in that each interview took place after two research lessons in each LS procedure were completed. 4 classes of this teacher were also observed before and after her participation in the program. Finally, the recording of the group meetings and lesson plans prepared by the group were also included as a data collection tool.

3.5 Data Analysis

As the analytical framework, Clarke and Hollingsworth's (2002) model of Interconnected Model of Teacher Professional Growth was used for investigating the learning processes of the teacher attending the OPD program. However, some adaptations were made to this model due to the idiosyncratic features of the program. Firstly, the external domain, which refers to any external source of information provided to the teachers, was separated into two parts as 'webinars' and 'lesson study discussions' (LSD). The webinars served to represent the information presented to the teachers about language teaching methodology and web 2.0 tools for language classes. The lesson study discussions, on the other hand, referred to any peer-to-peer learning arising in group meetings when the teachers exchanged ideas with each other.

Secondly, some elaborations were made on the Domain of Practice to include the practice-oriented activities in lesson study procedure. Hence, Domain of Practice was divided into 'lesson planning', 'teaching', 'revising' and 'reteaching' which are the components of LS procedure and 'common teaching practice' outside these LS components. The common teaching practice was included in this domain in order to display professional experimentations in teachers' own classes as distinct from the LS procedure.

Data analysis process was initiated with the transcription of all data. Through constant comparison method (Merriam, 1998), all data were coded with the use of the coding scheme adapted from Zwart, Wubbels, Bergen and Bolhuis's (2007) study in order to uncover the changes in the teacher's cognition, beliefs and her common teaching practices after her participation in the OPD program. The frequency of the identified change sequences were provided for the teacher. Later, the pathways of change among the domains were identified with the use of Justi and Van Driel's (2006) criteria. The distinction between a change sequence and a growth network was made according to Clarke and Hollingsworth's (2002) descriptions in which growth networks are 'explicit evidence of lasting change in practice or in teacher knowledge or beliefs' (p. 958) unlike change sequences representing short-term changes. For inter-coder reliability as a means of ensuring trustworthiness, all data were also coded and analyzed by a colleague of the researcher. Only small differences were detected in two sets of analysis and an agreement was reached for these differences. Member checking was also done in that the researcher sent her analysis to the teacher and made some changes in her analysis after getting some feedback from the teacher.

4. Findings

The results pointed at many instances of short term learning on part of Sally which were represented by various change sequences initiated by different domains (See Table 1). It was found that most of the change sequences identified in Sally's data were initiated by the webinars in External Domain and Domain of Practice. It was seen that the webinars in External Domain led to change sequences especially when the ideas presented in the webinars were integrated into the research lessons. In addition, the Domain of Practice was also an efficient information source for the teachers since it provided the teachers with the opportunity to observe the effect of the research lessons and each other's teaching practices and techniques. The effect of the lesson study discussions in External Domain and Personal Domain were less paramount for the teacher since teacher contributions to the lesson study

discussions were not even with some teachers having few contributions to the discussions for lesson planning or revising.

Table 1

The frequency of the entry points of change

Domains	Frequency
Personal Domain	3
External Domain (webinars)	6
External Domain (lesson study discussions)	4
Domain of Practice	10

4.1 Change sequences initiated by Personal Domain

The change sequences initiated by Personal Domain were represented in two ways with only a slight difference in the pathway of change. In the first one, Sally offered a particular activity to the group as an active teacher in group meetings (PD→ED) and it continued with the acceptance of that idea by the other teachers for the research lesson (ED→ DP (lesson planning)). Following the teaching of the co-planned lesson, she reflected on the general quality of the lesson (DP (teaching) →PD). In the other change sequence, the ending point was Domain of Consequence in lieu of Personal Domain as the only difference. To illustrate the first one, in group meeting one, Sally suggested using Kahoot in the first research lesson: “We can use Kahoot at the end of the lesson. I use it for revising vocabulary” (Source: Group meeting 2). Kahoot is integrated into the lesson plan and in group meeting three, she evaluates the lesson as follows:

I have seen that Kahoot might be troublesome in some classes. The internet problems in the class culminated in utter chaos. Not sure it is good to use all times. It could have been better to think of a b plan when using technology in case of any technical problems. (Source: Group meeting 2)

4.2 Change sequences initiated by External Domain

The findings indicated that there were three different change sequences initiated by lesson study discussions or webinars in the External Domain. The first change sequence represented a reflection process on the learning that comes about as a result of the teacher’s participation in the lesson study meetings (ED (LSD) → PD) or in the webinars (ED (webinars) →PD). In the following quotes, for example, Sally indicated that holding lesson study discussions with colleagues was of great benefit for her:

I believe these kinds of meetings are very beneficial. Even if you don’t learn something new in every meeting -I believe I have learnt many new things though- it definitely has many contributions for you. At least, it helps you to think about your practices and what you can do about it. It is a thought- provoking process. (Source: Interview 1)

In a similar vein, the quote provided below illustrated Sally’s reflection process on the content covered in the webinars. She stated that she had increased knowledge about using web 2.0 tools in language classes as a consequence of her learning from the webinars:

In the webinars, I have learnt many Web 2.0 tools. Most of the tools covered in the webinars were new to me. In each webinar, I found many things to do research about. I started to search for other technologies. I thought more about what I can do with these technologies. (Source: Interview 2)

Another change sequence was initiated when the group members suggested some activities and some of these ideas were put into practice in research lessons (ED (LSD) → DP (lesson planning)). Following the teaching of the research lesson, the teachers reflected on how this lesson turned to be effective in terms of student learning outcomes (DP (teaching)-DC). In group meeting one, for example, Lucy proposed a warm up activity to the group as seen below:

For the first lesson, we can use a trailer to attract the students' attention as a warm up activity. We can ask the students some questions after watching the trailer to get them to guess the topic of the reading passage. (Source: Group meeting 1)

Upon Lucy's suggestion, the group decided to find a trailer to use before the reading passage about Agatha Christie. In the group meeting in which the teachers evaluated the class implementation, Sally reflected on that lesson as below:

The idea of using a trailer as a warm-up activity worked well in the classroom. I did not try it before. The students' interest in the reading passage increased since they were wondering about what to read. Doing some speaking about the trailer was a good introduction to the lesson for the students, I think. (Source: Group meeting 3)

Finally, Sally often transferred the know-how she gained in the webinars to her classes in an immediate fashion as different from the other teachers (ED (webinar) → DP (common practice)). The integration of new practices was followed by her reflection on the student outcomes of these lessons (DP (common practice) →DC). For example, in the fifth group meeting, in which teachers were discussing about the activities to be used in the forthcoming lesson, Sally talked about how she employed the treasure hunt activity she learnt in the webinars in her own class. In her reflection, she also explained the increased student interest in the lesson as given below:

I used QR code and treasure hunt game in one of my classes. I was teaching present tense. I prepared five questions and qr codes for each of them. I posted the qr codes on different parts of the school. I chose a leader from each group. There were seven groups in total. Some instructions on how many questions they needed to answer appeared when they scanned the qr code. They answered the questions as a whole group. The activity was very nice, but chaotic, at the same time. The students had so much fun but when it is a crowded class, it is not quite possible to get rid of the noise. (Source: Group meeting 5)

4.3 Change sequences initiated by Domain of Practice

There were two different change sequences initiated by Domain of Practice. The first of these included a reflection process on the student outcomes of class implementations in research lessons (DP (teaching) →DC). This can be seen in the following quote from a group meeting:

I really liked to see that doing individual listening instead of whole class listening can work well in the classroom. In the lesson, the students were able to concentrate better and did not get distracted as they did the other times. (Source: Group meeting 10)

The other change sequence included a reflection process on the student learning or motivation observed in the research lessons (DP (teaching) →DC) which was followed by a change in teachers' previous cognitions or beliefs about teaching or learning (DC→ PD). This change sequence is apparent in the following group meeting extract. Here Sally evaluated the consequences of the research lesson in terms of student outcomes by comparing it to a more traditional lesson. Later, her realization about the favorable outcomes in the lesson led her to make new decisions about trying out a different approach to teaching listening in her own classes:

I really liked to see that doing individual listening instead of whole class listening can work well in the classroom. In the lesson, the students were able to concentrate better and did not get

distracted as they did the other times. It is a pretty good practice we can use in our lessons. I plan to use it sometimes. (Source: Group meeting 10)

4.4 Growth Networks

The results of the study demonstrated that there was one particular pathway of change identified for Sally which represented long-term changes in her cognition and teaching practices. This growth network was initiated when Sally incorporated the newly learnt web 2.0 tools introduced in the webinars into her common teaching practice. Consequent to her implementation of that new teaching practice, she reflected on the favorable student outcomes, which resulted in a change of her prior cognition about teaching.

To illustrate, in the third interview, Sally talked about how she integrated Padlet in the fifth webinar into her own teaching. In the post-observation data, she was also seen to use Padlet for a writing activity in the classroom (ED (webinar) → DP (common practice)). Evaluating the lesson, she first referred to the effects of using Padlet on students' writing performance (DP (common practice) → DC). Second, she explained her intention to use it in further lessons (DC → PD).

In class, I sent my students a padlet link and they did writing activity about the topic we covered that day. They really liked using Padlet. They enjoyed looking at their own writings and those of their friends. This time, they were more careful about grammar and they wrote more enthusiastically than they did other times. I believe when Padlet is used, they give more attention to the writing activity and the product is better compared to the times when classical methods are used. I will keep using it. (Source: Interview 2)

As another finding, the analysis of the interviews and the group meetings also demonstrated that the changes she made in her classroom practice during the online PD program were centered not on language teaching methods or techniques but on the integration of web 2.0 tools. This finding was reaffirmed in the pre- and post- observations which showed that she maintained her common teaching practices consequent to her participation in PD. She only started to integrate more technological tools as the only difference between pre- and post-observations.

5. Discussion and Conclusion

The findings of the study demonstrated that different domains with the joint effect of the mediating processes of enactment and reflection had an influence on the participating teachers' development in the OPD program. The contextualized research lessons in which practice-focus and collegial sharing were paramount were found to contribute to teacher learning (Xu & Pedder, 2015) affirming the value of online lesson study for teacher professional development (Sharma & Pang, 2015; Yursa & Silverman, 2011). Additionally, webinars focused on web 2.00 tools served to increase the teacher's technological know-how and motivated her to use these tools not only in research lessons but also in her own classes.

Despite the presence of change sequences with different entry points, all of the long-lasting changes in the teacher's cognition and teaching behavior were found to be initiated by the webinars in the External Domain. The teacher integrated the newly gained information from the webinars into her regular practice. This was followed by her reflection on the student outcomes of the lesson, which resulted in a change in her earlier cognition and beliefs about teaching. As an important finding, these more permanent changes were found to be related to the teacher's increased knowledge and skills about incorporating web 2.00 tools into language classes. As an active group member, she was a teacher who already used some tools in her classes, which can account for her success in integrating new practices into her teaching. This finding could also be due to her positive attitude towards technology use in language classes (vanOlphen, 2007) since not all teachers are able to implement the teaching techniques introduced in the webinars in their own classes at satisfying degrees of success (Mai & Ocricano, 2017).

The absence of any growth network initiated by lesson study discussions in the External Domain showed that collegial learning was not intense enough to bring about long-lasting changes in this teacher. It is possible that the lesson study meetings were not enough to create cognitive conflicts

for this teacher which are considered necessary for teachers to change their firm beliefs about teaching and student learning (Clarke & Hollingsworth, 2002). It is also likely that the online lesson study group was not able to create a collaborative dialogue among the teachers as a typical problem in professional development programs (Akiba, Murata, Howard & Wilkinson, 2019). It is anticipated that the low level of social engagement and real collaboration within the group can be due to the lack of a facilitator who can interfere in group interaction and sustain a ‘regime of participation’ (Henderson, 2007). As an implication of the study for future OPD efforts, therefore, a facilitator can be included in the online lesson study group as a moderator of teacher discussions. Moreover, webinars can be provided to the teachers (Blitz, 2013) in OPD programs along with the opportunities to test out the newly gained information in teachers’ classroom practices and to reflect on the student outcomes of these implementations. Additionally, since reflection on student outcomes arose as an important means to teacher learning, it can be asserted that further PD efforts should focus on developing teachers’ reflection skills and help them gain better understandings of their students’ thinking. Finally, due to some limitations of this study related to the short duration and few number of cases, further research can examine the development of more teachers in a longer period of time.

References

- Akiba, M., Murata, A., Howard, C.C. & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, 77, 352-365.
- Barab, S. A., Kling, R., & Gray, J. H. (2004). *Designing for virtual communities in the service of learning*. Cambridge, NY: Cambridge University Press.
- Bickerstaff, S., & Cormier, M. S. (2014). Examining faculty questions to facilitate instructional improvement in higher education. *Studies in Educational Evaluation*, 46, 74-80.
- Blitz, C. L. (2013). *Can online learning communities achieve the goals of traditional professional learning communities? What the literature says* (REL 2013-003). Washington, DC: US Department of Education, Institute of Education Sciences.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching* 36(2), 81-109.
- Cajkler, W., Wood, P., Norton, J., & Pedder, D. (2014) Lesson study as a vehicle for collaborative teacher learning in a secondary school. *Professional Development in Education*, 40(4), 511-529.
- Cheung, W.M. & Wong, W.Y. (2015). Does lesson study work? A systematic review on the effects of lesson study and learning study on teachers and students. *International Journal for Lesson and Learning Studies*, 3(2), 137-149.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947-967.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches (2nd Edition.)*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Clivaz, S. & Ni Shuilleabhain, A. (2017). Analysing mathematics teacher learning in lesson study-a proposed theoretical framework, *Congress of European Research in Mathematics Education - CERME 10*, Dublin City University, Ireland, February 2017.
- Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, 60(1), 8-19.
- Doig, B., & Groves, S. (2011). Japanese Lesson Study: Teacher professional development through communities of inquiry. *Mathematics Teacher Education and Development*, 13(1), 77-93.
- Fernandez, C., Cannon, J., & Chokshi, S. (2003). A U.S.-Japan lesson study collaboration reveals critical lenses for examining practice. *Teaching and Teacher Education*, 19(2), 171-185.
- Fernandez, C. & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Gall, M.D., Gall, J.P., & Borg, W.T. (2003). *Educational research (7th Edition)*. White Plains, NY: Pearson Education.
- Hadfield, M. & Jopling, M. (2016). Problematizing lesson study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, 60, 203-214.
- Henderson, M. (2007) Sustaining online teacher professional development through community design. *Campus-Wide Information Systems*, 24 (3), 162-173.
- Hew, K. F. (2009). Determinants of success for online communities: An analysis of three communities in terms of members’ perceived professional development. *Behaviour & Information Technology*, 28(5), 433-445.
- Hollingsworth, H., & Oliver, D. (2005). Lesson study: A professional learning model that actually makes a difference. In J. Mousley, L. Bragg, & C. Campbell (Eds.), *Mathematics: Celebrating achievement*.

- (Proceedings of the 42nd annual conference of the Mathematical Association of Victoria, Melbourne, pp. 168-176). Melbourne: MAV.
- Hung, H., & Yeh, H. (2013). Forming a change environment to encourage professional development through a teacher study group. *Teaching and Teacher Education*, 36, 153-165.
- Justi, R. & Van Driel, J. (2006). The use of the interconnected model of teacher professional growth for understanding the development of science teachers' knowledge on models and Modelling. *Teaching and Teacher Education*, 22(4), 437-450.
- Lantz-Andersson, A., Lundin, M. & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and Teacher Education*, 75, 302- 315.
- Lewis, C. (2002). Does lesson study have a future in the United States? *Nagoya Journal of Education and Human Development*, 1, 1-23.
- Lewis, C. (2009). What is the nature of knowledge development in lesson study? *Educational Action Research*, 17 (1), 95 – 110.
- Lieberman, A. & Pointer Mace, D. (2008). Teacher learning: the key to education reform. *Journal of Teacher Education*, 59(3), 226-234.
- Mackey, J., & Evans, T. (2011). Interconnecting networks of practice for professional learning. *International Review of Research in Open and Distance Learning*, 12(3), 1-18.
- Mai, T., & Ocriciano, M. (2017). Investigating the influence of webinar participation on professional development of English language teachers in rural Vietnam. *Language Education in Asia*, 8(1), 48-66.
- Merriam, S.B. (1998). *Qualitative Research and Case Study Applications in Education*. San Francisco: Jossey-Bass Publishers.
- Powell, C.G., & Bodur, Y. (2019). Teachers' perceptions of an online professional development experience: Implications for a design and implementation framework. *Teaching and Teacher Education*, 77, 19-30.
- Prenger, R., Poortman, C.L., & Handelzalts, A. (2017). *Teaching and Teacher Education*, 68, 77-90.
- Sharma, S.A. & Pang, S. (2015). Creating new opportunities for lesson study in an online clinic. *Literacy Research: Theory, Method and Practice*, 64, 415-428.
- Schipper, T., Goei, S.L., de Vries, S., & van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of Lesson Study. *Teaching and Teacher Education*, 68, 289-303.
- Xu, H., & Pedder, D. (2015) Lesson Study: an international review of the research. In P. Dudley (Eds.), *Lesson Study: Professional Learning for our time* (pp. 24-47). London: Routledge.
- Van Olphen, M. (2007). Digital portfolios: Balancing the academic and professional needs of world language teacher candidates. In M. Kassen, R. Lavine, K. Murphy-Judy, & M. Peters (Eds.), *Preparing and developing technology-proficient L2 teachers* (pp. 265-294). San Marcos, TX: CALICO.
- Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., Mckenney, S., Pieters, J., & De Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education*, 27(8), 1235-1244.
- Wang, X., Kim, B., Lee, J.W.Y. & Kim, M.S. (2014). Encouraging and being encouraged: Development of an epistemic community and teacher professional growth in a Singapore classroom. *Teaching and teacher education*, 44, 12-24.
- White, A.L. & Lim, C.S. (2008). Lesson Study in Asia Pacific classrooms: Local responses to a global movement. *ZDM International Journal of Mathematics Education*, 40, 915 – 25.
- Whitehouse, McCloskey & Ketelhut, D.J. (2010). Online pedagogy design and development: New models for 21st century online In J. O. Lindberg, & A. D. Olofsson, (Eds.), *Online Learning Communities and Teacher professional development: Methods for improved education delivery* (pp.247-262). New York: Information Science Reference.
- Wideman, H. (2010). Online teacher learning communities: A literature review. Institute for Research on Learning Technologies, Technical Report: York University. Retrieved January 15, 2017 from <http://irlt.yorku.ca/reports/TechReport2010-2.pdf>.
- Witterholt, M., Goedhart, M., Suhre, C., & van Streun, A. (2012). The Interconnected Model of Professional Growth as a means to assess the development of a mathematics teacher. *Teaching and Teacher Education*, 28(5), 661-674.
- Yang, Y. (2009). How a Chinese teacher improved classroom teaching in Teaching Research Group: a case study on Pythagoras theorem teaching in Shanghai. *ZDM International Journal on Mathematics Education*, 41, 279 – 96.
- Yursa, H., & Silverman, J. (2011). Developing online lesson study community. Paper presented at the NCTM regional conference, Atlantic City, NJ. Retrieved May 15, 2017 from http://mathforum.org/workshops/ac2011/LS_AC2012.pdf
- Zwart, R.C, T. Wubbels, T. C., Bergen, M., and Bolhuis, S. (2007). Experienced teacher learning within the context of reciprocal peer coaching. *Teachers and Teaching: theory and practice*, 13(2), 165-187.

Validating an Instrument for EFL Learners' English New Media Literacy and the Relation to English Language Self-efficacy

Lin LUAN^{tab*}, Yan DONG^a, Shasha WANG^c, Yin LI^c

^a*Department of Foreign Languages, Beijing University of Posts and Telecommunications, China*

^b*School of Educational Technology, Beijing Normal University, China*

^c*QC Education Innovation Institute., China*

*luanlin32@126.com

Abstract: An increasing amount of research has been conducted on the integration of new media in English language learning and teaching, while limited studies have explored English as a foreign language (EFL) learners' English new media literacy. This quantitative study developed an instrument with two questionnaires for assessing EFL learners' new media literacy and English language self-efficacy. It further investigated the associations between the two constructs. The participants were 486 EFL learners at a comprehensive university in China, who took a compulsory language course for improving their overall English proficiency. The results validated the two questionnaires and further confirmed the correlations between the two constructs. It was very striking to see that both functional consuming and functional prosuming literacy played the most significant and positive role for predicting all the four aspects of learners' English self-efficacy. Moreover, critical consuming was found to act as a significant predictor of students' self-efficacy in listening and reading. This study sheds light on the role of new media literacy on EFL learners' self-efficacy, and also provides EFL instructors and policy-makers with pedagogical implications on the effective integration of new media literacy education into EFL teaching and learning.

Keywords: English new media literacy; English language self-efficacy; functional; critical; consuming; prosuming

1. Introduction

Along with the emergence of new media technologies, English media is serving as a unique and significant approach for Chinese EFL learners to improve their English proficiency (Guo & Yu, 2002). Although research on second language acquisition (SLA) have shown a growing research interest in applying new media tools in language teaching and learning (e.g. Rachels & Rockinson-Szapkiw, 2018), little attention has been paid to English new media literacy of EFL learners, particularly in the Chinese context. Hence, this study, situated in mainland China, aims to reveal the under-investigated intricate relationship between EFL learners' English new media literacy and their language self-efficacy, and further explore the predictive role of new media literacy on EFL learners' self-efficacy.

2. Literature Review

2.1 New media literacy

Considering both the technical and socio-cultural characteristics of new media, a preliminary framework of NML was proposed by Chen, Wu, and Wang (2011), consisting of two continua named as consuming and prosuming (producing and consuming). Based on the above two continua, four types of NML can be recognized. They are (a) functional consuming, (b) functional presuming, (c) critical

consuming, and (d) critical prosuming. The functional consuming literacy refers to the ability to gain access to created new media and understand what is being conveyed while functional prosuming literacy focuses on the ability to participate in the creation of media content. Meanwhile, the critical consuming literacy is concerned with media consumers' ability to study the social economic, political and cultural contexts of the media content, while critical prosuming literacy refers to individuals' contextual interpretation of the media content during their media participation activities. This framework was further refined with ten indicators by Lin et al. (2013) based on various discourses of media literacy and new media studies. In the subsequent study of Lee, Chen, Li and Lin (2015), an NML instrument was developed to provide a more comprehensive understanding of new media literacy.

2.2 English language self-efficacy

In the field of SLA, English language self-efficacy is considered as "one's belief about how well he/she can successfully carry out a task in English based on his/her past experience" (Wang, Kim, Bai and Hu, 2014). According to Bai, Hu and Gu (2014), language learner' self-efficacy is regarded as an important motivational variable influencing their' behaviors and learning performance. Because self-efficacy may vary depending on different contexts or domains (Bandura, 1997), it is suggested to be approached by using multifaceted instruments instead of general self-efficacy items to measure language learners' self-efficacy (Wang, Schwab, Fenn, & Chang, 2013).

2.3 New media in EFL teaching and learning

With the continuous development and prevalence of new media technologies, an increasing attention has been paid to the potential of integrating new media in a wide range of practices in EFL context. (e.g., Laire, Casteleyn, & Mottart, 2012; Yang, 2013). Considering both the affordances and constraints of new media (Zheng, Yim, & Warschauer, 2018), there was still controversy regarding the effect of new media on EFL learners' self-efficacy. For example, several researchers claimed that the use of new media technologies could enhance students' self-confidence in EFL learning process (e.g. Rachels & Rockinson-Szapkiw, 2018; Sun & Yang, 2015). Conversely, other researchers (e.g. Wu & Hsu, 2011) have cautioned that students' motivation of learning might decrease since studying in an open environment of social media will make some students feel anxious. Therefore, the present study aims at filling in some of these gaps by exploring the relationship between Chinese EFL learners' English new media literacy and their English language self-efficacy. The following two research questions will be addressed:

1. What is the relationship between EFL learners' English new media literacy and their self-efficacy?
2. What predictive role does EFL learners' English new media literacy have on their self-efficacy?

3. Method

3.1 Research context and participants

This investigation was undertaken in the context of an EFL language course at a public university in China. A random sample of 486 sophomores (around 20 years old) were involved in this study. Since most of the participants were majoring in telecommunications, computer science and electronic information science, male students (322 males) outnumbered females in our research. All the participants had received formal EFL education for more than twelve years.

3.2 Instruments

Two questionnaires were employed to evaluate EFL learners' English new media literacy and English language self-efficacy. Both of the questionnaires were measured on a five-point Likert scale from 1 'Strongly Disagree' to 5 'Strongly Agree'. The items in the instrument were translated into

Chinese since that was the first language of all participants. The following paragraphs give a brief account of the two questionnaires.

3.2.1 Questionnaire evaluating EFL learners' English new media literacy

Based on the previous theoretical framework of NML (Chen et al., 2011; Lin et al., 2013; Lee et al., 2015), we maintained four subscales to explore Chinese EFL learners' new media literacy. The questionnaire consists of 39 items under four subscales with eleven indicators: (1) functional consuming (consuming skill and understanding); (2) functional prosuming (prosuming skill, distribution and production); (3) critical consuming (analysis, synthesis and evaluation); and (4) critical prosuming (integration, feedback and participation). The sample questions of the four subscales are as follows:

(1) Functional consuming: I use search engine such as Google, Baidu or Yahoo to find English information.

(2) Functional prosuming: I can set up a social network account or blog account to send English messages to foreign friends.

(3) Critical consuming: When I read a Media Message on English websites, I find it is selling some values.

(4) Critical prosuming: I interact with other people in English in real time chat room, video conference such as WeChat, MSN or Skype.

3.2.2 Questionnaire assessing students' English language self-efficacy

The other instrument, the English language self-efficacy (ELSE) questionnaire, was from Su, Zheng, Liang and Tsai (2018)'s survey for measuring EFL students' self-efficacy in listening, speaking, reading, and writing skills. The questionnaire includes 25 items concerning students' beliefs about how capable they are to complete specific tasks in English. The four factors in the ELSE are described below, with one sample item each provided:

(1) Listening efficacy: Can you understand English lectures of general topics?

(2) Speaking efficacy: Can you ask people for help in English?

(3) Reading efficacy: Can you read English newspapers?

(4) Writing efficacy: Can you use accurate grammar when you write English essays?

3.3 Data collection and analysis

A total of 498 participants responded to the questionnaires anonymously in classroom setting at the end of the academic year. It took about 15 minutes for all the respondents to complete the survey. Finally, 486 students' answers remained valid and were further analyzed to address the research questions. The procedure of data analysis includes three phases. First, the exploratory factor analysis (EFA) was conducted to examine the component structure of the ENML and OSEL questionnaires. The reliability coefficient was also collected to ensure the internal consistency of the measurements as a whole. Second, the correlation analysis was then employed to analyze the relationship between the two finalized questionnaire factors. Third, stepwise regression analyses were conducted, in which the students' English new media literacy was considered as the predictors, while their English language self-efficacy was outcome variables.

4. Results

4.1 Factor analysis of the ENML questionnaire

To clarify the structure and validity of the ENML survey used in this study, an exploratory factor analysis was performed by using the principal component analysis with varimax rotation method. As suggested by Stevens (1996), items with the loadings weighed higher than 0.40 on the relevant factor and lower than 0.40 on all the other factors were maintained in the finalized ENML. Finally, 37 items were retained and grouped into four subscales. It also indicated that students showed strongest

agreement on functional consuming (Mean = 3.65, SD = 0.66), followed by functional prosuming (Mean = 3.50, SD = 0.72), critical consuming (Mean = 3.37, SD = 0.75), and critical prosuming (Mean = 3.31, SD = 0.67). The reliability coefficient (Cronbach's alpha) for each scale in this survey ranged from 0.63 to 0.83, with an overall reliability of 0.91. The results show the satisfactory reliability for measuring EFL learners' English new media literacy.

4.2 Factor analysis of the ELSE questionnaire

With regard to the ELSE instrument, we applied a similar process of exploratory factor analysis to identify its structure. Four factors with a total of 25 items were displayed in the finalized ELSE, explaining 68.48% of the total variance. Specifically, the students scored highest on writing (Mean = 3.77, SD = 0.67), followed by speaking (Mean = 3.67, SD = 0.71), reading (Mean = 3.53, SD = 0.73), and listening (Mean = 3.20, SD = 0.76). The reliability coefficients ranged from 0.61 to 0.81, with an overall alpha of 0.88, suggesting that the internal consistency was sufficient and the four scales were adequately reliable to assess students' English language self-efficacy.

4.3 Correlations between the factors of ENML and ELSE

To examine the hypothesis that there is link between the EFL learners' English new media literacy and their self-efficacy in the process of learning English, Pearson correlation analysis was conducted based on participants' responses to the two questionnaires. Statistically significant positive correlations were found between the scales in ENML and those in ELSE ($r = 0.37-0.60$, $p < 0.001$).

4.4 Stepwise regression analysis of predicting students' ELSE based on ENML

In order to fulfill our second research purpose, stepwise regression analyses was conducted to examine the relations between learners' English new media literacy and their English language self-efficacy. For the analyses, the EFL learners' ENML scales were processed as the predictors, while the ELSE factors were the outcome variables. First of all, we can clearly see that the 'functional consuming' played the most powerful role for predicting all the factors in the learners' English language self-efficacy. Moreover, functional consuming significantly and positively predicted the listening ($\beta = 0.21$, $t = 3.61$, $p < 0.001$), speaking ($\beta = 0.27$, $t = 4.78$, $p < 0.001$), reading ($\beta = 0.20$, $t = 3.43$, $p < 0.01$) and writing ($\beta = 0.29$, $t = 4.94$, $p < 0.001$) scales. Similarly, it was found that the score disclosed by the factor of functional prosuming in ENML has positive predictions for the factors of listening ($\beta = 0.33$, $t = 5.42$, $p < 0.001$), speaking ($\beta = 0.34$, $t = 5.99$, $p < 0.001$), reading ($\beta = 0.30$, $t = 4.74$, $p < 0.001$) and writing ($\beta = 0.25$, $t = 4.32$, $p < 0.001$) in ELSE. In addition, critical consuming was also found to have positive prediction for learners' self-efficacy in the two English language skills, namely listening ($\beta = 0.15$, $t = 2.59$, $p < 0.05$) and reading ($\beta = 0.13$, $t = 2.06$, $p < 0.05$).

5. Discussion

This study employed two questionnaires to explore the relationship between EFL learners' English new media literacy and their English language self-efficacy. Our findings reveal that both instruments are valid and reliable for assessing students' English new media literacy and their English language self-efficacy.

5.1 Correlations between EFL learners' English new media literacy and their self-efficacy

More importantly, this study explored the relationship between EFL learners' English new media literacy and their English self-efficacy. The correlation analysis indicates that there is a close interplay between EFL learners' ENML and their ELSE. This result is in accordance with previous findings about the positive link between new media literacy and self-efficacy (e.g. Rachels & Rockinson-Szapkiw, 2018).

5.2 Predictive roles of learners' English new media literacy on their self-efficacy

The regression analyses further reveal the following predictive roles of learners' English new media literacy for their English self-efficacy. First, functional consuming and functional prosuming are significant predictors for all the factors of students' English self-efficacy, namely, listening, speaking, reading and writing. Functional media literacy refers to the individual's competency to use media tools for accessing and creating media messages and understanding them at the textual level (Chen et al., 2011). Our research further echoes previous findings that audio-visual aids may help to reduce EFL learners' test anxiety in both listening and reading (Lee, Lee, Liao, & Wang, 2015). Similar evidence can also be found in Chartrand's (2012) and Lin, Warschauer and Blake's (2016) study with respect to the positive influence of social-networking sites on students' confidence in using their target languages.

This study also reveals that critical consuming is a major predictor for English language self-efficacy in both listening and reading. A critical consumer is capable to interpret the social economic, political and cultural contexts of the media content (Chen et al., 2018). It suggests that EFL learners with the ability to analyze, evaluate and synthesize media content tend to be more confident in their listening and reading activities.

6. Conclusion and Implications

The current study validates the research instrument for evaluating EFL learners' English new media literacy and self-efficacy in English learning. It further explores the interplay between learners' English new media literacy and their English self-efficacy and discloses the intricate relations between the two constructs in a Chinese higher education setting. According to the results, it is revealed that both functional consuming and functional prosuming are identified as the two critical factors for predicting students' sense of self-efficacy in English listening, speaking, reading and writing. Besides, students with higher critical consuming literacy tend to possess more confidence in listening and reading self-efficacy.

Given the increasing significance of emerging technologies in education, several pedagogical implications are also provided for fostering learners' English language self-efficacy. First, it is essential for EFL instructors to skillfully integrate a variety of new media tools into language teaching and deliberate over the socio-cultural elements of new media, so as to provide students with more timely and appropriate scaffolding tailored to their diverse need, abilities and proficiencies (Zheng & Warschauer, 2018). Second, this inquiry also serves as a reminder for English language educators and policy-makers of the necessity of integrating new media literacy education in EFL school curriculum. Besides, due to the insufficiency of self-report questionnaires, more in-depth qualitative studies through interviews or observations could shed more light on the complex interplay between language learners' English new media literacy and their self-efficacy. Finally, because of the limited generalizability of our findings, it should be replicated in different contexts to give a more complete picture of the relationships between English new media literacy and English language self-efficacy.

Acknowledgements

This research was supported by the teaching reform projects of Beijing University of Posts and Telecommunications (Grant 2018JY-B09 and 2019Y002), the graduate students' teaching reform projects of School of Humanities, Beijing University of Posts and Telecommunications (Grant RW-YJ-201903), and QC Education Innovation Institute., China

References

- Bai, R., Hu, G., & Gu, P. Y. (2014). The relationship between use of writing strategies and English proficiency in Singapore primary schools. *The Asia-Pacific Education Researcher*, 23(3), 355-365.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

- Chartrand, R. (2012). Social networking for language learners: Creating meaningful output with Web 2.0 tools. *Knowledge Management & E-Learning: An International Journal*, 4(1), 97-101.
- Chen, D. T., Wu, J., & Wang, Y. M. (2011). Unpacking new media literacy.
- Chen, D. T., Lin, T. B., Li, J. Y., & Lee, L. (2018). Establishing the norm of new media literacy of Singaporean students: Implications to policy and pedagogy. *Computers & Education*, 124, 1-13.
- Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. MIT Press.
- Laire, D., Casteleyn, J., & Mottart, A. (2012). Social media's learning outcomes within writing instruction in the EFL classroom: Exploring, implementing and analyzing storify. *Procedia-Social and Behavioral Sciences*, 69, 442-448.
- Lee, L., Chen, D. T., Li, J. Y., & Lin, T. B. (2015). Understanding new media literacy: The development of a measuring instrument. *Computers & Education*, 85, 84-93.
- Lee, S. P., Lee, S. D., Liao, Y. L., & Wang, A. C. (2015). Effects of audio-visual aids on foreign language test anxiety, reading and listening comprehension, and retention in EFL learners. *Perceptual and motor skills*, 120(2), 576-590.
- Lin, C. H., Warschauer, M., & Blake, R. (2016). Language learning through social networks: Perceptions and reality. *Language Learning & Technology*, 20(1), 124-147.
- Lin, T. B., Li, J. Y., Deng, F., & Lee, L. (2013). Understanding new media literacy: An explorative theoretical framework. *Journal of Educational Technology & Society*, 16(4), 160-170.
- Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1-2), 72-89.
- Su, Y., Zheng, C., Liang, J. C., & Tsai, C. C. (2018). Examining the relationship between English language learners' online self-regulation and their self-efficacy. *Australasian Journal of Educational Technology*, 34(3), 105-121.
- Sun, Y. C., & Yang, F. Y. (2015). I help, therefore, I learn: service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning*, 28(3), 202-219.
- Towndrow, P. A., & Pereira, A. J. (2018). Reconsidering Literacy in the 21st Century: Exploring the Role of Digital Stories in Teaching English to Speakers of Other Languages. *RELC Journal*, 49(2), 179-194.
- Wang, C., Kim, D. H., Bai, R., & Hu, J. (2014). Psychometric properties of a self-efficacy scale for English language learners in China. *System*, 44, 24-33.
- Wang, C., Schwab, G., Fenn, P., & Chang, M. (2013). Self-efficacy and self-regulated learning strategies for English language learners: Comparison between Chinese and German college students. *Journal of Educational and Developmental Psychology*, 3(1), 173.
- Wu, P., & Hsu, L. H. (2011, November). EFL learning on social networking site?: An action research on Facebook. In *Teaching & Learning with Vision Conference. Queensland, Australia*.
- Yang, P. L. (2013). Discourse Analysis of EFL College Learners' Online Social Interaction and Attitudes towards FACEBOOK. *International Journal of English Linguistics*, 3(6), 64.
- Zheng, B., Yim, S., & Warschauer, M. (2018). Social media in the writing classroom and beyond. *The TESOL Encyclopedia of English Language Teaching*, 1-5.

Investigating Chinese University EFL Learners' Self-Efficacy in a Blended Learning Environment

Jingjing LIAN

School of Humanities, Beijing University of Posts and Telecommunications, China
lianjingjing@bupt.edu.cn

Abstract: This study examined the effect of a blended learning mode on students' English self-efficacy and self-regulation in a Chinese university, and further explored factors associated with their English listening self-efficacy. Results showed that after 16-week blended learning implementation, students' English listening self-efficacy, *Task Strategies* and *Self-Evaluation* have significantly improved. Students from more developed places reported higher level of English self-efficacy as well as the sources of English self-efficacy. Further OLS regression models identified *Social Persuasion* as a strong predictor of students' listening self-efficacy. The implications of the study are also discussed.

Keywords: Sources of self-efficacy, self-efficacy, self-regulation, EFL, blended learning

1. Introduction

Blended learning (BL), or the integration of face-to-face and online instruction, is widely adopted across higher education (Graham 2013). The "Guidelines on College English Teaching" issued in 2017 by the Ministry of Education of China urges College English to "integrate information technology with curriculum", and to "implement a blended teaching mode, in a way to foster students' active, autonomous and personalized learning".

In blended learning settings, autonomous learning is required as students have to complete learning tasks online before and after class, and set their own learning pace. Yet the lack of self-regulated competence and learning belief can lead to non-compliance issues, such as failure in watching online videos before class (He, Holton, Farkas, & Warschauer, 2016). Although much research has investigated learners' self-regulation and self-efficacy in blended learning settings (Barnard, Lan, To, Paton, & Lai, 2009; Shea & Bidjerano, 2010), few have focused on Chinese university-level EFL learners.

2. Literature Review

Self-efficacy refers to beliefs in one's capabilities to organize and execute the course of action required to produce given attainments (Bandura, 1997). Four principal sources of information constructed self-efficacy beliefs, namely, *mastery experiences*, *vicarious experiences*, *social persuasion* and *physiological states* (Bandura, 1997). Most researchers have used adapted versions of the Sources of Mathematics Self-Efficacy Scale developed by Lent, Lopez, and Bieschke (1991) to measure the sources of self-efficacy in academic settings (Usher & Pajares, 2008).

Self-regulation refers to the self-directive process through which learners transform their mental abilities into task-related academic skills (Zimmerman, 2001). Learners with stronger self-regulatory capacity are known to be more active and effective in academic task performance (Zimmerman & Schunk, 2011). And highly efficacious students are more likely to report the use of self-regulatory strategies (Kim, Wang, Ahn, & Bong, 2015).

As the construct of self-efficacy is domain specific and context specific (Zimmerman & Cleary, 2006), in order to better understand Chinese EFL learners' self-efficacy, Wang, Kim, Bai and Hu

(2014) have developed a self-efficacy questionnaire in an attempt of transferring the educational psychology constructs to the field of second language acquisition (SLA). Su, Liang and Tsai (2018) further validated the scale of the questionnaire and examined the positive relationship between EFL learners' online self-regulation and their English self-efficacy.

This study intends to further explore Chinese EFL learners' English self-efficacy and online self-regulation after a blended learning approach. Two research questions were proposed:

(1) What is the effect of blended College English approach on students' English self-efficacy and online self-regulation?

(2) What are the factors that may be associated with students' English listening self-efficacy?

3. Method

3.1 Participants

The study involved 135 sophomore (second year) students in a College English class at a Chinese comprehensive university in northern China. There are more male students (67%) in the class as all of the participants are engineering majors. Their ages ranged from 17 to 25 years, with an average age of 20.22 years (SD=0.77).

3.2 Instruments

The study employed three questionnaires. The first questionnaire is the Sources of EFL Learners' English Self-Efficacy (SESE) adapted from Zheng, Liang and Tsai (2017) with four dimensions included: *Mastery Experiences*, *Vicarious Experiences*, *Social Persuasion* and *Physiological States*. The second instrument is the Questionnaire of English Self-Efficacy (QESE) from Wang et al (2014). It is composed of four variables of *Listening Self-Efficacy*, *Speaking Self-Efficacy*, *Reading Self-Efficacy* and *Writing Self-Efficacy*. The third questionnaire is Online Self-regulated English Learning (OSEL) (Zheng, Liang, Yang, & Tsai, 2016). The variables are *Goal Setting*, *Environment Structuring*, *Task Strategies*, *Time Management*, *Help Seeking* and *Self-Evaluation*.

All the items in the three questionnaires were translated into Chinese and measured with a five-point scale, from 1 "I do not agree at all" to 5 "strongly agree", or from 1 "I cannot do it at all" to 5 "I can do it well". The Cronbach's alpha of the SESE, QESE and OSEL are 0.92, 0.93 and 0.88 separately. All the data analyses including *t*-test and OLS regression were processed in Stata 15.

3.3 Procedure

The research was conducted in the fall semester of 2017 in an integrated college English course aimed at improving students' listening, speaking, reading and writing. The course lasted for 16 weeks. The same lecturer was in charge of two parallel classes with 68 and 69 students.

Two identical sets of questionnaires were administered during class time at the second and the fifteenth week to guarantee completion rate. The lecturer designed a blended classroom approach with "iClass" (Blackboard) and an online learning platform "New Perspective" developed by Shanghai Foreign Language Education Press. The "iClass" platform was mainly used for curriculum notices and students' assignments. The "New Perspective" platform provides the digital version of the textbook, learning resources and online tests. In order to support students' online English learning self-regulation, the lecturer provided a variety of listening materials to foster students' learning interests, and also created a "WeChat class group" as an out-of-class communication platform.

The mid-term listening test is selected as it contains more listening questions compared with the final exam. It includes 25 multiple-choice questions testing students' listening comprehension of conversations, passages and news reports.

3.4 Data Analysis

To answer the first research question, paired-sample *t*-test of pre- and post- survey data was conducted. For the second research question, OLS regression of students' English listening self-efficacy on students' gender, high school location, sources of self-efficacy and online English learning self-regulation was reported.

Table 1 shows the descriptive statistics of students' age, gender, listening test and all the mean values of their post-survey variables by students' high school location. One-way ANOVA was conducted to compare the group differences and *F*-ratio (3, 131) is listed in the last column.

Table 1
Descriptive Statistics of All Analysis Variables by Students' High School Location

	Full Sample		Municipality		Provincial Capital		Other Cities		Town or Village		<i>F</i>
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Listening Test (Standard Score)	0.00	1.00	0.69	1.02	0.27	1.06	-0.23	0.92	-0.27	0.84	6.43***
Age	20.22	0.77	20.14	0.36	20.31	0.47	20.22	1.05	20.22	0.67	0.18
Male	0.67		0.67		0.5		0.75		0.70		1.64
<i>English Self-Efficacy</i>											
Listening Self-efficacy	2.92	0.60	3.27	0.65	3.14	0.53	2.81	0.59	2.71	0.52	6.06***
Speaking Self-efficacy	3.46	0.63	3.77	0.74	3.58	0.49	3.50	0.56	3.16	0.64	5.27**
Reading Self-efficacy	3.28	0.64	3.62	0.84	3.29	0.45	3.38	0.55	2.93	0.62	6.95***
Writing Self-efficacy	3.60	0.56	3.59	0.77	3.74	0.53	3.66	0.47	3.43	0.54	1.91
<i>Sources of English Self-Efficacy</i>											
Mastery Experiences	2.58	0.74	2.93	0.66	2.72	0.88	2.52	0.76	2.39	0.56	3.00*
Vicarious Experience	3.77	0.68	3.86	0.77	3.85	0.71	3.77	0.67	3.67	0.64	0.49
Social Persuasion	2.98	0.87	3.55	0.82	3	0.98	2.86	0.87	2.80	0.71	4.07**
Physiological States	2.69	0.94	2.35	1.03	2.41	0.88	2.75	0.98	3.01	0.78	3.31*
<i>Online English Learning Self-Regulation</i>											
Goal Setting	3.02	0.69	3.25	0.78	3.15	0.71	3	0.65	2.83	0.64	2.09
Environment Structuring	3.68	0.63	3.45	0.68	3.78	0.59	3.86	0.62	3.49	0.56	3.89*
Task Strategies	2.85	0.81	2.79	0.85	2.71	0.84	3.04	0.77	2.73	0.81	1.50
Time Management	3.03	0.77	2.76	0.92	3.31	0.63	3.13	0.75	2.85	0.71	2.81*
Help Seeking	3.20	0.81	3.38	0.60	2.92	0.82	3.38	0.89	3.04	0.73	0.29
Self-Evaluation	3.08	0.82	3.06	0.93	3	0.65	3.16	0.90	3.04	0.77	3.09*
Observations	135		21		26		51		37		

Note. All the survey items are measured on a 5-point rating scale and all the data except the listening test are from the post-survey. In mainland China, four cities are identified as municipality—Beijing, Tianjing, Shanghai and Chongqing.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. Results

4.1 Paired-Sample *t*-test

Table 2 shows the result of the paired-samples *t*-test of students' English self-efficacy and online English learning self-regulation. Only listening self-efficacy shows a significant improvement in post-survey ($t=2.89, p<0.01$). This result implies that after a semester of blended learning approach, students' listening self-efficacy has significantly improved, yet no significant change is found in their speaking, reading and writing self-efficacy.

As for their online English learning self-regulation, the result shows significant improvement in their *Task Strategies* ($t=4.80, p<0.001$) and their *Self-Evaluation* ($t=2.12, p<0.05$). The mean value of *Task Strategies* is the lowest among all the variables in the pre-survey (2.46, $SD=0.84$). After the treatment, the mean value of *Task Strategies* improved to 2.85 ($SD=0.81$).

Table 2
Paired Samples t-test between Pre-Survey and Post-Survey

Dimension		Mean	SD	<i>t</i>	<i>P</i>
<i>English Self-efficacy</i>					
Listening Self-efficacy	post-survey	2.92	0.60	2.89**	0.005
	pre-survey	2.79	0.61		
<i>Online English Learning Self-Regulation</i>					
Task Strategies	post-survey	2.85	0.81	4.80***	0.000
	pre-survey	2.46	0.84		
Self-Evaluation	post-survey	3.08	0.82	2.12*	0.036
	pre-survey	2.92	0.79		

Note. All the survey items are measured on a 5-point rating scale.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

4.2 OLS Regression Model

As only the listening self-efficacy shows a significant improvement after the blended learning mode, OLS regression models were used to further investigate factors associated with students' listening self-efficacy. In the first model of table 3, students' listening test score is a strong predictor of students' listening self-efficacy. In Model 2, comparing with students who are from municipality (the reference group), students from towns or villages on average show significantly lower listening self-efficacy, controlling for other variables.

In both Model 3 and Model 4, *Social Persuasion* is a strong predictor of students' listening self-efficacy ($p<0.01$). With one out of a five-point scale increase in students' *Social Persuasion* belief, their listening self-efficacy increases 0.18 point. In Model 4, students' *Goal Setting*, *Environment Structuring* and *Task Strategies* are significantly associated with students' listening self-efficacy ($p<0.05$), controlling for other variables. The *R*-squared value (0.43) in Model 4 indicates that the model explains 43.3% of the total variances in students' listening self-efficacy.

Table 3
English Listening Self-Efficacy Regressed on Students' Listening Test, Gender, High School Locations, Sources of English Self-Efficacy and Online English Self-Regulation

	Model 1	Model 2	Model 3	Model 4
Listening Test	0.08***(0.01)	0.06***(0.02)	0.03(0.02)	0.03(0.02)
Male	0.08(0.10)	0.10(0.10)	0.14(0.10)	0.20*(0.10)
<i>Students' High School Locations</i>				
Provincial Cities		-0.02(0.16)	0.05(0.15)	0.12(0.16)
Ordinary Prefectural Cities		-0.27(0.15)	-0.22(0.14)	-0.19(0.14)

Towns and Villages		-0.36*(0.16)	-0.29(0.15)	-0.26(0.15)
<i>Sources of English Self-Efficacy</i>				
Mastery Experiences			0.10(0.08)	-0.03(0.09)
Vicarious Experience			0.10(0.08)	0.09(0.08)
Social Persuasion			0.18**(0.07)	0.20**(0.07)
Physiological States			0.04(0.05)	0.03(0.05)
<i>Online English Learning Self-regulation</i>				
Goal Setting				0.21*(0.09)
Environment Structuring				-0.21*(0.08)
Task Strategies				0.13*(0.06)
Time Management				0.04(0.07)
Help Seeking				-0.02(0.07)
Self-evaluation				0.04(0.07)
Constant	1.64***(0.26)	2.05***(0.32)	1.20**(0.43)	1.14*(0.44)
R-squared	0.18	0.23	0.34	0.43
Observations	135	135	135	135

Note. All the survey items are measured on a 5-point rating scale and all the data except the listening test are from the post-survey. Standard errors are in parentheses. Municipality is the reference group in high school location. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5. Discussion

5.1 The Blended Approach

The results from the paired samples *t*-test showed that only students' listening self-efficacy had improved over the 16-week implementation of the blended approach. During the semester, students spent more time doing listening activities and tests online, as blended learning mode provides more high-quality multimedia resources to train their listening comprehension than traditional classrooms.

In terms of students' online English learning self-regulation, only *Task Strategies* and *Self-Evaluation* showed significant improvement. Students on average were more likely to read aloud the English instructional materials online to fight against distractions, and took more thorough notes when learning online. They have also reported improvement in finding how they were doing in online learning by communicating with teachers and classmates. Generally speaking, more targeted and specific treatment is needed to improve EFL students' speaking, reading and writing self-efficacy as well as their online English self-regulation.

5.2 Factors Associated with Students' English Listening Self-Efficacy

As shown in Table 1, students from more developed cities on the whole reported higher level of English self-efficacy as well as the sources of English self-efficacy. The difference in mean values between "municipality" and "provincial capital" is not as sharp as the others because most provincial capitals are also well-developed big cities. In China, students in bigger cities have better English learning resources and enjoy more communicative-driven teaching pedagogies in high school. On the contrary, high school English teaching is more test-driven in small cities or towns.

However, the effect of students' high school location disappears with more controlling variables added to the OLS regression model. Table 3 shows that *Social Persuasion* is a strong predictor of students' listening self-efficacy, which corresponds with the finding in Zheng et al (2017). Encouragement and praise from parents, teachers and peers whom the students trust can boost their confidence in their English ability, and this is more prominent in China given the collective cultural and social context.

6. Conclusion

This study has further validated the instrument for measuring Chinese EFL learners' self-efficacy and provides some preliminary findings in a blended learning approach carried out in a Chinese university. It provides a clearer picture of the relations among students' English listening proficiency, English self-efficacy and their online English learning self-regulation.

However, this study has many limitations. Further research with more sophisticated design is needed to get a better understanding of the effect of blended learning approach on learners' self-efficacy and self-regulation.

Acknowledgements

This paper is supported by the Beijing University of Posts and Telecommunications Teaching Reform Project (2019JY-C08).

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.
- Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The internet and higher education, 12*(1), 1-6.
- Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education*, (3rd ed., pp. 333–350). New York: Routledge.
- He, W., Holton, A., Farkas, G. A., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction, 45*, 61-71.
- Kim, D. H., Wang, C., Ahn, H. S., & Bong, M. (2015). English language learners' self-efficacy profiles and relationship with self-regulated learning strategies. *Learning and Individual Differences, 38*, 136-142.
- Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of counseling psychology, 38*(4), 424.
- Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education, 55*(4), 1721-1731.
- Su, Y., Zheng, C., Liang, J. C., & Tsai, C. C. (2018). Examining the relationship between English language learners' online self-regulation and their self-efficacy. *Australasian Journal of Educational Technology, 34*(3), 105-121.
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of educational research, 78*(4), 751-796.
- Wang, C., Kim, D.-H., Bai, R., & Hu, J. (2014). Psychometric properties of a self-efficacy scale for English language learners in China. *System, 44*, 24–33.
- Zheng, C., Liang, J. C., Yang, Y. F., & Tsai, C. C. (2016). The relationship between Chinese university students' conceptions of language learning and their online self-regulation. *System, 57*, 66-78.
- Zheng, C., Liang, J. C., & Tsai, C. C. (2017). Validating an Instrument for EFL Learners' Sources of Self-Efficacy, Academic Self-Efficacy and the Relation to English Proficiency. *The Asia-Pacific Education Researcher, 26*(6), 329-340.
- Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. E. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.
- Zimmerman, B. J., & Cleary, T. J. (2006). Adolescents' development of personal agency: the role of self-efficacy beliefs and self-regulatory skill. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 45–69). Greenwich, CT: Information Age Publishing.
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. J. Schunk (Eds.), *Handbook of self-regulated learning and performance* (pp. 1–12). New York, NY: Routledge.

Multitask Learning for Chinese Grammatical Error Detection

Yu-jie ZHOU^a, Yong ZHOU^{b*}

^{a,b} *Department of Education Information Technology, East China Normal University, Shanghai*

*yzhou@ied.ecnu.edu.cn

Abstract: Chinese as a Foreign Language (CFL) learners often make grammatical errors such as missing words, selecting wrong words and wrong word order due to language negative migration. In this paper, we propose a neural sequence labeling model with a supplementary objective for Chinese grammatical error detection. We use the manually labeled dataset written by CFL learners to train the models. This multitask learning model has better performance than other sequence labeling model because it can learn the bias in the label distribution and learn richer features for semantic composition.

Keywords: Grammatical error detection, Chinese as a Foreign Language, Multitask learning

1. Introduction

The number of Chinese as a Foreign Language (CFL) learners has continuously increased these years. Unlike English or some other languages, Chinese sentences are composed with the string of characters without spaces to mark word boundaries. Also, Chinese has quite flexible expressions and loose structural grammatical, so it has been regarded as one of the most difficult languages in the world (Bo Zheng et al., 2016). CFL learners often make grammatical errors such as missing words, selecting wrong words and wrong word order due to language negative migration, over-generalization, teaching methods, learning strategies and other reasons.

Automated Grammatical error detection and correction system are very essential and invaluable to language learners because manual correction is time-consuming and laborious (Leacock et al., 2010). GEC for English has been studied for many years, with many shared tasks such as CoNLL-2013 (Ng et al., 2013) and CoNLL-2014 (Ng et al., 2014), however, few grammatical correction applications have been developed to support CFL learners because of the limited labeled data and the complexity of Chinese. The exist Chinese grammatical errors detection applications are based on a range of techniques, such as statistical learning (Chang et al, 2012; Wu et al, 2010; Yu and Chen, 2012), rule-based analysis (Lee et al., 2013) and Deep Learning-based models (Gaoqi et al., 2017; Lee et al., 2016, 2015; Yu et al., 2014).

In this paper, Chinese grammatical errors detection is considered as a sequence labeling problem which assigns each Chinese word in a target sentence with a tag indicating the error types. With limited labeled data, we use multitask learning to solve this problem. To be specific, we propose a neural sequence labeling architecture which consists a supplementary objective of predicting surrounding words in addition to labeling each token to encourage the framework to learn richer features for semantic composition without requiring additional training data.

2. Related Work

Grammatical error detection is a sub-task of sequence labeling in natural language processing which assigns semantic label to each word of the input sentence. Our work builds on previous research exploring sequence labeling model on grammatical error detection.

The researchers used many different methods to study the grammatical error detection task and achieved good results (Tou et al., 2017). As for English, Marcin et al., used phrase based translation

optimized for F-score using a combination of kb-MIRA and MERT with augmented language models and task-specific features, and got a good result(Junczys-Dowmunt and Grundkiewicz, 2014). As a universal language model, the Long Short-Term Memory network (LSTM) has achieved good results in many tasks in natural language processing in recent years, including text classification tasks, machine translation tasks, and sequence annotation tasks(Hochreiter and Schmidhuber, 1997). Rei et al., used the Encoder-Decoder model similar to neural machine translation to process the English Grammatical(Rei, Yuan and Briscoe, 2017).

Compared with English, the research time of Chinese grammatical error diagnosis system is short. In recent years, the Natural Language Processing Techniques for Educational Applications (NLPTEA) workshops have hosted a series of shared tasks for Chinese grammatical error diagnosis. Some researchers have done some works based on the given dataset. Bo et al., propose a CRF+BiLSTM model based on character embedding on bigram embedding, on the CGED-HSK dataset of NLP-TEA-3 shared task, their system presents the best F1-scores in all the three levels (Zheng et al., 2016).Ruiji et al., improved the model of bidirectional Long Short-Term Memory with a conditional random field layer (BiLSTM-CRF) with several new features and adopted a probabilistic ensemble approach. This model achieved the best F1 scores on NLPTEA-2018 CGED(Ruiji et al., 2018). Chen et al., proposed a two-stage hybrid system which combined the BiLSTM-CRF model along with some handcraft features and three GEC models which achieved the highest precision(Chen et al., 2018).

3. Task Definition

Table 1

Two errors are found in the sentence below, one is word selection error (S) at positon 8, the other is word ordering error (W) from position 9 to 12.

	国 ¹ 家 ² 不 ³ 应 ⁴ 该 ⁵ 盲 ⁶ 目 ⁷ 的 ⁸ 经 ⁹ 济 ¹⁰ 发 ¹¹ 展 ¹² 。 ¹³	
Error Type	S	W
Error Position-start	8	9
Error Position-end	8	12
Correction	国家不应该盲目地发展经济。	

The grammatical errors made by CFL learners are varied. The NLPTEA shared task for Chinese Grammatical Error Diagnosis (CGED) which has been hosted for years has divided those errors into four types, including redundant words (denoted as R), missing words (M), word selection errors (S), and word ordering errors (W) (Gaoqi et al., 2017,2018). The goal of this task is to detect these four types of grammatical errors. The input sentence may contain one or more grammatical errors. Example sentence and corresponding notes are shown in Table 1.

4. Methodology

In this section, we introduce the proposed multitask neural CRF sequence labeling model (MTN-CRF). First, we will introduce the sequence labeling model used in our main task. Then we will introduce the neural language model used in our auxiliary task.

4.1 Bi-LSTM grammatical error detection model

Our Bi-LSTM grammatical error detection model can efficiently use past input features via an LSTM layer. The model is shown in Figure 1, we build Bi-LSTM blocks (h1) for each input word, and concatenate these block in two directions to form the forward Bi-LSTM and backward Bi-LSTM. These Bi-LSTM blocks are parameter sharing.

Compared with traditional grammatical error detection models, Bi-LSTM based model mainly have two advantages. The first is that Bi-LSTM model can capture the context features for each word. For example, consider the input sentence “我 中文 学” in Figure 1, when predicting the tag for the Chinese word “中文”, the Bi-LSTM model can using the features from the whole sentence (i.e., “我” and “学”). While traditional models usually cannot capture such features since the data sparse issue and the computation is costly. For a given input length n , the Bi-LSTM model can learn such features in time complexity $O(n)$, while traditional models will have a time complexity $O(n^2)$. The second advantage is that Bi-LSTM model is an end to end model, which can extract features automatically. In traditional models, the model performance largely depends on the feature engineering. Doing such feature engineering is exhaust for the system developer.

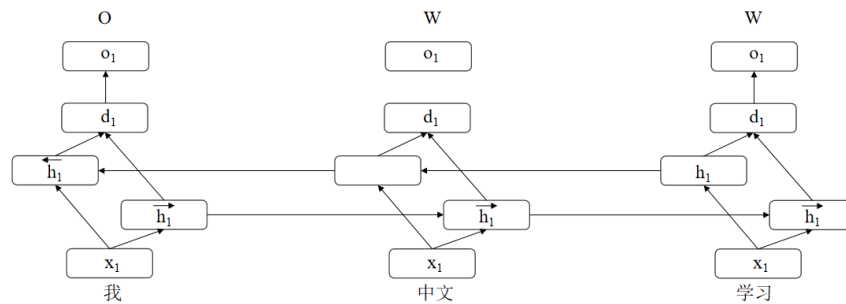


Figure 1. Bi-LSTM grammatical error detection model

4.2 Bi-LSTM language models

In this work, we also combine our grammatical error detection model with a Bi-LSTM language model to enhance the model performance. The Bi-LSTM language model is shown in Figure 2. The training of the Bi-LSTM is a unsupervised learning. It takes an natural language text as input, and models the probability of the next word at each time step. Consider the example in Figure 2, the input of the Bi-LSTM language model is a sentence without any labels (i.e., “我学中文”). For the first time step, the input of the LSTM unit is “我”, and the output of the Bi-LSTM is the probability of the next word. Suppose the word vocabulary is 3, includes “我”, “学”, “中文”, then the output of the first time step will be key-value pairs like (“我”: 0.2), (“学”: 0.5), (“中文”: 0.3). The key is the vocabulary, the value is the probability to occur at next time step. If the model is correctly trained, the word “学” will have a higher probability then others. This model is quite similar to the traditional n-gram language model in natural language processing. They both models the probability distribution of the next time step given each input word. In practice, the performance Bi-LSTM language model is better than the traditional n-gram language model.

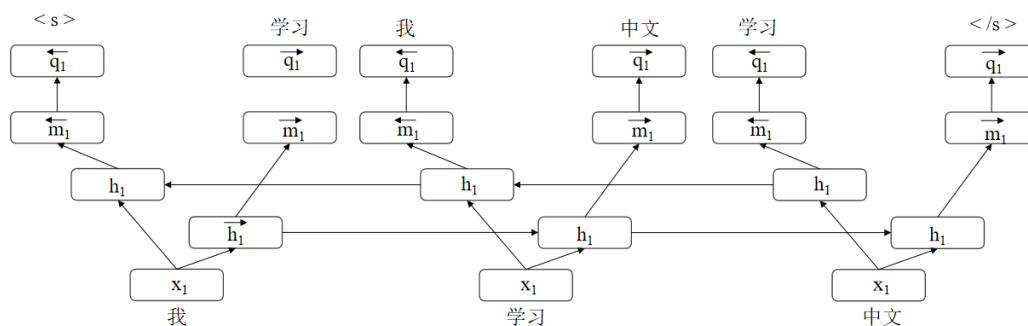


Figure 2. Bi-LSTM language models

4.3 Multitask learning framework

In this section, we will introduce how we combine the grammatical error detection model with the language model to form the multitask learning framework. The multitask learning framework is shown in Figure 3.

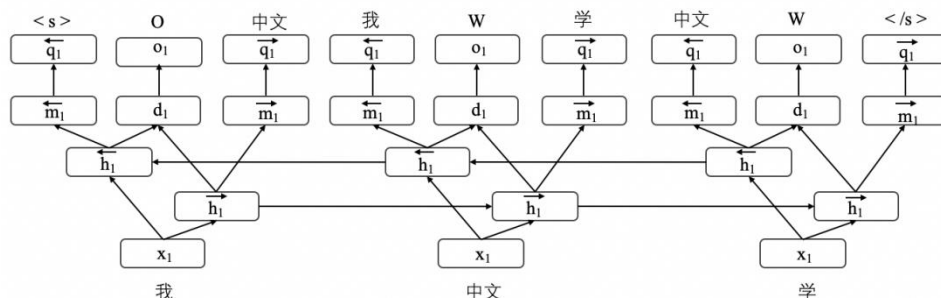


Figure 3. Multitask learning framework

The LSTM units are sharing on this two tasks. The output of the LSTM units contains both language model outputs and the grammatical error detection outputs. That is to say the model needs to predict both surround words and the grammatical error type at the same type. The loss of the mode is as bellow:

$$\text{Loss}_{\text{all}} = \text{Loss}_{\text{lm}} + \alpha \text{Loss}_{\text{ged}}$$

Where Loss_{all} is the joint loss of the multitask learning framework, Loss_{lm} is the loss of the language model, and the Loss_{ged} is the loss of the grammatical error detection model, α is the weight parameter of the Loss_{ged} which we set 1, and automatically tuned through training of the multitask learning framework.

5. Experiments and Evaluation

5.1 Dataset

The dataset we use was from the NLPTEA (Natural Language Processing Techniques for Educational Applications) shared task for Chinese Grammatical Error Diagnosis. The corpora used in CGED task were taken from the writing section of HSK (Pinyin of Hanyu Shuiping Kaoshi, Test of Chinese Level) which is for Chinese as a Foreign Language(CFL) learners. The data set which contains the origin sentence, the error type, the position of the error and the correction of the sentence is manually labeled by the experienced teachers.

Table 2 shows the distributions of error types in the training set, validation set and testing set. The ratio of training set size to validation set size is about 10:1. Besides the sentences with grammatical errors, there are over 40% of the sentences contain no error which was simulated the sampling in the writing sessions in HSK to test the performance of the systems in false positive identification.

For the supplementary objective, we use an external dataset Lang-81 to train the model, which contains more than 700,000 items, and each item consists of an original sentence and corresponding corrected sentences.

Table 2

The distributions of error types in datasets

	#R	#M	#W	#S
Training Set	10671(22.55%)	11955(25.26%)	3516(7.43%)	21178(44.75%)
Validation Set	574(21.95%)	682(26.08%)	171(6.54%)	1188(45.43%)
Testing Set	795(21.97)	928(25.64%)	281(7.76%)	1615(44.63%)

5.2 Results and discussion

The evaluation of the test results is determined three levels including detection-level, identification-level and Position level. Detection-level is binary classification of the given sentence, that is, correct or incorrect, should be completely identical. All error types will be regarded as incorrect. Identification-level could be considered as a multi-class categorization problem to identify the error type. Position-level judges the occurrence range of the grammatical error. Table 11 Table 12 and Table 13 shows the evaluation result for detection level, identification level and position level of multitask learning architecture on grammatical error detection datasets. The two baseline results are from our previous work in 2018 CGED shared task (Yujie et al., 2018).

Table 3

Evaluation results of sequence labeling architectures on Detection Level

Detection Level			
	Precision	Recall	F1
Baseline			
CRF	0.5923	0.5445	0.5993
BiLSTM-CRF	0.8202	0.5652	0.6692
Our Work			
Multitask learning	0.8314	0.5932	0.6924

Table 4

Evaluation results of sequence labeling architectures on Identification Level

Identification Level			
	Precision	Recall	F1
Baseline			
CRF	0.4452	0.2740	0.3392
BiLSTM-CRF	0.6068	0.4183	0.4952
Our Work			
Multitask learning	0.6342	0.4723	0.5414

Table 5

Evaluation results of sequence labeling architectures on Positon Level

Positon Level			
	Precision	Recall	F1
Baseline			
CRF	0.3532	0.1346	0.1949
BiLSTM-CRF	0.4631	0.2568	0.3303
Our Work			
Multitask learning	0.4735	0.2984	0.3661

According to experiment results, we found that Multitask learning model has better performance. The CRF baseline is low, because CRF model largely depends on feature engineering. It is hard to do feature engineering in grammatical error detection, because the training data is sparse. And it is also difficult to find certain feature to capture a specific error type. The BiLSTM-CRF model performs slightly better than CRF model, since it can automatically extract features for CRF models rather than handcraft feature engineering. But it still suffers from the data sparse issue. The multitask learning model performs better than Bi-LSTM-CRF on all three level. This is because the sequence labeling model is only optimized based on the labels contains information. While in our test set, over 40% of the sentences in the test set contain no error and 84% of all tokens have the label O (correct). So

many of the tokens in the dataset contribute very little to the training process. Multitask learning architecture which contain a supplementary objective is able to learn this bias in the label distribution without obtaining much additional information from the majority labels. It allows the model to make full use of the training data and get better results than other sequence labeling task in Grammatical error detection task.

6. Conclusion and Further Work

In this paper, we propose a neural sequence labeling model with a supplementary objective for Chinese grammatical error detection. We use the manually labeled dataset written by CFL learners to train the models. This multitask learning model has better performance than other sequence labeling model because it can learn the bias in the label distribution and learn richer features for semantic composition. For further work, we plan to address more complex errors in addition to the four-main error type in this paper and focus on Chinese grammatical error correction which may involve Machine Translation models.

References

- Lafferty J, Andrew M, and Fernando P, Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the eighteenth international conference on machine learning, In proc, ICML, 2001.
- Lample G , Ballesteros M , Subramanian S , et al. Neural Architectures for Named Entity Recognition, In proc. NAACL, 2016.
- Ma X, and Hovy E, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, In proc. ACL, 2016.
- Marek Rei, Semi-supervised Multitask Learning for Sequence Labeling, In proc. ACL, 2017.
- Helen Yannakoudakis, Marek Rei, Øistein E. Andersen and Zheng Yuan. 2017. Neural Sequence-Labeling Models for Grammatical Error Correction. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2795–2806
- Gaoqi Rao Qi Gong Baolin Zhang Endong Xun. 2018. Overview of NLPTEA-2018 Share Task Chinese Grammatical Error Diagnosis. In Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications, pages 42–51
- Grundkiewicz, R., & Junczys-Dowmunt, M. (2014). The WikEd Error Corpus: A Corpus of Corrective Wikipedia Edits and Its Application to Grammatical Error Correction. *Advances in Natural Language Processing*.
- Hochreiter, S. , & Jürgen Schmidhuber. (1997). Flat minima. *Neural Computation*, 9(1), 1.
- Rei, M., Felice, M., Yuan, Z., & Briscoe, T. (2017). Artificial error generation with machine translation and syntactic patterns.
- Bo Zheng, Wanxiang Che, Jiang Guo, and Ting Liu. 2016. Chinese grammatical error diagnosis with long short-term memory networks. In Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, pages 49–56.
- Yujie Zhou, Yinan Shao, Yong Zhou. 2018. Chinese Grammatical Error Diagnosis Based on CRF and LSTM-CRF model. In Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications, pages 165–17.
- Po-Lin Chen, Wu Shih-Hung, Liang-Pu Chen, and Ping-Che Yang. 2016. Improving the selection error recognition in a Chinese grammatical error detection system. *International Conference on Information Reuse and Integration*, pages 525-530.
- Yajun Liu, Yingjie Han, Liyan Zhuo, and Hongying Zan. 2016. Automatic grammatical error detection for Chinese based on conditional random field. In Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, pages 57–62.
- Wanxiang Che, Zhenghua Li, and Ting Liu. 2010. LTP: A Chinese language technology platform. In Proceedings of the Coling 2010: Demonstrations, pages 13-16.
- Zhiheng Huang, Wei Xu and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *Computer Science*. Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via Bi-directional LSTM-CNNs-CRF. In Proceedings of ACL, pages 1064-1074.

Improving Summary Writing Performance via Theory-based Learning System

Chiou Sheng CHEW^{a*}, Wen-Chi Vivian WU^b & Norisma IDRIS^c & Er Fu LOH^d

^a*Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, Melaka, Malaysia*

^b*Department of Foreign Languages and Literature, Asia University, Taichung, Taiwan*

^c*Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia*

^d*Academy of Language Studies, Universiti Teknologi MARA, Melaka, Malaysia*

[*cschew501@gmail.com](mailto:cschew501@gmail.com)

Abstract: While summary writing has been acknowledged by many scholars as an integral skill that builds the background knowledge of learners, and enhances writing outcomes, only a handful of theory-based online summary writing tools have been developed to date. In view of this, the Summary Writing-PAL (SW-PAL) was developed, rooted in several learning and education theories, in the attempt to assist ESL non-English majors in improving their summary writing skills. This study presents both the development and the evaluation of SW-PAL. The three primary features of SW-PAL are prior knowledge activation, summarising strategies instruction, and scaffolding. Prior knowledge activation applies a concept mapping tool as an advance organiser that activates the students' prior knowledge while comprehending the text. The worked examples tool is meant to aid students in acquiring the essential summarising strategies. The self-generated feedback provided by the tool serves as a scaffolding tool to assist students through the summarising process. Pre- and post-tests, as well as an interview session held with students, ascertained the effectiveness of the tool and perceptions of its users. The study outcomes revealed that the developed SW-PAL had managed to improve their summary-writing ability at a significant level.

Keywords: Summary writing, computer assisted learning, summarizing strategies, learning theories

1. Introduction

Most tertiary and secondary education institutions learning, particularly those established across ASEAN countries, view summary writing as a significant aspect of assessment to evaluate one's ability to comprehend texts written in the English language (Abdi, Idris, Alguliyev, & Aliguliyev, 2016; Idris, Baba, & Abdullah, 2011; McDonough, Crawford, & De Vleeschauwer, 2014; Wichadee, 2014). Students, due to poor text comprehension and summary writing skills, may end up rewriting parts of the original excerpt in a haphazard manner. Writing apprehension stems partly due to lacking of writing skills (Wichadee, 2014). Typically, summary writing is taught as follows: 1) identifying the main idea and deleting unimportant content (deletion); 2) identifying umbrella terms or general words based on the main idea and supporting details (generalisation); and 3) identifying and rephrasing the main idea to improve the sentence (construction). Students with exceptional summarising skills may better understand the text, and hence, produce better summaries.

1.1 Reading Comprehension

Comprehending reading texts simply refers to understanding the gist of the written texts by applying cognitive skills. In order to comprehend a text effectively, one would require good decoding skills (Kintsch, 1988) and prior knowledge (Best, Ozuru, Floyd, & McNamara, 2006). Activated prior knowledge aids learning, while the opposite hinders learning (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). Prior knowledge is vital while reading a text to gain better understanding of the gist,

while summary writing has been proven to connect old and new knowledge effectively so as to enhance reading comprehension (Marzec-Stawiarska, 2016).

1.2 Summary Writing

The essentials of summary writing are generating a concise text comprising of only important information while discarding explanatory and supporting details from the original source. Summary writing skills are imminent to understand the gist of a text, which are tested across schools in ASEAN countries (Idris et al., 2011). The key to writing a good summary is by integrating important ideas into a single paragraph in accordance to summarising or macro rules. Brown and Day (1983) listed five macro rules for summarising a text, which are: elimination of unimportant information, deleting redundant information by rewording, and restating several vital sentences. Additionally, similar words are substituted with synonyms and the gist found in the source text is rephrased in the writer's own words (Idris et al., 2011; Lemaire, Mandin, Dessus, & Denhière, 2005). Despite introducing these strategies to the students, some continue to fail in writing exceptional summary (Idris et al., 2011; McDonough et al., 2014).

1.3 Relationship between Reading Comprehension and Summary Writing

The Zone of Proximal Development (ZPD), Transactional, Meaningful Learning, and Cognitive Load theories are significant upon assessing the correlation between reading comprehension and summary writing (Ausubel, 1963; Rosenblatt, 1988; J. Sweller, 1988; Vygotsky, 1978). Typically, transaction takes place between the source text and the students' prior knowledge. Activation of prior knowledge is essential to understand the gist of a text, which in turn, facilitates learning (Mason, Ariasi, & Boldrin, 2011). Vast prior knowledge enhances reading comprehension skills (Calisir & Gurel, 2003). Summary writing can only take place after understanding the gist of a text. In enabling students to write summary effectively, scaffolding must be integrated with ZPD and cognitive load should be decreased. A computer-assisted learning tool that addresses these issues may enhance the summary writing ability among students.

The conceptual framework displayed above motivated the researchers to build a theory-based computer-assisted summary writing learning tool called Summary Writing-Pal (SW-PAL) to aid students learn summary writing skills. The purpose of this study is two-fold. First, it determined if SW-PAL can improve the performance of students to write summary, and second, it analysed the perceptions of students who were exposed to SW-PAL.

The research questions of this study are listed in the following:

1. Is there a significant difference in ESL students' summary writing performance after using the SW-PAL?
2. What are the students' perceptions towards the use of SW-PAL?

2. The Summary Writing-PAL (SW-PAL)

The three primary components in SW-PAL are Prior Knowledge Activation (PKA), Summarising Strategies Instruction (SSI), and Scaffolding (SC). The first component, PKA, activates knowledge for students to understand the text effectively with the aid of their prior knowledge prior to summary writing. Here, the concept of mapping tool serves as an advance organiser to improve both text comprehension and summary writing (Sung, Liao, Chang, Chen, & Chang, 2016). The second component, SSI, exposes students to a range of summarising strategies through the use of worked example instructional approach. According to Sweller, Ayres, and Kalyuga (2011), the effect of split-attention may be obtained by integrating various information sources. The last component, SC, is a feedback tool that identifies strategies for the students to check their summary writing strategies. It is vital to ensure the correct use of summary writing strategies. This SC feedback tool assures that students learn and practice writing summaries independently, thus minimising the teacher's workload.

3. Methods

3.1 Subjects and research design

Twenty-five Malaysian undergraduate students, four males and twenty-one females, were exposed to SW-PAL to learn summary writing. A pre-test and post-test experimental design was adopted to assess the impact of SW-PAL on the performance of the subjects in writing summaries. This study took five weeks, with a week for pre-test, three weeks for SW-PAL intervention, and the final week for post-test.

3.2 Instruments

The scores retrieved from pre- and post-test determined whether the subjects had improved their performance in writing summary. The sample text was selected by using the Flesch Reading Ease (FRE) readability index. An expository excerpt was applied for both pre- and post-test, while the other text types had been used for the SW-PAL intervention period. Two worked examples of the summary were used for every text, along with a range of strategies for summary writing prepared via SW-PAL.

3.3 Procedure

A pre-test, six SW-PAL practice sessions, a post-test, and a semi-structured focus-group interview had been carried out for all the selected subjects in the study. The pre-test took place in the first week, which required the subjects to summarise a text after reading it. Next, six practice sessions were carried out between week two and week four to teach the subjects SW-PAL operation and summary writing via worked examples, as prepared by the instructor. The subjects practiced with a text for every 90-minute session. Lastly, the post-test was performed in the fifth week after completing the practice phase. All procedures implemented in the post-test were similar to those used for pre-test.

By using the Grading Rubric for summary (Desoiza, 2011), two ESL lecturers assesses the summaries retrieved from both pre- and post-test. The Grading Rubric was composed of five criteria: main ideas, accuracy, words and style, organisation, and length. Next, the marking scheme had four-level grading: exemplary (4 points), proficient (3 points), adequate (2 points), and needs to improve (1 point). Upon summing the scores for every criterion, score for cumulative summary performance was retrieved (5-20 points). Lastly, the scores were assessed using paired samples t-test.

3.4 Semi-structured focus-group interview

Five subjects were selected for a semi-structured focus-group interview by adhering to the interview protocols devised by the researchers. The interview was performed for several purposes, including: (1) to assess their learning experiences with SW-PAL, (2) to determine if they liked the SW-PAL features, and lastly, (3) to gain feedback about SW-PAL features. The interview was recorded, transcribed, and analysed in accordance to qualitative study design.

4. Findings

4.1 Effectiveness of SW-PAL

Research question 1 determined the effectiveness of SW-PAL in improving the summary writing performance displayed by the subjects. The descriptive statistics (see Table 1) showed that the mean score of the post-test (M=12.88) exceeded the mean score of the pre-test (M=10.04) significantly. Based on the paired-sample t-test (see Table 2), the mean variance between pre-and post-test was 2.84; reflecting significant improvement among the subjects at 5%, $t = -11.70$, and $p < 0.005$. Hence, SW-PAL had significantly enhanced the students' performance in their summary writing.

Table 1

Descriptive statistics of subjects

Test	N	Mean (M)	Std. Deviation
Pre-test	25	10.04	2.09
Post-test	25	12.88	1.72

Table 2

Paired-Sample t-test between Pre-test and Post-test scores of subjects

Test	Mean Difference	t	df	Sig. (2-tailed)
Pre-Test Score – Post-Test Score	-2.84	-11.70	24	.000

4.2 Students' perceptions towards SW-PAL

Research question 2, which refers to students' perceptions towards SW-PAL, was addressed via semi-structured focus-group interview. The subjects admitted that they felt motivated to use the tool and agreed that the SW-PAL had greatly aided and enhanced their performance in writing summaries. According to them, the concept mapping was helpful in logically organising thoughts and concepts before they proceeded with summary writing, besides activating their prior knowledge while reading the text. The subjects claimed that they preferred applying the worked examples in order to learn the various strategies to write a summary effectively. As for the feedback tool, the subjects found it beneficial in checking their summary writing strategies.

5. Discussion and Conclusion

A theory-based computer-assisted learning tool called SW-PAL was successfully developed to aid summary writing, by integrating three primary aspects, namely: concept mapping, worked examples, and feedback features. This study determined the effectiveness of SW-PAL in enhancing the subjects' summary writing performance, including their perceptions towards the SW-PAL features.

A total of 25 ESL undergraduate students participated in this study to assess the developed SW-PAL via pre-test post-test experimental design. The study outcomes revealed that SW-PAL was indeed an effective tool that enhanced the subjects' ability to write summaries. In fact, the findings are in agreement with those reported in several past studies (Chiu, 2015; Wade-Stein & Kintsch, 2004), which highlighted improvement in language learning through the use of learning tool.

The subjects claimed that they preferred using SW-PAL for summary writing, particularly the worked examples that exposed them to a range of summary writing strategies. A few subjects admitted that they had anxiety at the initial stage when using the concept mapping, which served as advance organiser to activate their prior knowledge. The subjects also found the feedback tool beneficial in improving their language proficiency. The intent of the researchers is to perfect this developed SW-PAL in future.

A number of limitations were detected in this study. First, the sample size in this study (25 subjects) is insufficient to generalise the efficacy of SW-PAL to other populations. Second, the impacts of SW-PAL could be generalised merely to ESL population that share similar attributes, but not other ESL students with varied English language proficiency, age group, and learning institutions (university, college, matriculation, foundation). Lastly, the validity of the study outcomes is questionable due to the short empirical period – merely five weeks.

The main objective of this study is to assess the worked examples of SW-PAL. Future investigations may examine the impacts of concept mapping and feedback aspects or even their combination on the summary writing performance amongst a range of subjects, besides ESL students.

Additionally, this study calls for more assessments that analyse the intelligent feedback in SW-PAL, so as to enhance the ability of writing summaries.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References

- Abdi, A., Idris, N., Alguliyev, R. M., & Aliguliyev, R. M. (2016). An automated summarization assessment algorithm for identifying summarizing strategies. *Plos One*, *11*(1), e0145809.
- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.
- Ausubel, D. P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education*, *14*(2), 217–222.
- Best, R., Ozuru, Y., Floyd, R., & McNamara, D. (2006). Children's text comprehension: Effects of genre, knowledge, and text cohesion. *Proceedings of the 7th International Conference on Learning Sciences*, 37–42. Retrieved from <http://portal.acm.org/citation.cfm?id=1150040>
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 1–14.
- Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior*, *19*(2), 135–145.
- Chiu, C. (2015). Enhancing reading comprehension and summarization abilities of EFL learners through online summarization practice. *The Journal of Language Teaching and Learning*, *1*, 79–95.
- Desoiza, R. A. (2011). *Pictorial map effects on learning how to summarize*. Doctoral dissertation. University of San Francisco.
- Idris, Baba, & Abdullah. (2011). Identifying students' summary writing strategies using summary sentence decomposition algorithm. *Malaysian Journal of Computer Science*, *24*(4), 180–194.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension - A construction integration model. *Psychological Review*, *95*(2), 163–182.
- Lemaire, B., Mandin, S., Dessus, P., & Denhière, G. (2005). Computational cognitive models of summarization assessment skills. *Proceedings of the 27th Annual Conference of the Cognitive Science Society (CogSci'2005)*, 1266–1271.
- Marzec-Stawiarska, M. (2016). The influence of summary writing on the development of reading skills in a foreign language. *System*, *59*, 90–99.
- Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction*, *21*(1), 137–151.
- McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing*, *24*(1), 20–32.
- Rosenblatt, L. M. (1988). Writing and reading: The transactional theory. *Reading and Writing Connections*, *20*(fall), 153–176.
- Sung, Y. T., Liao, C. N., Chang, T. H., Chen, C. L., & Chang, K. E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers and Education*, *95*, 1–18.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(1), 257–285.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. *Cognitive Load Theory* (Vol. 1).
- Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Mind in Society The Development of Higher Psychological Processes*, *Mind in So*, 159.
- Wade-Stein, D., & Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362.
- Wichadee, S. (2014). Developing reading and summary writing abilities of EFL undergraduate students through transactional strategies. *Research in Education*, *92*(Nov 2014), 59–71.

Enhancing EFL College Students' Language Performance via eBook Supported Learning

Fang-ying Riva LO^{a*}, Wen-Chi Vivian WU^b, Cheng-Huan CHEN^c & Jun Scott CHEN HSIEH^d

^aCenter for General Education, Asia University, Taiwan

^bDepartment of Foreign Languages and Literature, Asia University, Taiwan

^cDepartment of M-Commerce and Multimedia Applications, Asia University, Taiwan

^dGraduate Institute of Network Learning Technology, National Central University, Taiwan

*flo@asia.edu.tw

Abstract: This study aims to analyze which particular learning behaviors with eBook supported learning are associated with high-achievers in language learning. This specifically focuses on English output (refers to English writing in this study) performance. Furthermore, to compare learning with and without an eBook, this study observes the correlation between self-regulated English learning (SREL), English self-efficacy (ESE), and learning achievements. This is done as a means to suggest which particular form of learning autonomy awareness should be explicitly addressed to boost language learning outcomes. This semester-long study was conducted with EFL students from the College of Management in a “General English for Specific Purposes” (GESP) course in a four-year comprehensive university in Taiwan. Pre-class reading materials were uploaded to an eBook platform and features of the platform were introduced beforehand to facilitate self-learning prior to class. Through analyzing the students’ 1) learning logs on the eBook platform, 2) scores on SREL and ESE questionnaires, and 3) learning achievement, the results revealed that online reading behaviors of adding markers and memos had a significant, positive correlation with students’ learning achievements. Meanwhile, the experimental group students outperformed those in the control group in open-ended, short essay style questions in terms of written statement quantity. With regards to SREL, scores in both the control and experimental groups increased throughout the study.

Keywords: Learning analytics, eBook, Self-regulated English learning, English self-efficacy, Language output

1. Introduction

The acquisition of a foreign language consists of language perception and language production. Language perception can be assessed through comprehension questions, and language production can also be assessed by various rubrics to determine the level of capacity. Evaluations of language perception and production are a standard practice used to ensure that the desired effectiveness of language teaching is met. Notably, the process before achieving language perception is often neglected. How second language learners process language input is an essential area of study, which can assist instructors in intervening at an early stage to offer the necessary help at the right time (LaBrozzi, 2016). Processing language input is a personal decoding process that used to be unrecorded when media was paper based. With the advance in technology, eBook supported learning allows instructors to possess a better understanding and influence on learners’ learning outside of the classroom (Lai, 2015). Unlike first language reading, second or foreign language reading requires skills in both decoding the language and the content. Whilst reading, second language readers regularly deal with the two sets of skills, simultaneously. Therefore, the significance of this study is to identify positive connections between second language reading behaviors and language performance. Through the adoption of eBook platforms, the connections of the two in question can be revealed partially if not wholly.

eBook reading behaviors can be associated with several learning models from the perspective of self-regulated learning (SRL). According to Zimmerman’s (2000) model, students undergo the stages of forethought, performance, and self-reflection, thus acquiring skills through monitoring their use of the platform. In Winne and Hadwin’s (1998) model, rooted in the Information Processing Theory (Greene & Azevedo, 2007), students acquire tactics and strategies through their interactions with the

platform. Several eBook reading behaviors have been analyzed as a means to effectively predict students' learning outcomes. Consequently, this allows the implementation of in-time intervention from instructors to facilitate students' learning. For example, Yamada and Goda (2016) discussed the relationship between procrastination in eBook reading and learning performance. By analyzing students' learning logs, instructors are able to focus on the specific behaviors that learners produce, which provide them with evidence-based insights for future classroom activity designs. BookRoll, the eBook platform utilized in this research, can track the following activities: opening a file, closing a file, adding bookmarks, deleting bookmarks, adding a marker, deleting a marker, adding a memo, deleting a memo, changing a memo, clicking the next page, clicking the previous page, jumping a page, and using the search function. The feature known as "Marker" allows learners to highlight important points in red and confusing words or statements in yellow. Moreover, "Memo" enables learners to annotate quick notes, self-reflection points, questions, comments, and the like.

While SRL refers to the ability to self-teach and monitor, self-efficacy (SE) refers to self-evaluation and reflection on learners' ability to complete tasks. Bandura and Schunk (1981) emphasized that there is a strong correlation between students' self-efficacy and actual ability. Studies of Second Language Acquisition have indicated a positive correlation between English self-efficacy (ESE) and learning performance. Finally, as Su et al. (2018) concluded, according to the results of questionnaires on self-regulated English learning (SREL) and ESE, SRL has a strong positive correlation with SE.

Based on the above introduction, this study aims to analyze the relationships among SREL, ESE, eBook supported learning, and language learning achievement. Additional attention will be placed on students' writing fluency as demonstrated through the word counts of short essay writing. Therefore, three research questions are established to evaluate the effectiveness of eBook learning activities for further instruction design in language curricula:

RQ 1. Which eBook supported learning behaviors are positively associated with language learning achievement?

RQ 2. What is the correlation between students' ESE and language learning achievement?

RQ 3. How do SREL and ESE evolve after a semester of study in both the control group and the experimental group?

2. Methodology

2.1 Subject and Course

This study was conducted in a university setting in Taiwan, with 82 sophomore students in two classes undertaking a required semester-long language course in "English for Business Communication." The control group consisted of 43 students, while the experimental group comprised of 39 students. The English proficiency of both groups were similar according to their English scores on the General Scholastic Ability Test taken prior to entering the university. Their average English proficiency was close to CEFR A2 to B1. While undertaking this required course, students were advised to read supplementary reading materials before class. In-class activities included weekly short essay writing assessments, group discussions, and micro-presentations. Short essay writing assessments were used to reflect their understanding of the assigned readings, while also used to demonstrate their critical thinking skills. Grading was weighted more on the content and the flow of expression, rather than the technical aspects of grammar or spelling. Notably, points were not removed due to grammatical or spelling mistakes unless such mistakes hindered the understanding of the essay. In-class group discussions would adopt the same open-ended questions as essay prompts. These were designed to boost students' collaborative critical thinking skills. After discussion, each group would appoint a representative to deliver a micro-presentation. Both the control and experimental groups had an identical course which consisted of pre-reading, short essay writing, group discussion, and micro-presentations to enhance their language skills in different aspects. The differences between the two groups included the following aspects: 1) the control group was given access to downloadable PDF files of reading materials, whereas the experimental group was introduced to BookRoll, an eBook platform. On BookRoll, students' reading behaviors could be tracked by the instructor, 2) in the experimental group, the instructor started each class with a quick review of the platform dashboard, which illustrated an overview of the class' reading status. This was, not only to ensure expectations of material previews

were met, but also to respond directly to questions or confusing concepts students marked or annotated on the platform. At the end of this course, students were evaluated by their weekly assessments scores, mid and post-course learning achievement tests, and group discussions. The learning activities and collected datasets are illustrated in Figure 1.

2.2 Data Collection

Data was collected from both the experimental group and the control group, which includes scores and word counts of the weekly assessments, pre and post-achievement tests, and SREL and ESE questionnaires. During this study, word counts of the weekly essay assessments were used as an indicator of learners' writing fluency. eBook learning logs were also collected for the experimental group.

Zheng et al. (2016) designed the SREL questionnaire regarding self-regulated learning, which focused on English learning. The questionnaire contains six components, referring to self-regulation theories in educational psychology proposed by Barnard et al. (2009) and Zimmerman & Schunk (2011). Along with the SREL questionnaire, Su et al. (2018) designed an ESE questionnaire which was a modification of Wang's et al. (2014). This was designed as a means to assess students' English self-efficacy in four skills: listening, speaking, reading, and writing.

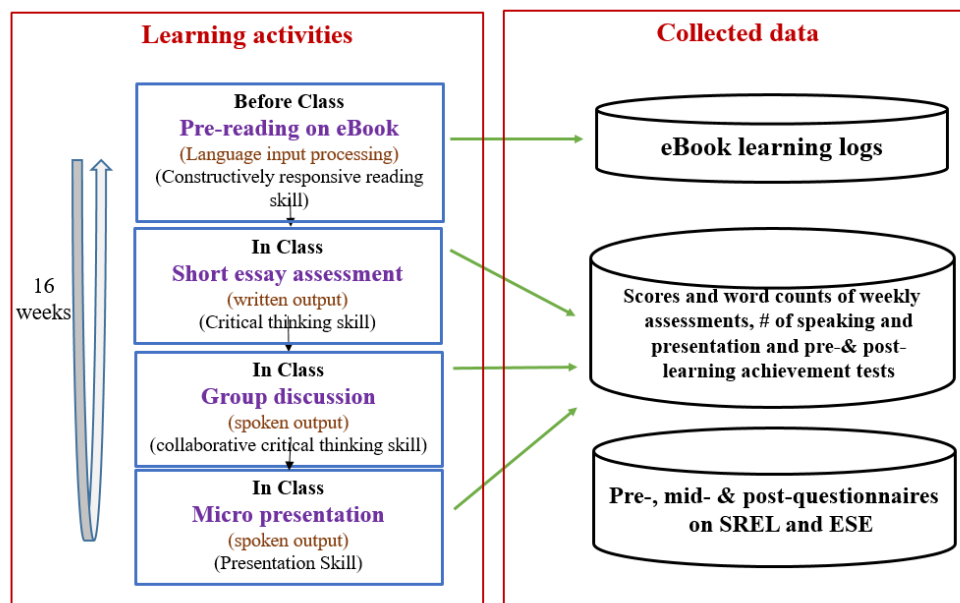


Figure 1. Learning activities and collected data from the experimental group and control group.

3. Results

3.1 Correlation between eBook Learning Behaviors and Academic Performance

According to the Spearman correlation analysis, we found that online reading behaviors that used the functions “add marker,” “delete marker,” and “add memo” on the eBook platform have a strong positive correlation with the students' learning achievement within the experimental group. Students used the marker function to highlight important points or unclear statements and added memos on the side of reading passages to note down translations of keywords or to raise questions. Table 1 shows that the features of “add marker” and “add memo” on eBook platforms may help students “consciously register” input and then transfer the input to be intake.

Table 1

Correlation between eBook Learning Behaviors and Academic Performance

	Feature	4-1 score	4-2 Score	5-1 score	5-2 score	U6 score	Final total	Final A	Final B	Course score
eBook file	Open	0.346*	0.199	0.505**	0.251	0.572***	0.510**	0.443**	0.516**	0.542**
	Close	0.222	0.204	0.230	0.134	0.368*	0.328*	0.239	0.414*	0.437**
Bookmark	Add Bookmark	0.349*	0.157	0.125	-0.147	0.232	0.255	0.252	0.199	0.226
	Delete Bookmark	0.365*	0.130	0.160	-0.102	0.204	0.262	0.280	0.200	0.244
Marker	Add Marker	0.465**	0.345*	0.245	0.226	0.285	0.612***	0.642***	0.350*	0.379*
	Delete Marker	0.386*	0.121	0.259	0.243	0.180	0.456**	0.444**	0.325*	0.377*
	Marker	0.478**	0.342*	0.261	0.234	0.288	0.622***	0.664***	0.350*	0.391*
Memo	Add Memo	0.413*	0.269	0.075	0.215	0.171	0.375*	0.434**	0.190	0.403*
	Delete Memo	0.167	0.104	0.160	0.211	0.069	0.170	0.220	0.132	0.193
	Change Memo	0.272	0.203	0.024	0.216	0.136	0.285	0.326*	0.175	0.247
	Memo	0.394*	0.242	0.061	0.218	0.129	0.357*	0.427**	0.171	0.364*
eBook page	Next	0.177	0.300	0.305	0.296	0.450**	0.333*	0.337*	0.245	0.349*
	Prev	0.317	0.029	0.252	0.304	0.418*	0.238	0.243	0.251	0.261
	Jump	0.179	-0.291	-0.133	-0.167	-0.397*	-0.166	-0.200	-0.041	-0.100
	Search	-0.018	0.097	0.090	0.250	0.093	0.236	0.206	0.184	0.108

3.2 Correlation between English Self-Efficacy and Academic Performance: On the Quantity of Written Output

Table 2 shows the t-tests for the control group and experimental group on writing fluency. Students in the experimental group outperformed the control group in terms of the flow of their writing.

Table 2
T-test of Control Group and Experimental Group on Writing Fluency/ Word Count

	Control group		Experimental group		t-test
	Mean	Std	Mean	Std	
1-1 word count	46.12	33.74	47.62	28.10	0.213
1-2 word count	43.55	28.18	41.89	33.13	-0.240
2-1 word count	22.45	20.59	37.22	33.92	2.369*
2-2 word count	22.19	22.05	29.38	25.36	1.348
3-1 word count	24.26	16.23	34.24	25.7	2.089*
mid word count	37.19	31.83	74.49	56.38	3.675***
4-1 word count	31.38	26.20	36.54	31.09	0.800
4-2 word count	16.31	16.21	29.78	31.56	2.420*

Table 3 indicates that the experimental group's results on the final ESE questionnaire have a stronger correlation with their academic performance, especially in reading, writing, and internet skills. The results also indicate that the correlation between ESE and academic performance continued to intensify over the pre-, mid-, and post-course ESE questionnaires. Students' use of an eBook platform which allows them to "add memo" may have had a positive influence in helping them raise their writing ESE, thereby enhancing the flow of their written output in open-ended essay questions.

Table 3
Correlation between Students' English Self-Efficacy and their Academic Performance (Experimental Group- Post-test)

	Listening	Speaking	Reading	Writing	Internet
Pre-test A (mid)	0.205	0.367*	0.433**	0.490**	0.275
Pre-test B (final)	0.336	0.297	0.477**	0.399*	0.154
1-1 score	0.280	0.211	0.303	0.339*	0.291
1-2 score	0.089	0.184	0.277	0.210	0.341*
2-1 score	0.383*	0.380*	0.447**	0.284	0.348*
2-2 score	0.214	0.227	0.205	0.167	0.067
3-1 score	0.005	0.069	0.067	0.040	0.247
mid total	0.255	0.356*	0.494**	0.420*	0.487**
mid A (multiple choice)	0.182	0.222	0.397*	0.355*	0.323
mid B (short essay)	0.299	0.423**	0.492**	0.414*	0.523**

4-1 score	-0.014	0.063	0.005	0.105	0.233
4-2 score	0.168	0.120	0.216	0.151	0.151
5-1 score	-0.205	-0.040	0.095	0.183	0.312
5-2 score	-0.005	0.007	0.017	-0.070	0.160
U6 score	0.008	0.121	0.119	0.135	0.331*
Final total	0.079	0.237	0.425**	0.397*	0.458**
final A (multiple choice)	-0.058	0.102	0.280	0.271	0.358*
final B (short essay)	0.252	0.366*	0.472**	0.482**	0.420*
Semester score	0.109	0.221	0.351*	0.276	0.395*

Table 4 shows the correlations between students' ESE post-tests and their academic performance in writing fluency in the experimental group. Students in the control group do not exhibit a distinctive correlation between the two, whereas students in the experimental group show a positive correlation in listening, speaking, reading, and internet skills.

Table 4
Correlation between Students' English Self-efficacy (Post-test) and their Academic Performance in Writing Fluency (Experimental Group)

	Listening	Speaking	Reading	Writing	Internet
1-1 word count	0.282	0.327*	0.290	0.274	0.241
1-2 word count	0.114	0.191	0.256	0.184	0.441**
2-1 word count	0.349*	0.340*	0.423**	0.211	0.377*
2-2 word count	0.126	0.200	0.098	0.135	0.044
3-1 word count	0.125	0.094	0.085	0.077	0.280
mid word count	0.288	0.313	0.367*	0.264	0.445**
4-1 word count	0.059	0.113	0.142	0.169	0.255
4-2 word count	0.158	0.060	0.195	0.072	0.084
5-1 word count	-0.109	-0.032	0.193	0.119	0.289
5-2 word count	0.078	0.066	0.051	-0.071	0.194
U6 word count	0.042	0.108	0.162	0.088	0.327*
final word count	0.275	0.186	0.303	0.196	0.198

3.3 T-tests of SREL and ESE pre- and post-tests

Tables 5 and 6 reveal that SREL scores on “task strategies” and “self-evaluation” increased from the pre-test to the post-test in the control group; whereas the scores on “help-seeking” and “self-evaluation” increased in the experimental group. Both groups improved their scores on “self-evaluation.”

Table 5
T-test of SREL and ESE (Control Group's Pre-test and Post-test Scores)

	Pre		Post		t-test of Pre and Post
	Mean	Std	Mean	Std	
Goal setting	0.71	0.14	0.73	0.16	0.502
Environment structuring	0.74	0.14	0.77	0.13	1.206
SREL Task strategies	0.64	0.13	0.71	0.16	2.349*
Time management	0.66	0.15	0.69	0.17	0.801
Help seeking	0.68	0.15	0.74	0.16	1.752
Self-evaluation	0.64	0.13	0.73	0.16	2.632*

Table 6
T-test of SREL and ESE (Experimental Group's Pre-test and Post-test Scores)

	Pre		Post		t-test of Pre And Post
	Mean	Std	Mean	Std	
Goal setting	0.72	0.12	0.74	0.13	0.807
Environment structuring	0.73	0.14	0.74	0.14	0.460
SREL Task strategies	0.63	0.14	0.69	0.15	1.781
Time management	0.65	0.14	0.71	0.15	1.602
Help seeking	0.64	0.17	0.73	0.14	2.317*
Self-evaluation	0.62	0.14	0.69	0.16	2.059*

4. Conclusion and Future Works

eBook platforms possess great potential in elevating the abilities of instructor to proactively monitor students' progress of acquiring language input through reading. The results of the study demonstrate that the eBook functions of "add marker," "delete marker," and "add memo" have a strong positive correlation with academic achievement. This indicates the importance of developing active reading strategies that could empower learners to be more constructive and responsive whilst reading. Future applications in teaching may include the introduction of active reading activities such as paraphrasing, predicting or commenting to increase the productive usage of the "add memo" tool. The study also indicates that the monitoring of reading activities on an eBook platform by the instructor could potentially help students. Such method could assist by transferring reading input to written output more effectively, as more in-time interventions and clarifications are proactively provided for confusing points. Further research could direct attention to spoken output analysis. Moreover, such studies could observe how identical eBook supported learning processes can affect output in the spoken form.

Acknowledgements

We would like to thank Professor Hiroaki Ogata's team from Kyoto University, Japan for granting the usage of BookRoll platform in this study. Special thanks also go to Meng-Wei Mavise Lin from Asia University in sharing EFL instruction insights and a team led by Professor Stephen J.H. Yang from National Central University, specifically Dr. Anna Y.Q. Huang, for assistance in data analysis.

References

- Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, 41, 586-598.
- Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *Internet and Higher Education*, 12(1), 1-6.
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, 77, 334-372.
- LaBrozzi, R. (2016). The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. *Language Teaching Research*, 20, 75-91.
- Lai, C. (2015). Modeling teachers' influence on learners' self-directed use of technology for language learning outside the classroom. *Computers & Education*, 82, 74-83.
- Su, Y., Zheng, C., Liang, J., & Tsai, C. (2018). Examining the relationship between English language learners' online self-regulation and their self-efficacy. *Australasian Journal of Educational Technology*, 34(3), 105-121.
- Wang, C., Kim, D., Bai, R., & Hu, J. (2014). Psychometric properties of a self-efficacy scale for English language learners in China. *System*, 44, 24-33.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.
- Yamada, M., & Goda, Y., (2016). How does self-regulated learning relate to active procrastination and other learning behaviors?. *Journal of Computing in Higher Education*, 28(3), 326-343.
- Zheng, C., Liang, J. C., Yang, Y. F., & Tsai, C. C. (2016). The relationship between Chinese university students' conceptions of language learning and their online self-regulation. *System*, 57, 66-78.
- Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1-12). New York: Routledge.

Augmented Reality and 3D Model for Children Chinese Character Recognition - Hong Kong Primary School Education

Ka-Yan FUNG^{ab*}, Kwong-Chiu FUNG^{ab} & Dr. Sally, Wai-Yan WAN^c

^aGabi Education Limited, Hong Kong,

^bThe Hong Kong University of Science and Technology

^cChinese University of Hong Kong, Hong Kong

*gabrieliefung@gabi.hk

Abstract: Unlike other phonetic alphabet languages, the pronunciation and writing of Chinese are different learning modes. This study intends to motivate primary students to learn Chinese characters and improve learning performance by integrating Augmented Reality (AR) technology with 3D models. The paper first describes the exertion of AR and 3D models to construct a Chinese character recognition application and then presents the findings of the pilot test in assessing the effectiveness of AR and 3D models in learning Chinese characters. Data collection methods are pre-test, post-test, teacher focus group, and student focus group. The study reveals that the prototype of this AR learning application enables students to have a better understanding of the recognition of Chinese characters. This enhances the learning motivation, concentration, and interest of school children. In summary, the study shows an interactive and interesting way to promote Chinese character learning via the AR environment and 3D models. Implications for future development in using AR and 3D models in supporting Chinese language learning discusses at the end of the paper.

Keywords: Augmented Reality (AR), 3D models, Chinese character recognition, e-learning, game-based learning, learning effectiveness

1. Introduction & Literature Review

Unlike other phonetic alphabet languages, the learning of Chinese is an abstract learning mode for primary school students. The pronunciation and writing of Chinese are different learning modes. We pronounce “big” as “dà”, but we write “big” as “大”. Besides, many Chinese characters are similar (Lou, 2000), such as “大” (big), “太” (super) and “犬” (dog). Students easily write wrong Chinese characters by adding “more stroke” (e.g. write 大 as 太) or “less stroke” (e.g. write 犬 as 大). In schools, students need to copy each Chinese character three times in doing homework in Chinese language subject. When they write a Chinese character wrong, teachers will punish them by copying the correct Chinese character five times. So, students are frustrated to learn Chinese characters due to lack of motivation. This is the reason why the study is conducted. This study aims to examine the efficiency of employing Augmented Reality (AR) technology with three-dimensional (3D) models in Chinese language learning. We would like to find out if the technologies could motivate Hong Kong primary students to learn Chinese characters better thus improve their ability in Chinese character recognition. The paper first describes the exertion of AR and 3D models to construct a Chinese character recognition application and then presents the findings of a pilot test in assessing the effectiveness of the application of AR and 3D models in learning Chinese characters. Data collection methods are pre-test, post-test, teacher focus group, and student focus group.

1.1 Chinese Language Medium

In Hong Kong, we are generally taught Chinese language subject by using Cantonese and traditional Chinese writing. According to the HKSAR Legislative Council Panel on Education (HKSAR, 2018). Putonghua has been encouraged by the local government to be the education medium of the Chinese

language subject since 2008. Currently, about 71.7% of elementary schools and 36.9% of secondary schools in Hong Kong have adopted Putonghua as a medium of instruction in teaching Chinese.

1.2 Background of Implementing Chinese Language Education

Chinese education has encountered drawbacks in Hong Kong. Hong Kong Education Bureau established a uniform standard for the handwritten Chinese characters in 1986. Teachers generally followed the rules listed on “List of Graphemes of Commonly-used Chinese Characters” to grade students’ handwriting. However, some parents complained that teachers were too strict with the students’ handwriting. The “correct” writing of Chinese characters’ was varied to teachers’ judgments (Li, 2018). Students felt frustrated from learning Chinese as they might be punished by not writing a “proper” Chinese characters fulfilling to the invisible and vague teachers’ standard (Li, 2018).

1.3 Common Difficulties in Learning Chinese Language

Chinese learners always face three difficulties in mastering the Chinese language – (1) many Chinese characters and difficult to recognize, (2) many shapes and difficult to write, and (3) many pronunciation and difficult to pronounce (Tse, 2000). As local primary school students, they usually encountered the first two difficulties in studying the Chinese language - character recognition and character writing. Unlike English, the Chinese word, Hanzi, can either operate solely or integrate with more than one character to compose a word. A word can be interpreted as a “multicharacter” word, such as 好 (good) is integrated with 女 (girl) and 子 (son). We usually compose a story to remember the multicharacter word (i.e. If a lady has a daughter and a son, she is good). However, primary students may find the Chinese characters difficult to distinguish because they do not have a systematic learning mode. If students could not recognize the meaning of the Chinese characters, it is even more difficult for them to memorize the script of the words (Jyun, 2017). A research conducted by Chan and Nunes (1998) shown evidence that the writing strokes arrangement of Chinese characters could consolidate the comprehension of learners about the Chinese characters to a certain extent.

1.4 Effectiveness of Technology-Assisted Education

The complementarity of technology in education has been proven its constructive effect on learning. The application of technology was generally associated with positive influence in language education. Zhao (2003) found that it could effectively raise the value of the input, efficiency of message transmission and ensure the relevance and validity of the feedback. Golonka, Bowles, Frank, Richardson & Freynik (2014) also pinpointed technology as an impressive instrument to boost learning desire and language-learning performance. Nevertheless, the learning outcome of practicing technology also hinge upon the practical application of technology and the teaching environment. Hence, more research regarding the proper technology operation format and context was encouraged to be done (Salaberry, 2001).

1.5 Augmented Reality as Language Learning Medium

Learning languages can be experienced through a wide variety of media. Students traditionally learned language through interaction with teachers and peers. They would learn from non- interactive media such as textbooks, copybooks, and instructional broadcasting programs. In past decades, digital media has been used in enhancing language learning efficiencies such as power-point slides, projectors, and interactive whiteboards. Furthermore, the ways of interaction with language learning experiences are changing. Students interact with on-screen content with fingers, paper, and real-world environments. Augmented Reality (AR) is an interactive experience of a real-world environment to offer perceptually enriched learning experiences. AR is used as a complementary tool in terms of text, graphics, video, and audio to provide students a real-time learning environment in language education institutions (Yuen, Yaoyuneyong & Johnson, 2011). Interactive AR media strongly influences students’ learning in terms of visual 3D functions and specific content to scaffold student learning (Radu, 2014). Researchers researched AR technology assisting in different language learnings, such as English (Chang, Lee, Wang & Chen, 2010; Liu & Tsai, 2013), Thai (Thiengtham & Sriboonruang, 2012) and French (Perry, 2015). AR technology improved students’ understanding and learning motivation in language learning.

2 Application of Augmented Reality (AR) technology with 3D models

An application provides a scenario for students to recognize Chinese characters at their own pace. The application provides games in 6 learning sections. Each section has 3 learning exercises. Students took a bone card to scan through a back-end camera of an electronic device. After scanning the “bone card” (Figure 2), a 3D animal model (Figure 4) popped up. Then, students pressed an animal cartoon (Figure 5) to listen to the animal sound and pressed a Chinese character (Figure 5) to listen to the pronunciation. Then, students entered the Game 1 (Figure 6) to recognize the pronunciation of the character and the animal cartoon. After finishing Game 1, students automatically entered Game 2 (Figure 7). Students pressed the underscore “_” and a window popped up (Figure 8). Students picked and scanned a relevant pinyin card (Figure 3). When students answered correctly, students automatically entered Game 3 (Figure 9). In Game 3, students pressed the animal to listen to the pronunciation and matched with the correct Chinese character. Students repeated the games with different “bone cards” (Figure 2).

3 Pilot Test on Chinese Character Recognition



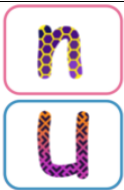
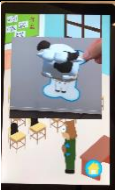


This is an AR interactive language learning application. The application provides a systematic language learning experience for primary school students. A pilot test is conducted to examine the effectiveness of the application in assisting primary school students to recognize Chinese characters.




3.1 Participants and Procedures

A local primary school participated in a pilot test. Mandarin is the language of instruction in the Chinese language subject. Students learned traditional Chinese characters in the school. Three teachers and 30 students recruited to participate in the pilot test. After the pilot test, the teachers and six students participated in focus groups separately. Before the execution of all pilot tests, a parental consensus for the information collection of the participants was collected.

Three teachers shared a similar experience in Chinese language subject teaching. The students were native Chinese speakers. They were from 6-year-old to 7-year-old, who were beginners in Chinese learning. 10 students were randomly selected for three groups, including one control group and two experimental groups.

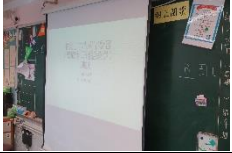
The pilot test lasted 40 minutes each session. We did one session for each group in the school. All groups spent the first 5-minute in a pre-test. Then, they spent the next 30-minute to learn six characters. Finally, they spent the last 5-minute in a post-test. The pre- and post-test were administered by students individually. Each focus group lasted 15 minutes.

		
<p>Figure 1. Gabi for Chinese Character Recognition</p>	<p>Figure 2. Learning Material (Bone Cards)</p>	<p>Figure 3. Pinyin Card</p>
		
<p>Figure 4. AR Technology & 3D Model for Chinese Character Recognition</p>	<p>Figure 5. Interaction Animation and Interactive Sound Attraction</p>	<p>Figure 6. Learning Material</p>

		
Figure 7. Interaction Animation and Interactive Sound Attraction	Figure 8. AR Technology for Chinese Character Recognition	Figure 9. Interactive Sound Attraction and Learning Material



3.2 Grouping Arrangement

The same set of pre- and post-test was used in the three groups. Group 1 was a traditional class (control group), which teacher (Miss YIP) taught students by using PowerPoint with the blackboard. Group 2 was a semi-traditional class (experimental group), which teachers (Miss LEE) taught students by using PowerPoint with the Augmented Reality Technology with 3D Models Application. Group 3 (experimental group) was a non-traditional class, which teachers (Miss LEUNG) taught students by using the Augmented Reality (AR) Technology with 3D Models Application.

		
Group 1	Group 2	Group 3

3.3 Curriculum Materials & Pilot Test Materials

There were nine modules in a curriculum. We used Module One in the pilot test. It was about *Twelve Zodiacs (Appendix)*, which was integrated with AR technology, bone cards, animal cartoons, 3D animals' models, sound attraction and pronunciation. The goal of this pilot test was to examine the effectiveness of using Augmented Reality (AR) technology with 3D models in motivating primary students to learn Chinese characters and improve learning performance. The curriculum design, user interface (UI) design and user experience (UX) design were produced by the first author. Twelve Zodiacs (Figure 1) were chosen as the content of the curriculum because the application wants to bring out the topic of Chinese culture. Six zodiacs were chosen for this pilot test. The six zodiacs were six Chinese characters to represent six animals in Chinese, including Cow (牛), Sheep (羊), Rabbit (兔), Tiger (虎), Snake (蛇) and Dog (狗). The pre- (Figure 10) and post-test (Figure 11) assessment we used consisted of 1 assessment item with six zodiacs, cow, sheep, rabbit, tiger, snake, and dog.

	
Figure 10. Pre-test	Figure 11. Post-test

4. Findings

4.1 Analysis of learning performance

About *Table 1*, the sample size of each group was ten students. The full scores of each test were 6. The average pre-test scores of Group 1 were 5.8. The average pre-test scores of Group 2 were 5.9. The average pre-test scores of Group 3 were 5.0. The average pre-test scores of Group 1 and Group 2 were extremely similar and exhibited a difference of 0.1. The average pre-test scores of Group 1 were quite obvious to Group 3 and exhibited a difference of 0.8.

Table 1: Mean scores, sample size and improvement for pre-test and post-test

Variables	Pre-test	Post-test	Difference	Improvement	Sample Size
	Mean	Mean	Mean	%	N
Group 1	5.8	6.0	0.2	3.45	10
Group 2	5.9	6.0	0.1	1.69	10
Group 3	5.0	5.8	0.8	16	10

After all learning activities were completed, the average post-test scores of Group 1 and Group 2 were 6.0 respectively. The average post-test scores of Group 3 were 5.8. The average post-test scores of Group 1 and Group 2 were the same. The percentage of scores in Group 1 increased by 3.45%. The percentage of scores in Group 2 increased by 1.69%. The percentage of scores in Group 3 increased by 16%. By comparing the improvement of total score gain in three groups, Groups 3 exhibited the most significant improvement after performing the learning activities by the full employment of AR technology with 3D models.

Regarding Table 2, the improvement of each character was recorded among all three groups. After the intervention of tutorial guidance, students' learning performance was generally improved. Nevertheless, the non-traditional class achieved considerable progress. For those Chinese characters with more strokes, such as horse (馬), snake (蛇), and dog (狗), the improvement was relatively significant by the employment of AR technology with 3D models. The improvement rate was from 10% to 20%.

Table 2: Chinese Character Recognition Improvement

	Cow牛	Rabbit兔	Horse馬	Snake蛇	Dog狗	Sheep羊
Group 1	--	--	--	+10%	+10%	--
Group 2	+10%	--	--	+10%	+10%	--
Group 3	--	+10%	+20%	+20%	+20%	+10%

4.2 Student Focus Group

All interviewees expressed their endorsement to the technology employed in the pilot test as they found the 3D animal models fantastic. Majority of interviewees mentioned that they have used similar applications to learn English, but they never use to learn Chinese. Moreover, all interviewees were able to tell all the words they learned. For Group 3's interviewees, they could also connect the related features of the words they learnt, such as the animals' sound. It demonstrated that the application could effectively consolidate learners' memories. Students suggested including functions that could show them the common mistakes they made in writing.

4.3 Teacher Focus Group

Teachers agreed that the technology exerted in the lesson could efficiently boost students' learning motivation. Students were keen to finish the writing and listening exercises correctly as they wanted to collect the whole serial of the 3D animal models. The application enabled learners to master the Chinese language in an enjoyable environment. However, teachers raised their concern in using electronic devices during lessons. When students had technical problems in using the application, schools may not have enough technical assistance. Besides, teachers may need to spend extra energy on class discipline to be in order. If the application acts as a supplementary tool for learning, it could effectively tackle the problem of individual learning difference. Furthermore, teachers mentioned the difficulties and insufficient support provided by the school's management team.

5. Conclusion and Discussion

The study reveals that the prototype of this AR learning application enables school children to have a better understanding of the recognition of Chinese characters, thus strengthen the interest of study, concentration, and improvement of Chinese character recognition. The pilot tests provide a satisfactory result on Chinese character recognition. The result supports the fact that Augmented Reality (AR) technology with 3D models could motivate students to learn Chinese characters and improve their

learning performance. Both teachers and students provided positive feedback to the 3D animal models as powerful tools to enhance students the learning motivation. No doubt, the larger sample size will be more convincing.

Some teachers prefer using traditional teaching methods in the lesson because it is easier to control the lessons' orders and follow the lesson plan. When teachers adopt electronic devices in teaching, more man-power is needed to control class discipline. Besides, teachers may be easier to lag behind the planned schedule due to individual learning differences. All interviewed teachers agreed that technology could enhance learning efficiency and effectiveness. We still need to discuss how we could take the balance in between. In the future, we will conduct pilot tests with a larger sample size. So, the performance will be more significant. Apart from Chinese character recognition, we will conduct pilot tests in Chinese character writing in future research.

Acknowledgements

We would like to thank Dr. Richard LUI, Miss LEUNG, Miss LEE and Miss YIP to arrange and assist in the pilot tests. We would like to thank Miss Doreen CHONG to provide us valuable feedback.

Appendix

Twelve Zodiacs, known as Sheng Xiao, is based on a twelve-year cycle. Each year in that cycle related to an animal sign. These signs in order are the mouse, cow, tiger, rabbit, dragon, snake, horse, sheep, monkey, chicken, dog and pig. It is calculated according to Chinese lunar calendar.

References

- Chan, L., & Nunes, T. (1998). Children's understanding of the formal and functional characteristics of written Chinese. *Applied Psycholinguistics*, 19(1), 115-131.
- Chang, C. W., Lee, J. H., Wang, C. Y., & Chen, G. D. (2010). Improving the authentic learning experience by integrating robots into the mixed-reality environment. *Computers & Education*, 55(4), 1572-1578.
- Liu, P. H. E., & Tsai, M. K. (2013). Using augmented-reality-based mobile learning material in EFL English composition: An exploratory case study. *British Journal of Educational Technology*, 44(1), E1-E4.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer assisted language learning*, 27(1), 70-105.
- HKSAR. (2018). LCQ21: The use of Putonghua as the medium of instruction for teaching the Chinese Language Subject in primary and secondary schools. Retrieved from <https://www.info.gov.hk/gia/general/201802/07/P2018020700609.htm?fontSize=1>
- Lou, Ch. K. (2000). 掌握漢字特點的識字教學方法——分析比較. Retrieve from http://www.cuhk.edu.hk/ics/clrc/crcl_53/lou.pdf
- Jyun, Z. (2015). 袁進. 糾結的香港中小學語文教學。《灼見名家》 Retrieve from <https://www.master-insight.com/糾結的香港中小學語文教學/>
- Perry, B. (2015). Gamifying French Language Learning: a case study examining a quest-based, augmented reality mobile learning-tool.
- Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6), 1533-1543.
- Salaberry, M. (2001). The Use of Technology for Second Language Learning and Teaching: A Retrospective. *Modern Language Journal*, 85(1), 39-56.
- Thiengtham, N., & Sriboonruang, Y. (2012). Improve template matching method in mobile augmented reality for thai alphabet learning. *Int. J. Smart Home*, 6(3), 25-32.
- Tse, S.K (2000). 高效識字。載於中文教育 <http://www.chineseedu.hku.hk>
- Yuen, S. C. Y., Yaoyuneyong, G., & Johnson, E. (2011). Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1), 11.
- Zhao, Y. (2013). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO journal*, 21(1), 7-27.

Diagnostic Language Assessment: Lessons Learned from Rapid Prototyping

Allan NICHOLAS & John BLAKE

Center for Language Research, University of Aizu, Japan

anich@u-aizu.ac.jp

Abstract: In this paper, we provide the rationale for a computerized diagnostic language assessment (DLA) of email writing for Japanese undergraduates. The development of a rapid prototype of the DLA system is then described, and lessons learned from the process are shared. The DLA focuses on *pragmatic competence* – the ability to alter one’s language choices relating to politeness, formality and structure, depending on the social context of the communication. While tools now exist to assist with learners’ grammar problems and vocabulary, few exist for this pragmatic aspect of communication. Furthermore, traditional language assessments typically provide an overall score of learner ability without identifying specific strengths and weaknesses. The DLA seeks to address this issue via a three-stage process: 1) assessment administration; 2) provision of feedback to the learner, and 3) provision of further individualized instruction based on the assessment results. To gain a better perspective of the problems and to start to climb the learning curve as quickly as possible, rapid prototyping using Axure RP was selected. This high-fidelity prototype of the DLA was designed for some use cases. Numerous lessons were learned. Notably, rapid prototyping is not necessarily rapid. Early feedback on usability and the user experience were received from the rapid prototyping, helping the developer gain a better understanding of the user needs and discover any unexpected usability issues prior to actual coding. Another ongoing advantage is that rather than describing the expected functionalities of the software to stakeholders or interested parties, the functionalities of the fully-fledged DLA can be shown using the prototype. The prototype functions as an interactive visual aid, reducing miscommunication and providing a focus around which discussions can be based.

Keywords: EFL, diagnostic language assessment, pragmatics, rapid prototyping, Azure RP

1. Introduction

As part of an ongoing larger project, in this paper we describe and reflect on the development of a rapid prototype of a computerized diagnostic language assessment (DLA) for Japanese EFL learners’ English email writing. The particular focus for the DLA is on the pragmatic aspect of learners’ email writing – the relationship between the social contexts of the emails and the language choices the learners make. We put forward a rationale for the use of DLA in the EFL classroom, and the ways in which computerization can enhance the potential of this form of assessment. We then describe the steps taken to develop the basic DLA prototype, and conclude with the lessons learned by the developers from the prototype development process.

2. Background

2.1 Diagnostic Language Assessment (DLA)

Traditional language assessments typically focus on scoring overall learner ability, without analyzing specific strengths and weaknesses, and so may not be helpful to teachers or students (Poehner, Zhang & Lu, 2015). This issue is addressed with recent diagnostic language assessment (DLA) methods that implement a 3-stage process: 1) administration of the assessment; 2) provision of feedback to learners, and 3) further individualized instruction based on the assessment results. Therefore, the purpose of a

DLA is to assess learners' strengths and weaknesses, and provide tasks to help them improve. Practical DLA implementation faces three key challenges: 1) designing DLAs suitable for different learner skills; 2) identification of the most effective types of feedback for learners (Lee, 2015), and 3) efficient DLA administration (Poehner, 2008). One possible way to achieve efficiency is through computerization; however, typically this has meant learners receive only generic, rather than individualized, feedback (Harding, Alderson & Brunfaut, 2015).

2.2 English Email Writing and Pragmatic Competence

A faculty survey carried out at the higher education institution in which the current study is taking place has identified English email writing as a key task the undergraduate students should be able to perform (Kaneko et al., 2018). Upon graduating, this skill can also be beneficial for their future careers in a globalizing world. However, many students struggle with email writing in English. In particular, the pragmatic aspect of communication- in which social contextual variables such as the interlocutor's relative social status, social distance (how well the interlocutors know each other) and the potential imposition of the email communication upon the receiver (Brown & Levinson, 1987) can affect language choices related to politeness, formality and structure. This can be especially challenging for EFL learners (LoCastro, 2012). In addition, the pragmatic aspect of communicative competence is frequently undertaught in the EFL classroom; when it is addressed, it is often taught in an ineffective manner (McConarchy & Hata, 2013). While there are now various tools available that can help learners with the formal aspects of language such as grammatical difficulties (Grammarly or Google Translate for example), there is little help available for the pragmatic element of email communication. Further, while recently there have been efforts to develop DLA for reading and listening comprehension (Poehner et al., 2015; Yang & Qian, 2019), little DLA-related research has been conducted investigating writing ability. Specifically, applying DLA principles to communicative writing in varying social contexts, such as in email writing, has yet to be explored.

2.3 DLA and EFL email writing

We suggest focusing on two elements of this problem with pragmatic competence: 1) identifying precisely which pragmatic aspects of email writing learners find problematic; and 2) helping the learners improve their performance. Within the context of classes with large enrollments, the only feasible way to provide assessment and feedback is to implement a computerized DLA system; further, it should be able to provide individualized, specific feedback. However, little research has been done on developing a DLA of email writing skills that can help address problems students have with adapting their language choices to suit different social situations.

3. Developing a Basic DLA Prototype for EFL Learners' Email Writing

3.1 Context

As part of a larger study of DLA, Japanese EFL learners and email writing, the development of a basic computerized DLA prototype is one of the central tasks of the project. Figure 1 shows the overall development process for the DLA. The development of this prototype is the first part of the fourth step, Build DLA.



Figure 43: Core phases of the DLA development process

3.2 Rationale for Rapid Prototyping

Development of full-working versions of complex software packages is costly in terms of time, labor and finance. Prototypes or mock-ups are frequently used to gauge requirements, test design ideas and elicit feedback from stakeholders. Many organizations develop prototypes that can be trialed with stakeholders before starting work on the actual code. This enables developers to test the depth and breadth of their understanding of the requirements prior to developing fully-fledged code (Käpyaho & Kauppinen, 2015).

According to Tate (2010, n.p.), there is a “continuum of fidelity” that needs to be considered, starting from sketching concepts on paper to visual mock-ups and on to functional prototypes. The two most common types of rapid prototypes are low-fidelity and high-fidelity. Low-fidelity prototypes are very quick to create, but may lack either functionality or design, while high-fidelity prototypes appear professional and are fully functional for a number of use cases. High-fidelity prototypes aim to mimic the functions that users experience. For example, when creating a simple submission form, event handlers, such as mouseover and onclick need to be replicated. This is achieved by creating states and using interactions to switch between them, thus mimicking website behavior. The main downside to high-fidelity prototypes is the amount of time needed to create complex interactions.

Rapid prototyping was selected to create a high-fidelity prototype of the DLA system due to it being fast and economical (Chasanidou, Gasparini, & Lee, 2015). There are three key advantages to rapid prototyping: the early discovery of issues, its use as a communication aid, and its ability to test user interfaces and user experience.

- **Early discovery of issues**
Low-fidelity prototypes can enable developers to gain insight into basic issues much earlier. High-fidelity prototypes force stakeholders to make decisions on more complex implementation issues earlier in the process.
- **Communication aid**
Another advantage of rapid prototyping is that it is easy to show the software developer of the fully-functional version how the users are expected to interact with the interface. This can save time and reduce miscommunication between the developers and the clients. The developers can see how the clients expect users to interact using the prototype, and then aim to re-create the same feature and/or functions in the actual version. In short, high-fidelity prototypes provide developers with “living specifications” and enable showing rather than telling.
- **Tool to gain insight into the user-interface (UI) and user experience (UX)**
This initial test of the graphical user interface (GUI) can help identify aspects of the user experience that can be enhanced. Multiple aspects of the user interface need to be considered, including layout, theme and colour. UX focuses on the whole experience of using the website. Simply put, UI focuses on what the user sees while UX focuses on what the user feels. The instrumental, experiential and emotional aspects of a GUI combine together to create the UX (Hassenzahl & Tractinsky, 2006). UI can be visualized through sketches on paper and stakeholder feedback gained on those sketches, but it is difficult for users to imagine how they might feel when using a website if they cannot actually use it. With a high-fidelity prototype, users can provide detailed feedback on UX.

3.3 Procedure for Rapid Prototyping

Use case models were created to envisage how actors will interact with the DLA. Table 1 shows some of the use cases considered in the development of the rapid prototype. A primary persona specification was created to represent a typical user.

Table 1
Sample use cases for the Rapid DLA Prototype

Use Case	Actor	Task
1	Student	A new user registers to use the DLA and logs off
2	Student	An existing user takes a DLA task
3	Student	An existing user receives a report after the fifth DLA task

Based on the use case models and discussions with stakeholders, a brief requirements analysis audit was conducted and a software requirements specification (SRS) was drawn up. The required functionalities were then listed. The prototype was developed based on the SRS and the list of required functionalities, taking into account the primary persona specification.

3.4 Prototyping software: Axure RP 9

Axure RP 9, a desktop prototyping tool with a built-in cloud-based shared facility, was selected based on its availability, suitability and ease of use. Axure is one of the most popular prototyping software programs (Carter & Hundhausen, 2010). According to Axure Software Solutions Inc. (2016, n.p.), “86% of the Fortune 100” companies create prototypes with their software and over a million prototypes are shared on their cloud-based platform. Axure RP has an intuitive GUI (see Figure 2).

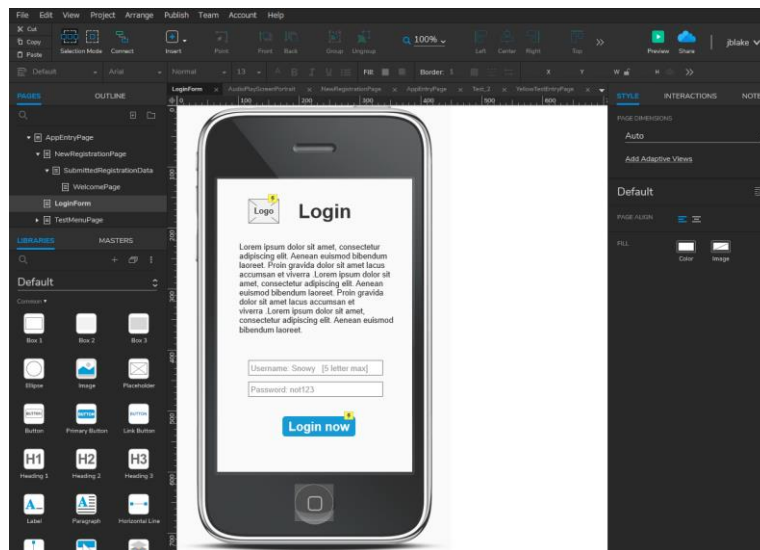


Figure 2. The user-interface for Axure RP showing a login screen

Creating a low-fidelity prototype of the DLA that can run in standard website browsers is straightforward when using this GUI. Figure 3 shows a screenshot of the login page in the high-fidelity prototype of the DLA created using Axure RP. The screenshot is taken from the use case in which the learner attempts the first DLA task. At this stage of this project, the nature of the tasks, expected pragmatic errors and content of the feedback are unknown. Placeholders of dummy text are used to provide a visual representation of the layout and quantity of text as seen through differently sized viewports on different devices.

4. Lessons learned

As with any new technology, novice users climb a learning curve. Five lessons were learned by the developer during the prototype design, development and initial *ad hoc* usability trials. The lessons described below aim to raise awareness of issues for other educators who may be considering prototyping their technology-enhanced language learning projects using Axure RP.



Diagnostic Language Assessment

User: Taro Suzuki

logged in



Task 1

Read the task carefully.

You work for a small computer supply company. You started work at the company two weeks ago. You would like to take a day off work next week. You need permission from your manager. Write an email requesting permission. Give the reason and a specific date.

Write your email below .

Figure 3. The prototype DLA GUI for task one created in Axure RP

4.1 One technology

When creating web-based tools, it is necessary to switch between various technologies, languages and libraries. For example, websites may use html, css, JavaScript and JQuery. PHP might be used to interact with a server and MySQL for a database. In Axure, however, the developer does not need to switch between different web technologies, reducing the need for mastering multiple technologies and speeding up the prototyping process.

4.2 Rapid prototyping is a misnomer

Low-fidelity rapid prototypes are aptly named, but high-fidelity rapid prototyping may take a considerable amount of time. The cost-benefit trade-off needs to be seriously considered if trying to replicate complex events. For example, the DLA system is required to automatically identify pragmatic errors that users make when writing in a submission form. To mimic this for multiple scenarios, it was necessary to create multiple regular expressions and make extensive use of states and interactions. A low-fidelity version, however, might only work for one use case which the user needs to input. Despite the apparent misnomer, the development of the prototype is rapid when compared to creating the full-code version.

4.3 Incomplete software requirements specifications

Some requirements that were needed by the developer were ambiguous while others were completely absent in the SRS. Should the full code have been outsourced using the identical SRS, this would undoubtedly have resulted in additional time and labor costs.

4.4 Usability issues with DLA prototype GUI

During the early stage of development, one Japanese speaker tested the interface, but struggled to understand what to do. This was because the instructions and interface were written in English only. Given that the DLA aims to test the use of English, this may be acceptable. However, as other users may face the same problem, the necessity for a bilingual interface needs further consideration.

4.5 User experience

The initial feedback from users was that the interface felt like a test. Although the DLA is an online assessment system in which users complete tasks, the purpose is to evaluate their ability to write emails in English and then provide guidance based on their performance. Yet, the task looked more like a test rather than writing an email in a typical email client, such as Gmail, or Outlook Express. Features that Gmail affords, such as generated short responses, were absent. Users need to suspend disbelief and pretend that they are writing an email. Perhaps, it is viable to replicate an email client within the DLA.

5. Conclusion

The development of the DLA prototype was a steep learning curve for the developer, but was a very positive experience. Not only did the developer create a working prototype in a short timeframe, but there were many tangible outcomes that have a positive effect on the full-code version of the DLA. Most importantly the process forced numerous operational decisions to be made earlier than anticipated. Secondly, feedback on the UI and UX at such an early stage in this multi-year project means that resources can be dedicated earlier to improving these aspects. Thirdly, the prototype looks and behaves in a similar manner to the actual DLA and so can be used as an effective communication aid showing how the DLA works rather than describing how it should work. Therefore, it is suggested that the interactive nature of the prototype promotes understanding in a way that written descriptions or pictures cannot.

Acknowledgements

This research is supported by a JSPS KAKENHI (Grants-in-aid for Scientific Research), Grant Number 19K00826, “Developing a diagnostic language assessment of Japanese EFL learners' English language email writing”.

References

- Axure Software Solutions Inc. (2016, April 15). Axure RP 8 takes Prototyping & Diagramming to the Next Level. CISON PR Newswire.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Use*. Cambridge: Cambridge University Press.
- Carter, A.S., & Hundhausen, C. D. (2010). How is User Interface Prototyping Really Done in Practice? A Survey of User Interface Designers. *Proceedings of 2010 IEEE Symposium on Visual Languages and Human-Centric Computing*, 38, 207-211. <https://doi:10.1109/vlhcc.2010.36>
- Chasanidou, D., Gasparini, A. A., & Lee, E. (2015, August). Design thinking methods and tools for innovation. In *International Conference of Design, User Experience, and Usability* (pp. 12-23). Springer, Cham.
- Harding, L., Alderson, J.C., Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing* 32(3), 317-336.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience – a research agenda. *Behaviour & Information Technology*, 25(2), 91-97. <https://doi.org/10.1080/01449290500330331>
- Kaneko, E., Park, M., Wilson, I., Roy, D., Heo, Y., Yasuta, T., Blake, J., Nicholas, A. (2018). English curriculum innovation for computer science majors in the Japanese EFL context: From needs to tasks. *Proceedings of the 2018 IEEE International Professional Communication Conference*, 84-89.
- Käpyaho, M., & Kauppinen, M. (2015, August). Agile requirements engineering with prototyping: A case study. In *Proceedings of the 2015 IEEE 23rd International requirements engineering conference (RE)* (pp. 334-343). IEEE.
- Lee, Y-W. (2015). Diagnosing diagnostic language assessment. *Language Testing* 32(3), 299-316.
- LoCastro, V. (2012). *Pragmatics for Language Educators: A Sociolinguistic Perspective*. New York: Routledge.
- McConachy, T., & Hata, K. (2013). Addressing textbook representations of pragmatics and culture. *ELT Journal*, 67(3), 294–301.
- Poehner, M. (2008). *Diagnostic Assessment: A Vygotskian Approach to Understanding and Promoting L2 Development*. Berlin: Springer.
- Poehner, M., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337–357.
- Tate, T. (2010, July 29). Concerning fidelity in design. UX Booth. <https://www.uxbooth.com/articles/concerning-fidelity-and-design/>
- Yang, Y. & Qian, D. (2019). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, 1-25. <https://doi.org/10.1080/09588221.2019.1585882>

Effects of Learning Activities Based on Augmented Reality on Students' Understanding and Expression in an English Class

Yang YANG^a, Shumin WU^a, Dantong WANG^b, Yaohua HUANG^b & Su CAI^{a,c*}

^a*VR/AR+Education Lab, School of Educational Technology, Faculty of Education, Beijing Normal University, China*

^b*Tsinghua University Primary School, China*

^c*Beijing Advanced Innovation Center for Future Education, Beijing Normal University, China*

*caisu@bnu.edu.cn

Abstract: In English as foreign language(EFL) learning, students' understanding and expression of the context are very important. However, teachers usually meet difficulties in explaining scientific phenomenon with easily understandable English words when teaching text related to science. It leads students to mechanically understand and memorize according to their own imagination. In this study, a series of learning activities based on Augmented Reality(AR) technology was carried out to cope with this problem. A specially developed AR application created a virtual-reality combined environment for students to observe and interact with the sun and the earth in grade 4. Moreover, diverse research methods were used to analyze students' understanding about learning content and oral expression about learning topic. The experimental results reveals that learning activities based on AR can assist students in the construction of understanding and expression about learning contents. Furthermore, by interviewing the students, the mechanism of how AR worked in this class was reported. Students have more time to learn and communicate in AR-based activities. They acquire more knowledge by interacting with other people in light of learning materials, instead of passively listening to the teacher or reading books. In this regard, they get more authentic and profound learning experience to support better understanding and expression of learning contents. These findings can be valuable references for those who intend to implement learning activities based on AR or those who want to improve students' understanding and expression in EFL classes.

Keywords: augmented reality, understanding, expression, English as foreign language learning, EFL

1. Introduction

In recent years, Augmented Reality (AR) has gradually been closed to people's daily life and aroused widespread concern. In the field of education, mobile-based AR has been focused on by a crowd of researchers and recommended to teachers for teaching (Akçayir & Akçayir, 2017). With the adoption of AR in learning activities, it is assumed that primary school students will have better performance in motivation, confidence and related dimensions (Chiang, Yang, & Hwang, 2014; Han, Jo, Hyun, & So, 2015). Based on its advantages, AR has been regarded as a breakthrough in the solution of some teaching difficulties, such as teaching of abstract concepts (Crandall et al., 2015).

In this study, we try to use AR in English as a foreign/second language (EFL) learning field. In a primary school in China, the EFL teachers had difficulties in creating attractive situations when giving lessons about an interdisciplinary integration content. According to AR's characteristics on enhancing physical real world by adding virtual computer-generated information to it, we presented a special situation for teachers and students through an AR application. And we did a new instructional design to contribute to a good integration of AR and courses at the classroom level.

2. Current states and related works

AR is a technology that supplements the real world with virtual objects and appears to coexist in the real world (Azuma et al., 2001; Rt, 1997). AR has been used in EFL learning for a long time.

During the early time, researchers used AR to directly afford models, audios or videos when learners catching right labels in real environment. In these years, AR was mainly used to teach the meaning and pronunciation of words (Amaia, Inigo, Jorge, & Enara, 2016; He, Ren, Zhu, Cai, & Chen, 2014). There were also some researchers taking notice of promoting learners to interact with special labels or the AR system (Barreira et al., 2012). These studies primarily showed AR's effects on improve students' motivation, satisfaction or attention. Further researches later confirmed AR's effectiveness in teaching children and attempted to use AR in higher education to improve learners' language skills (Safar, Al-Jafar, & Al-Yousefi, 2017). Richardson (2016) presented that AR can be used to arrange tasks for advanced level language learners to improve their motivation and learning performance. Yang and Mei (2018) founded that learners have positive perception and attitude towards the immersive language learning experience afforded by the AR programming.

In summary, previous studies have proved that AR used in EFL learning is able to ensure better attitude and performance through more interesting activities or immersive experience. However, almost all of these studies focus on informal learning neglecting the learning difficulties happen in class. Wu, Lee, Chang, and Liang (2013) suggested that researchers need to explore the possibilities and solutions of integrating AR into regular school curricula. Therefore, we hope this study can show AR's effects on solving teaching difficulties in a formal lesson based on this technology's features and previous researches' results.

3. Method

The main aim of this study is to test the effects of learning activities. On the basis of a well-designed AR application, students' understanding and expression in a regular 50-minute EFL class was recorded and analyzed. A pre-test and a post-test were carried out to trace students' change. And interviews with students were conducted at the next step to investigate how AR assists students in class.

3.1 Research Questions

Based on literature review and our prediction, the research questions of this study are presented below.

Q1: Can leaning activities based on AR benefit student' understanding in regular EFL class?

Q2: Can leaning activities based on AR benefit student' expression in regular EFL class?

Q3: How does AR assist students learning in regular EFL class?

3.2 Participants

92 students in grade 4 and their English teacher in Beijing, the capital of China, participated in this study. These students and their teacher all had experience of using mobile devices in class before this study. And these students experienced other AR applications before this class to learn how to use camera to observe and interact with virtual objects in an AR environment with location-based technology and multisensory technology.

3.3 Instruments

The pre-test consisting of 5 short-answer questions aims to evaluate students' prior knowledge of sun and earth. The first three questions are used to test students' English vocabulary and expression skills about the sun. And the last two questions are used to test students' cognition about cycles of day and night as well as four seasons. The last two questions are also used as post-test to acquire students' conceptual changes after class. Two teachers were required to score the last two questions with standards below in Table 12:

Table 12*The Standards for Students' Cognition about the Sun and the Earth*

Score	Standard
1	Blank
2	Response containing error messages
3	Response containing irrelevant messages
4	Response containing incomplete messages
5	Response containing complete messages

Semi-structured interviews were used to examine if and how AR helped students to learn in students' own view. The outline of the interviews is shown below.

- What activity do you think is particularly interesting in class?
- Do you remember how we saw the sun and the earth in class?
- You learned through AR today, and is there any difference from this class and your previous English classes?
- Could you tell me what abilities you have improved in class?
- Do you want to continue to use AR to learn in the future?
- How do you think AR can be better used in this class?

4. AR Application Design

In this study, the application containing three main scenes is designed and implemented. The mobile-based AR application, called the Sun and the Earth, which can run on Android devices with cameras and gyroscopes. This application is developed based on Unity 3D engine, and Vuforia SDK is used to link virtual objects with the real environment. This application contains three main scenes: (1) See the Sun, (2) Day and Night, and (3) Four Seasons. These scenes are designed and implemented to make learning activities in class more vivid and interactive.

In the first scene, users can see the solar system around by lifting the device to different directions with gyroscope. There is a thermometer on the right side of the screen. It gets redder and taller when users make the sun bigger. And if the sun becomes big enough, the thermometer will explode and a voice will say so hot.

In the second scene, users can see the sun and the earth from a side-view, so that they can directly observe the cycles of day and night on earth. Users can start or stop the earth's rotation by clicking on the earth and feel about the cycles of day and night directly by observing the girl in pink and the boy in blue standing in different hemispheres on the earth.

In the third scene, users can see the earth's revolution around the sun from an overlooking perspective. With gyroscope in device, users can see earth in different seasons when facing different directions. There is a tree in the northern hemisphere showing the changing process of four seasons clearly, and a text message at the top of the screen present the season information. Users can start or stop the changing process by clicking on the earth.

5. Instructional Design

In this study, an English reading lesson for fourth grade students, named "The Sun", containing some learning activities with mobile-based AR applications is designed. This class should be an extended lesson which is implemented after students have learned basic vocabulary to describe the sun.

The basic goal of this class is to introduce the sun and its influence on the earth logically in English. Students dialogues will be carried out to help students understand the content better and perfect their oral expression. Students should learn at their own pace in dialogues, and they also develop the autonomous learning ability through the chatting activities. However, the sun and the earth are huge objective things that can hardly be observed and controlled by humans. The use of AR offers contextual support for students' dialogues by presenting corresponding phenomenon directly and allowing interaction with visual objects.

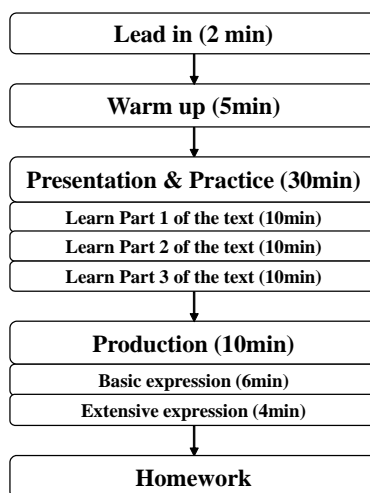


Figure 44. Structure of This Class

The basic structure of this class is shown in Figure 44. In this class, the AR application is used in leading in and learning about Part 2 and Part 3 of the text. Part 2 claims why earth has day and night and Part 3 states why earth has four seasons. These are difficult parts to understand and express in the text. In warming up, teacher first asks students to guess the sun according to some descriptions and enjoy a song about the sun to review some related words. Then they use AR to see the sun closer to intuitively feel the heat of the sun.

AR in learning activity of Part 2 help students practice these sentences: (1) Earth takes one day to spin around, (2) Yao Li is in the day, (3) Mike is at night, (4) China faces the sun, and (5) America faces away the sun. AR in learning activity of Part 3 help students to act as the sun and the earth and practice these sentence: (1) You take one year to go around me, and (2) It's spring/summer/fall/winter. After finishing the group work, volunteers can present in class.

6. Results

6.1 Students' understanding

One objective of this study is to examine the effectiveness of learning activities based on AR in terms of students' understanding in a regular EFL class. All the students were required to take the pre-test before class and post-test after class, but some of them didn't finish both tests for personal factors. Basically, the number of students is 73. By analyzing the changing process of students' cognition towards why we have day and night or four seasons. The *t*-test results show that students understand the content significantly better after the class as shown in Table 13.

Table 13

t-test Results of the Students' Understanding of Contents and Score Gap between 2 Parts

Item	Mean	SD	t	N
Day & Night	pre-test	3.14	-7.50***	73
	post-test	4.25		
Four seasons	pre-test	1.64	-10.62***	
	post-test	3.49		

(*** $p < 0.01$)

In terms of why we have day and night, the means of pre-test and post-test are 3.14 and 4.25, as SD's are 1.35 and 1.06. There are 43.8% and 84.9% of the students getting more than 4 score in pre-test and post-test. The *t*-test result also confirms that there are significant differences between their scores in pre-test and post-test. This implies that the students get better understanding about the sun gives us day and night through the learning activity based on AR.

In terms of why we have four seasons, the means of pre-test and post-test are 1.64 and 3.49, as SD's are 1.20 and 1.00. There are 12.3% and 67.1% of the students getting more than 4 score in pre-test and

post-test. The *t*-test result also shows there are significant differences between their scores in pre-test and post-test. These results meant learning activity based on AR that related to why we have four seasons help students construct their understanding effectively. However, some messages might not be presented clearly enough, so students tried to explain the phenomenon with some incorrect or irrelevant information. If the messages about the formation of seasons are presented more clearly, students will have better performance. It also inspires us the amount and clarity of information in an AR application is very important for the effects of using it.

6.2 *Students' expression*

3 questions in pre-test were used to test students' vocabulary and expression ability before class. Most of the students are able to describe the sun in terms of shape and heat. The common answer is that the sun was big and hot. The color of the sun is also a factor that many students talked about. Some students said the sun was like a ball. Few students referred to the relationship between the sun and the earth. Even if the earth appeared in students' answers, it was usually presented in a comparison of size or heat. So the students' expressions before class were simple and illogical.

Formative evaluation was used to test students' expression in class. Almost all of the students can introduce the sun in a comprehensive and logical way as the activities going on. Students were required to introduce the sun with following sentences: (1) It is a star, (2) It is very big, (3) It is made of superhot gas, (4) It can give heat and light, (5) It can give day and night, (6) It can give four seasons, and (7) It is the most important part of our solar system. Students practiced saying these sentences for several times and were checked by their partner in group and teacher in class. Some of the students could explain why the sun give day and night as well as four seasons. They can say earth takes one day to spin around and earth takes one year to go round the sun.

The improvement in students' expression was established in learning activities based on AR. One of the teaching difficulties in this class was to break and reconstruct students' understanding of the sun. The presentation and interaction with AR helps students get better understanding, and then it becomes much easier for them to expand own expression about the sun.

6.3 *How did AR work*

5 students performing well in class were interviewed according to the semi-structured outline to show how AR assisted students to learn in this EFL class. Students thought the visual presentation of AR brought about more authentic experience and more frequent interaction with others. The use of AR extended the teaching space from the podium to the entire classroom. As for abilities, they thought AR benefits in the improvement of collaboration ability, inquiry ability and autonomous learning ability. They thought there was a clearer division of work in the activities based on AR. And they were able to discovery some new knowledge when they use AR to learn. In summary, students will have more time to learn and communicate by using AR in class. In this way, they can get more authentic and rich learning experience to support better understanding and expression of learning contents.

7. Conclusion

The experimental results in this study revealed that learning activities based on AR was able to assist in the construction of understanding and expression about learning contents. Furthermore, by interviewing the students, the mechanism of how AR worked in this class was reported. Students would have more time to learn and communicate in AR-based activities. They needed to acquire knowledge by interacting with learning material and other people instead of just listening to the teacher or reading books. In this way, they get better understanding and expression of learning contents.

This study proves that AR takes advantages in EFL learning not only in informal learning but also in formal learning. First, AR can let learning experience be more authentic and interesting, which improve students' positive attitude and motivation(Amaia et al., 2016; He et al., 2014; Yang & Mei, 2018). Second, complex language skills are also improved in learning activities based on AR(Liu & Tsai, 2013; Safar et al., 2017). This study found the authentic presentation and frequent interaction are the

cores of using AR in class. Only when the use of AR enriches the learning environment and changes students learning style to autonomous learning, students can really improve their abilities through the use of the technology.

The results of analyzing in this study also inspire us to notice the amount and clarity of knowledge in an AR application. When we design and develop the AR applications, we choose what objects and relationships will be presented in them. Proper presentation may help students pay attention to knowledge that they ignore in daily life. And inappropriate presentation may bring misunderstanding of the learning content. It's worthy to be further discussed how to choose and present knowledge in AR applications.

Acknowledgements

Our work is supported by the National Natural Science Foundation of China (61602043) and 2019 Comprehensive Discipline Construction Fund of Faculty of Education, International Joint Research Project of Faculty of Education, Beijing Normal University.

References

- Akçayir, M., & Akçayir, G. (2017). Advantages and challenges associated with augmented reality crossMark for education: A systematic review of the literature. *Educational Research Review*, 20, 1-11. doi:10.1016/j.edurev.2016.11.002
- Amaia, A. M., Inigo, A. L., Jorge, R. L. B., & Enara, A. G. (2016). *Leihoa: A window to augmented reality in early childhood education*. Paper presented at the International Symposium on Computers in Education.
- Azuma, R., Baillet, Y., Behringer, R., Feiner, S., Julier, S., & Macintyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6), 34-47. doi:10.1109/38.963459
- Barreira, J., Bessa, M., Pereira, L. C., Adao, T., Peres, E., & Magalhaes, L. (2012). *MOW: Augmented Reality game to learn words in different languages: Case study: Learning English names of animals in elementary school*. Paper presented at the Information Systems & Technologies.
- Chiang, T. H. C., Yang, S. J. H., & Hwang, G. J. (2014). An Augmented Reality-based Mobile Learning System to Improve Students' Learning Achievements and Motivations in Natural Science Inquiry Activities. *Educational Technology & Society*, 17(4), 352-365.
- Crandall, P. G., Engler, R. K., Beck, D. E., Killian, S. A., O'Bryan, C. A., Jarvis, N., & Clausen, E. (2015). Development of an Augmented Reality Game to Teach Abstract Concepts in Food Chemistry. *Journal of Food Science Education*, 14(1), 18-23. doi:10.1111/1541-4329.12048
- Han, J., Jo, M., Hyun, E., & So, H. J. (2015). Examining young children's perception toward augmented reality-infused dramatic play. *Etr&D-Educational Technology Research and Development*, 63(3), 455-474. doi:10.1007/s11423-015-9374-9
- He, J., Ren, J., Zhu, G., Cai, S., & Chen, G. (2014). *Mobile-Based AR Application Helps to Promote EFL Children's Vocabulary Study*. Paper presented at the IEEE International Conference on Advanced Learning Technologies.
- Liu, P. H. E., & Tsai, M. K. (2013). Using augmented-reality-based mobile learning material in EFL English composition: An exploratory case study. *British Journal of Educational Technology*, 44(1), E1-E4. doi:10.1111/j.1467-8535.2012.01302.x
- Richardson, D. (2016). Exploring the Potential of a Location Based Augmented Reality Game for Language Learning. *International Journal of Game-Based Learning*, 6(3), 34-49. doi:10.4018/ijgbl.2016070103
- Rt, A. (1997). A survey of augmented reality. *Presence*, 6(4), 355-385.
- Safar, A. H., Al-Jafar, A. A., & Al-Yousefi, Z. H. (2017). The Effectiveness of Using Augmented Reality Apps in Teaching the English Alphabet to Kindergarten Children: A Case Study in the State of Kuwait. *Eurasia Journal of Mathematics Science and Technology Education*, 13(2), 417-440. doi:10.12973/eurasia.2017.00624a
- Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41-49. doi:10.1016/j.compedu.2012.10.024
- Yang, S., & Mei, B. (2018). Understanding learners' use of augmented reality in language learning: insights from a case study. *Journal of Education for Teaching*, 44(4), 511-513. doi:10.1080/02607476.2018.1450937

Developing an Integrated system of Robots and Toys with Internet of Things for Children's Language Development

Ya-Wen CHENG^a, Yuping WANG^b, Ming-Rong Liang^a, Nian-Shing CHEN^c & Ya-Fei YANG^{d*}

^a *Department of Information Management, National Sun Yat-sen University, Taiwan*

^b *School of Humanities, Languages and Social Science, Griffith University, Australia*

^c *Department of Applied Foreign Languages, National Yunlin University of Science and Technology, Taiwan*

^d *Department of Digital Design and Information Management, Chung Hwa University of Medical Technology, Taiwan*

*cillia55@gmail.com

Abstract: This study aims to create an interactive and immersive language learning environment for children with the support of robots and IoT sensors. In this paper, we first presented our key design principles for a system that integrates robots and IoT sensors into children's learning of a language while playing. The intergraded system was developed following the procedure of requirement analysis, prototyping and system testing and evaluation. Following the identification of learning needs and system requirements, a prototype of the system was built and evaluated using the cognitive walkthrough method. Although some usability issues were reported by the participants, the overall results confirm that learning tasks were successfully completed using this integrated system, and it is an easy-to-use system in that non-technical professionals were able to design scripts to control the robot and the IoT sensors.

Keywords: robot, IoT toys, robot for language learning, script editor

1. Introduction

Second language (L2) and/or foreign language learning has been driven by many theories and approaches such as the communicative, interactionist and intercultural approaches. With the adoption of technology in language learning, the validity of these theories and approaches have been further confirmed. At the same time, learning a second or foreign language through such a conscious learning process (Krashen, 1982; Schmidt, 1983) has also been criticized for their lack of effectiveness and native-like proficiency on the part of the learner in comparison with native language acquisition (Chang, Lee, Chao, Wang, & Chen, 2010; Pot, Monceaux, Gelin, & Maisonnier, 2009; Valk, 2010). Consequently, the advantages of acquiring multiple native languages at an early age in an immersive multilingual environment have received much academic attention. However, the reality is that most children lack access to such an environment during their critical period of language development. The current research aims to design such an immersive learning environment for children, taking advantage of the robotic technology and Internet of Things (IoT).

Robot-Assisted Language Learning (RALL) was introduced in early 2000 to facilitate learning foreign languages. Studies have shown that incorporating robots in young children's education may result in various benefits in terms of children's cognitive and behavioral development as well as language learning performance (Kanero et al., 2018; Kuhl, 2011; Mathur & Reichling, 2016; Morse & Cangelosi, 2017; Vogt, De Haas, De Jong, Baxter, & Kraemer, 2017). With the help of sensors, robots are able to perceive and react to the surrounding environments, providing proper responses through sounds or emotional expressions. In addition to affording anthropomorphic interactions, it is also an easy way to record and track the learning process. However, the incorporation of robots and sensors in language education can present various challenges to educators who do not have a technical background.

This study aims to design an easy-to-use system for language professionals who want to create an interactive and immersive language learning environment using robots and IoT sensors. To develop such a system is the end goal of the current study. Thus this study is guided by the following two research questions: 1) what are the design principles for developing an easy-to-use integrated system for non-technical background professionals to incorporate robots and IoT sensors in language education? and 2) what are the usability issues of such a system that system designers should be aware of?

2. Literature Review

2.1 Robot Assisted Language Learning (RALL)

Social robots, providing human like interaction through sounds and gestures have made great contributions to early language learning (Kennedy, Baxter, & Belpaeme, 2015; Vogt et al., 2017). Incorporating robots in language education is found to have positive effects on enhancing students' willingness to learn, students' learning motivation and learning performance (Kanero et al., 2018; Kuhl, 2011; Meltzoff et al., 2009; Morse & Cangelosi, 2017). Due to its humanoid features such as the appealing appearance, gestures and movements, a robot'' interaction with children can result in many positive effects on learning outcomes, such as increased vocabulary and willingness to speak in English (Alemi, Meghdari, & Ghazisaedy, 2015). However, most of the existing research into the use of robots in language learning only concern with the use of individual robots in a certain aspect of language learning in the context of traditional approaches (e.g., communicative approach) to L2 or foreign language learning. Different from the existing studies, our research approaches the use of robots from the perspective of language learning through immersion, with an emphasis on using the target language as the medium of communication and interaction between robots and children, much like the learning of one's mother tongue. Thus this research aims to create an immersive language learning environment for children, supported not only by robots but also by IoT sensors.

2.2 Script editor

Many studies have devoted to the simplification of programming for novice users. Visual programming language provides an interface that allows the user to create programs by manipulating program elements graphically rather than by specifying them textually. With the help the graphical display of programming elements, users are able to contemplate, reason and learn to program (Chao, 2016). In fact, many educational robots were designed to teach programming adopting the graphical approach, e.g., LEGO Mindstorms (Valk, 2010), Choregraphe for NAO (Pot et al., 2009), EV3 programmer app, Zenbo app builder, and scratch for programming Arduino robot. However, these apps were developed to support the teaching of programming, in which the users are expected to have basic prior knowledge of programming and to acquire more advanced programming skills after using the programming applications. Form-based user interface is widely adopted in the apps for mobile devices and information systems. Similar to graphical programming, form-based application provides a more intuitive user interface, through which users can accomplish tasks by simply using the controls in the form, e.g., dropdown buttons. These types of tasks do not require prior knowledge of programming. Informed by this technique, we adopted Windows Forms (WinForms) to develop a form-based script editor for non-technical professionals to incorporate robots in education.

3. Method

3.1 Research Design

The proposed system was developed through a non-linear and iterative process, which sought to identify users' needs and to develop the system through iterative cycles of testing and refinements. The process consists of three major phases: learning needs and system requirement analysis, prototyping, and testing and evaluation. Phase 1 was to understand users, to identify critical problems and to propose solutions to the users' needs. To this end, focus group discussions were held with the participation of two TELL (Technology Assisted Language Learning) professionals and the research team consisting of robot experts and curriculum designers. The two TELL professionals were invited because of their strong interest in incorporating robots and IoT sensors in language learning. Two themes characterized these discussions: 1) essential elements that are critical for children's language learning through playing

with robots, and 2) the requirements of an integrated system that allows non-technical professionals to manage robots and IoT sensors to meet their teaching needs. The recordings of these discussions were coded and themes were identified to inform our needs and requirement analysis. This analysis led to the creation of the specifications of the proposed system. Phase 2 was to develop a prototype according to the specifications identified in Phase 1. To facilitate the iterative cycles of testing and refinements, we used 3D printing to build a prototype of a robot and used a plastic box to build a prototype of a toy. Phase 3 was to test the system and to identify usability issues in the early developmental stage. This study adopted the Cognitive Walkthrough method to evaluate how easy it is for new users to accomplish tasks using the system we developed. Cognitive Walkthrough is a usability evaluation method in evaluators work through a series of tasks to identify system usability issues. We invited four non-technical professionals and parents, who are interested in using robots for children’s language education, to participate in the testing of the system. The participants were first given a 10 minutes introduction of the system and the tasks they were to perform. The participants were then asked to design a script to specify interactions between the robot and the IoT sensors using the system we developed. At the end, the participants were to report the usability problems they encountered and give a severity score to the problems.

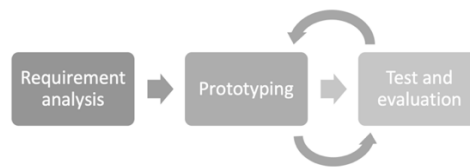


Figure 1. The developmental process of the proposed system.

3.2 The specifications of pedagogical needs and system requirements

Informed by the results from the focus group discussions, the following four specifications of pedagogical needs were identified. First, a toy should be used as the focus of the joint attention of the robot and children. That is, the robot was to play a toy with the children and provide linguistic input and output as parents would do. Second, the core functionality of the robot should be to provide linguistic cues, instead of performing the real tasks, whereas the tasks should be performed by the children. Third, toys should be included to better engage the children and help develop their cognitive skills while playing and receiving linguistic exposure. Fourth, the toy should be meaningful to the children while carrying multiple learning contents and functions to maintain children’s engagement.

The focus group discussions also led to the inclusion of four key system requirements. The first requirement is the development of an easy-to-use script management system that allows users to program desired interactions between the robot, the IoT sensors and the children. As children’s language cognition progresses rapidly with age, the learning contents should also be updated regularly and timely. It is thus crucial that the system allows language professionals to adjust the learning content in a timely manner in accordance with the developmental needs of the children. Second, a non-text programming environment, e.g., a graphical or iconic programming environment, is also deemed necessary to lower the technical barriers of incorporating robots in language education for language educators. Third, toys with IoT sensors should be able to be used together with the robot or independently from the robot. Fourth, scalability should also be an essential requirement for this system, allowing more sensors and modules to be included. This will ensure that the system is flexible enough to provide educators with rooms for catering to different children’s individual needs, and for updating and creating learning contents for different playing scenarios and learning environments.

4. Results and Discussion

4.1.1 The Design of the Robot and Toys

This study used Raspberry Pi 3 as the hardware for the development of robots and toys. Owing high computing power and low power consumption wireless transmission devices, such as on-board WiFi and Bluetooth, Pi 3 is able to work well with IoT and is capable of adding on sensors. As the goal of the system was to facilitate children’s language learning, providing contextual linguistic cues was considered as the first priority requirement of the system. Therefore, context awareness and providing corresponding linguistic information including speaking and facial cues were considered two key

system requirements. Based on these considerations, a prototype of a robot was designed as shown in Figure 3. It was a light and portable robot that kids could move easily. The robot was equipped with a web camera to monitor children's motions, a microphone to receive voice commands and a screen to display learning materials and facial expressions. In order to increase the scalability, a breadboard was embedded inside of the robot, allowing more sensors to be added on to receive contextual information such as humidity, temperature, location etc. In order to maintain cost-effectiveness and quick turnaround of the iterations of usability tests and refinements, we used a 3D printer to build a prototype of the robot. The inner structure of the robot was designed to address the concerns of safety, a power supply and overheating. The robot was composed of the following components: a Raspberry Pi, a web camera with a microphone, a speaker, a 7 inches screen, a power bank, a breadboard and a humidity sensor. An expression management sub-system was developed to control the robot's facial expressions. In addition, the system also allowed the user to monitor the system's status, such as being disconnected with the cloud server and no script running at a specific moment, by observing whether the robot was displaying the animation of being confused or sleepy.



Figure 3. The design of the language learning robot and IoT toy.

The toy was designed as an interactive toy box, providing interactions through pressing the flashcards on the side of the toy box. Whenever the child pressed the flashcards on the side, either the toy or the robot would generate instant audio feedback to interact with the child, so as to engage he or she in playing. The toy was designed using a touch sensor board that connected to 12 objects, allowing children to learn through exploring nearby objects and to practice motor skills while listening and speaking the target language with the robot. Different from existing interactive sound books, of which the learning content is fixed and irreplaceable, this toy box was designed to provide flexibility to accommodate the needs of replacing learning contents whenever needed. Another reason for the inclusion of toys was to provide a focal object that drew the attention from both the robot and children, while the robot played toys with children in order to facilitate children's language development in an interactive and immersive environment. Scenarios were created using the toys as a stimulus for the robot to engage a conversation with children in the target language. With the help of the sensors on the toy, the robot was able to detect what children were doing and prompt instant feedback as parents would do when playing with their children.

4.1.2 The Design of Script Editing System

The script management sub-system provided the following four functionalities: allowing users to create and edit learning materials, to manage scripts, to view and search log files and to adjust system configurations. As shown in Figure 6, the script management sub-system was designed with a graphical programming interface, including filling-in-blank boxes and drop-down navigation bars. It also allowed users to perform different actions by selecting pages on the top. With these graphical programming features, this sub-system could ease users' cognitive loads. The first component of this sub-system was a content management sub-system, enabling users to create and edit learning materials by uploading pre-recorded mp3 files or simply applying the text to speech function to generate a spoken sentence. It also allowed users to delete or to modify the learning materials they created. The second component was a script management sub-system. It allowed the educators to edit or delete a script that they created. The third component was a log file management sub-system. It allowed users to view, search, sort and download the log files. The fourth component was system configurations, which allowed users to adjust

system settings, e.g. the Internet access setting. This study also developed a graphical script editor. As shown in Figure 4, the script editor adopted Windows Forms, allowing users to program a robot and a set of touch sensors by clicking a few buttons. In order to generate multiple interactions using the 12 sensors, we designed two interaction modules to enrich the variety of interactions: a single sensor module and a multiple sensor module, allowing users to design the responses by specifying a set of sensors.

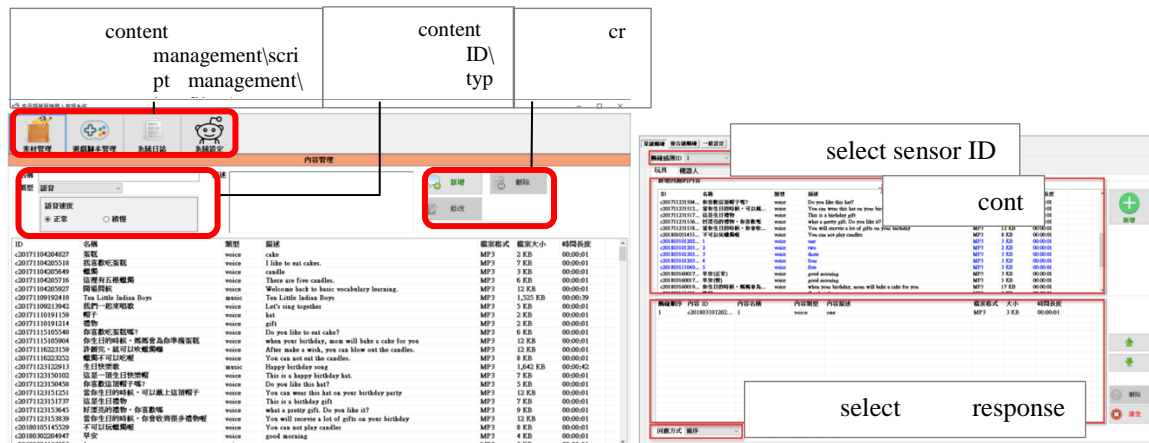


Figure 4. The interface of script management sub-system and script editor.

4.2 Preliminary Result of System Usability Evaluation

This study conducted a usability evaluation using the Cognitive Walkthrough method. The preliminary results from four participants are presented below. All of the participants, assuming the role of the end users, were able to complete the given task using the integrated system. 10 usability problems were reported, as shown in Table 1. The usability problems were categorized into six categories, including poor hardware design, system scalability, error prevention and recovery, system learnability, visibility of system status and efficiency of use. The participants also made two suggestions to improve the system, speech control and scheduling. With regard to affective speech, since the system was designed for children's language education, the TTS function provided by the system should be able to adjust the tones and speech speed of the speaker in order to effectively engage children's attention. As far as scheduling is concerned, the participants suggested that the system should include a scheduling function, e.g. executing good morning script before 9 am, and bedtime story script after 9 pm.

Table 1

The List of Usability Problems

Problem description	Numbers of participants	Average severity (1~5)	Problem Category
Bad connections	4	4	Poor hardware design
The way of creating and editing a script is not intuitive	3	3.7	System Learnability
Nonstop playing learning contents when the same button was mistakenly pressed.	2	3.5	Error prevention and recovery
Complex WiFi setting	4	3.25	System Learnability
Limited sensor slots	2	3	System scalability
Unable to add new functions to the script editing system	4	3	System scalability
Unable to adjust the sound volume of the toy and the robot from hardware	2	2.5	Poor hardware design
No information about the cause of system crash	3	2.3	Error prevention and recovery
Unable to observe battery status	2	2	Visibility of system

		status
Lack of accelerators to indicate which learning content has been selected or used	1 2	Efficiency of use

5. Conclusion

This study has discussed some critical issues relating to the design of creating an interactive and immersive language learning environment using robots and IoT sensors. The pedagogical concern of this study is to create a play scenario for children to learn a language through playing toys with a robot. To achieve this goal, we developed an intergraded system for non-technical professionals to control and specify interactions between robots and IoT sensors. In this preliminary study, we presented the design of the system architecture, system components and the employment of a graphic programming editor. As presented in this study, the system can be used in various learning scenarios, allowing educators and researchers to incorporate robots and IoT sensors in education effortlessly for language studies.

This study reports the preliminary results of an ongoing project with the aim to identify usability problems in the early developmental stage of the system, which will be addressed in our next round of design and improvement. We recognize that the complexity of designing an easy-to-use integrated system for children's cognitive and language development requires a systematic and iterative approach to the evaluation and improvements of the system design. We are planning to solve the identified usability issues and conduct more evaluations to improve the design of the system. Eventually the system will be evaluated in an authentic learning environment in which children will be invited to play and interact with the robot and the IoT toys, with the aim to develop a more comprehensive evaluation taxonomy for an intergraded systems of robots and IoT devices.

Acknowledgements

This research was supported by the National Science Council, Taiwan under project numbers 108-2511-H-224 -008 -MY3, MOST-107-2511-H-224-007-MY3 and MOST-106-2511-S-224-005-MY3.

References

- Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, 7(4), 523-535.
- Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y., & Chen, G.-D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Journal of Educational Technology & Society*, 13(2), 13-24.
- Chao, P.-Y. (2016). Exploring students' computational practice, design and performance of problem-solving through a visual programming environment. *Computers & Education*, 95, 202-215.
- Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, 12(3), 146-151.
- Kennedy, J., Baxter, P., & Belpaeme, T. (2015). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, 7(2), 293-308.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*: Oxford: Pergamon.
- Kuhl, P. K. (2011). Social mechanisms in early language acquisition: Understanding integrated brain systems supporting language. *The Oxford handbook of social neuroscience*, 649-667.
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, 146, 22-32.
- Morse, A. F., & Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Science*, 41, 32-51.

- Pot, E., Monceaux, J., Gelin, R., & Maisonnier, B. (2009). *Choregraphe: a graphical tool for humanoid robot programming*. Paper presented at the RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication.
- Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. *Sociolinguistics and language acquisition*, 137, 174.
- Valk, L. (2010). *A Beginner's Guide to Building and Programming Robots*: starch press.
- Vogt, P., De Haas, M., De Jong, C., Baxter, P., & Kraemer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Frontiers in human neuroscience*, 11, 73.

Automatic Vocabulary Study Map Generation by Semantic Context and Learning Material Analysis

Brendan FLANAGAN^{a*}, Mei-Rong Alice CHEN^a, Louis LECAILLIEZ^b, Rwitajit MAJUMDAR^a, Gökhan AKÇAPINAR^{ac}, Patrick OCHEJA^b & Hiroaki OGATA^c

^a*Academic Center for Computing and Media Studies, Kyoto University, Japan*

^b*Graduate School of Informatics, Kyoto University, Japan*

^c*Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*

*flanagan.brendanjohn.4n@kyoto-u.ac.jp

Abstract: Learning English as a foreign language is a core part of K-12 education for many countries in which English is not the main spoken language, and especially in Asia. One of the fundamental tasks that students encounter is to learn vocabulary that is a part of the assigned curriculum. These are often sourced from reference materials or assigned vocabulary lists and may not consider the learner's current proficiency or the semantic context of words that were recently learnt. By suggesting vocabulary that have similar proficiency or semantic contexts to what a student has recently studied could improve and support vocabulary learning. In this paper, we propose a method for recommending words that have similar difficulty and semantic context with previous words learnt based on the analysis of prescribed textbooks for Japanese junior high school students. This research could be used to guide a student learning English by helping them select a sequence of vocabulary that is appropriate.

Keywords: English as a foreign language, vocabulary learning, semantic context, proficiency

1. Introduction

When studying vocabulary, students are often faced with the challenge of selecting appropriate words to learn at their current level of proficiency. Often students will refer to reference materials or an assigned vocabulary list that are part of their current curriculum. However, these materials do not consider the learners previously learnt vocabulary, which can give insight into the semantic context of their current study.

Nation & Hunston (2013), suggest that learning from context is one of the most important sources when studying vocabulary, and should be an integral part of second language learning. This highlights the strength of the role that context plays in helping learners not only understand the meaning of a word from its surrounding sentence, but also the similarities in contexts of words which are close to each other semantically. Wolter (2006) investigated the differences in native language lexical networks and the learner's perception of the foreign language lexical network. It was found that learners often rely on their knowledge from the native language lexical network. Borodkin et al. (2016) found that foreign language learners do not possess lexical networks that are as well-organized as the network of their native language. Therefore, learning vocabulary from a recommended vocabulary map might help improve the lexical network of the EFL or ESL learners.

In this paper, we propose a method for recommending English vocabulary that a learner could study based on a map of the whole curriculum. The map has been organized with the intention of recommending words that are closely related both semantically and according to the student's current level of acquisition. The analysis of three English as a foreign language textbooks that are widely used at Japanese junior high schools are examined as case studies of the proposed method.

Previous research into the use of semantic context analysis in language learning has focused on the substitution of difficult or unknown words to support reading. Azab et al. (2015) analyzed existing semantic maps, such as: WordNet (Miller, 1995), Roget, and Encarta, to find synonymous that could be used as substitutes to describe words the learner could not understand with reading text with a web

browser. While similar methods could be employed to the problem presented in this paper, there are numerous words that excluded from such semantic maps, and the relations are not derived from the actual context in which the word is used. To overcome these limitations, we propose that a word2vec corpus trained using the actual contexts of words would be better suited to the task of recommending vocabulary for learners.

2. Method Overview

2.1 Data Collection: Learning Materials

In this research, we focus on the analysis of three-year levels of learning materials that are used extensively in English Language classes at Japanese Junior High-Schools. A list of prescribed vocabulary is available for each textbook, and was used as a mask to focus the analysis in this paper. However, it should be noted that it is also possible to extract such information automatically from a set of textbooks by comparing the unique words in each book. An overview of the books and vocabulary is shown in Table 1.

Table 1

An overview of the three different levels of EFL learning materials that were targeted in this paper.

<i>Junior High School Year Level</i>	<i>Vocabulary Studied</i>	<i>Pages</i>
<i>1</i>	<i>389</i>	<i>124</i>
<i>2</i>	<i>225</i>	<i>116</i>
<i>3</i>	<i>341</i>	<i>106</i>

An ordered list of the vocabulary was extracted from each of the textbooks by finding the first mention of a word in the book. If words occurred on the same page, then words that were higher up the page were considered to have occurred before words that were lower down the page as the book is in single column layout.

2.2 Data Collection: Semantic Similarity Corpus and Vector Training

Word2vec was proposed by Mikolov et al. (2013) to effectively learn word embedding representation by examining the context of surrounding words in sample sentences. In essence, the resulting model is trained to predict a word in a sentence based on the surrounding words within a specified window. The weights of the resulting model are then used as a vector representation of the predicted word in a semantic space. Therefore, words that are around a similar area within the vector space are words that usually occur in similar contexts, and therefore have a similar meaning. This technique has been successfully applied to several semantic problems in previous research (Mikolov et al., 2013; Lau & Baldwin, 2016), and could be used to find semantically similar words for language learners to study.

For the purpose of this research, we trained a word2vec vector space model using a corpus based on the full collection of English Wikipedia. The trained model contains a total of 2.5m unique words, with each word being represented by a 300-dimension vector. A subset of this model was extracted for each of the three textbooks that are analyzed in this paper.

2.3 Recommending Vocabulary to Study

In this section, we introduce a method of creating a map of the vocabulary for each textbook. A word in the vocabulary is represented as a node, and the weight of edges between each pair of words represents the semantic context distance between the 2 words. The vector representations of vocabulary that were obtained from the word2vec model were used to measure the context distance by calculating the cosine distance between the vectors of two words. As a number of the items in the vocabulary list contain multiple words that would result in an equal number of vectors which could not be compared using standard techniques. A method of representing these words as a single vector is necessary, and the sum total of the word embedding vectors as proposed in Lau & Baldwin (2016) was used to represent these items. The map contains a large number of edges, and an optimal sub-map of the strongest relations was

extracted by the minimum spanning tree algorithm as described in Flanagan et al (2019). This results in a map that links words by the closest similarity while keeping the number of edges in the map at a minimum. The distribution of the similarity context distance of all of the edges of the maps before and after optimization are shown in Fig. 1 along with the mean and standard deviation in Table 2.

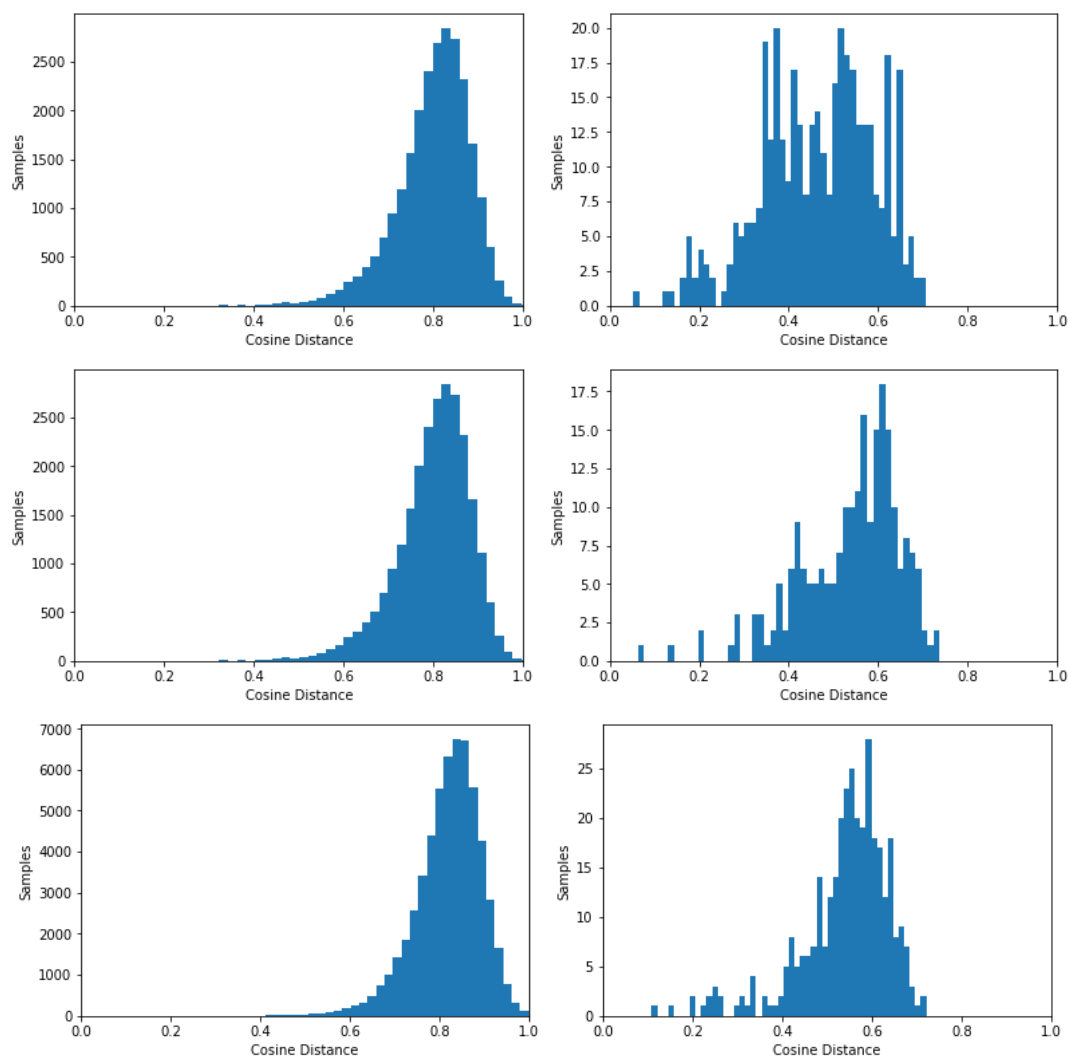


Figure 1. Distribution of semantic distance of all word relations (left) and of optimized word relations (right) for all year 1 (top) to year 3 (bottom) of junior high school.

Table 2

An overview of the three different levels of EFL learning materials that were targeted in this paper.

	Year 1		Year 2		Year 3	
	Mean	SD	Mean	SD	Mean	SD
All	0.7884	0.0911	0.7996	0.0824	0.8193	0.0740
Optimized	0.4645	0.1244	0.5384	0.1116	0.5415	0.0996

The distribution of the semantic context distance of vocabulary in the textbook for year 1 has a lower mean and greater standard deviation than higher years. This suggests that the range of vocabulary broadens as the target proficiency increases. Also, the mean distance of the optimized map is substantially lower than that of all relations, indicating that the relation of the vocabulary has greater semantic context similarity.

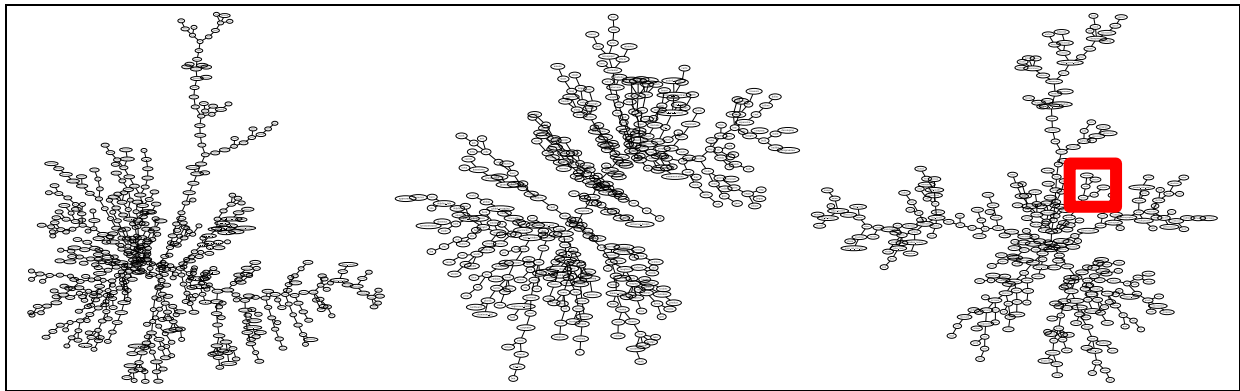


Figure 2. Overview of semantic vocabulary learning maps for year 1 (left) to year 3 (right) of junior high school.

An overview of the maps generated for three different textbooks using the proposed method based on cosine distance are shown in Fig. 1. Some of the branches in the map are long and are usually related to specific themes that are covered in the textbook, whereas shorter branches mostly cover specific language concepts.

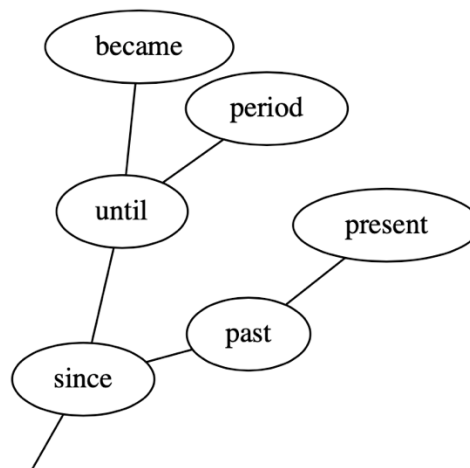


Figure 3. Details of a branch representing vocabulary that relate to time concepts that are studied in third year junior high school textbook.

An example of the types of vocabulary paths that the map can recommend a student to learn are shown in Fig. 3, which contains a detailed view of the red highlighted branch in Fig 2. It is made up of words that are related to the concept of time that students learning in the third year of junior high school. One sub-branch “past” and “present” represents the tense time concept, while other parts are related to time relations, such as: since, until, and period. An example of recommendations that could be made using this map would be if a learner has just studied the vocabulary “since” the system would then recommend that the learner next studies the words “until” and “past”. These vocabularies have the same semantic context of time in common with the parent word “since”. If the learner chooses to then study “past”, the system could then recommend that they should study the word “present” afterwards. The weights of the edges that are connected to the current node could be used to rank the recommendation, giving higher importance to vocabulary that have a close semantic context to previously learnt words.

In addition to recommending vocabulary based on the semantic context, recommendations could also be based on the difficulty of the word. As described in section 2.1, we extracted an ordered list of vocabulary from the textbooks based on the first occurrence of words. On the assumption that the textbooks would introduce increasingly difficult words as study progresses, this list can be analyzed to recommend vocabulary based on the difference of the position. For example the tense branch occurs in the following sequence within the textbook: “since” → “present” → “past”, which suggests that a learning should study the vocabulary in this order.

3. Conclusion

In this paper, we propose a method to automatically generate a map of the semantic context relations of vocabulary, and define how it can be used to recommend vocabulary items that a learner could study based on their learning history. The relation of the vocabulary items was determined by measuring the cosine distance between two words in a word2vec corpus that was trained on the sentence contexts of the English Wikipedia corpus. We applied the method to three textbooks that are frequently used in Japanese junior high schools when studying English as a foreign language and examined parts of the maps that were generated as case studies. The maps will be introduced to a knowledge mapping system (Flanagan et al., 2019) that is being deployed in Japanese K-12 institutions to support the study of English.

In future work, we plan to formally evaluate the use of the recommendation system in Japanese K-12 English as a foreign language class.

Acknowledgement

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

- Azab, M., Hokamp, C., & Mihalcea, R. (2015). Using word semantics to assist English as a second language learners. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 116-120).
- Borodkin, K., Kenett, Y. N., Faust, M., & Mashal, N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition*, 156, 60-70.
- Flanagan, B., Akçapınar, G., Majumdar, R., & Ogata, H. (2018). Automatic Generation of Contents Models for Digital Learning Materials. *Proceedings of the 26th International Conference on Computers in Education (ICCE2018)* (pp. 804-806).
- Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments, *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK19)*.
- Lau, J. H., & Baldwin, T. (2016). An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. *ACL 2016* (pp. 78-86).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Nation, I., & Hunston, S. (2013). Learning words from context. In *Learning Vocabulary in Another Language* (Cambridge Applied Linguistics, pp. 348-388). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139858656.010
- Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, 27, 4, 741-747.

Building a Confused Character Set for Chinese Spell Checking

Lung-Hao LEE^a, Wun-Syuan WU^b, Jian-Hong LI^a, Yu-Chi LIN^c & Yuen-Hsien TSENG^{c*}

^aDepartment of Electrical Engineering, National Central University, Taiwan

^bDepartment of Chinese, National Taiwan Normal University, Taiwan

^cGraduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan

*samtseng@ntnu.edu.tw

Abstract: In this paper, we describe the construction details of a confused character set for Chinese spell checking. The SIGHAN 2013-2015 bakeoff datasets are adopted to measure the performance of correct character suggestions. Our confusion set significantly outperforms the existing confusion set in candidate selection for automatic spelling checkers.

Keywords: Chinese spell checking, confusion set, shape similarity, pronunciation similarity

1. Introduction

Automatic Chinese spell-checking tools (like those for English in MS Word) are valuable for Chinese language learners. However, it is particularly challenging for Chinese, in part, because whether or not a Chinese character is correct depends on its context. For example, the character “偏” (“pian” in pronunciation) is correct in the word “偏見” (“pian jian4”; “prejudice” in English), but it is incorrect in the word “普偏” which instead should be “普遍” (“pu3 pian4”; “common”). Hence, a sentence “學生打工的情況很普偏” is incorrect and its corresponding correction is “學生打工的情況很普遍” (A student having a part-time job is very common). The possible error causes may be similar pronunciation and/or shape which are shared between the characters “偏” and “遍”.

Ideally, automatic spelling checkers should be capable of detecting character or word errors and making correction accordingly. In practice, they often first pin-point the locations of various types of spelling errors and then suggest a candidate list of correct characters for the detected error where the correct one should be as ranked as early as possible (favorably top 1 in the list) in order for automatic replacement. In other words, spell checking is often a two-step process: Once erroneous characters are identified, a confused character set (henceforth referred as “confusion set”) is applied for error correction/replacement. This paper focuses on the second part of the Chinese spell checking. In particular, we focus on building of a confusion set that is independent of the methods to spot the spell errors and that is better than the existing ones.

Empirical analysis shows that most Chinese spelling errors arise from phonologically similar, visually similar, and semantically confused characters (Liu et al. 2011). A confusion set contains a set of seed characters each accompanying with a list of corresponding confused characters. For example, a character “不” in the set can be used to find its confused character list such as “部, 步, 布, ...”. In general, a confusion set is built according to similar shape or pronunciation that are concerned. However, without a proper order of the confused character list, it is inconvenient for an automatic spelling checker to seek possible correction candidates effectively and efficiently. This observation motivates us to build a universal confusion set to facilitate automatic Chinese spell checking.

2. Chinese Confusion Set Construction

A first step for confusion set construction is to determine candidate characters as seeds. The Sinica Corpus is the balanced Chinese corpus with word segmentation and part-of-speech tagging. We obtain a word list with accumulated word frequency in Sinica Corpus 3.0. From this word list sorted in decreasing order of frequency, there are 4,743 distinct and commonly used Chinese characters. These characters are regarded as the seeds and stored according to their frequencies decreasingly.

The next step is to find easily confused characters for each seed character. The following three major methods are used and combined to yield a confused character list for each seed character:

Real-Error Frequency. In SIGHAN 2013-2015 bakeoffs, real spelling errors caused by Chinese native speakers or Chinese-as-a-foreign-language learners were collected and manually corrected to evaluate the performance of automatic spelling checkers (Tseng et al., 2015; Wu et al., 2013; Yu et al., 2014). In the training sets, there are 8,772 spelling errors, in which 3,394 characters are unique. We utilize these spelling errors (and their corrections) to find the confused characters for each corrected character (which is also in seed characters). For example, the most seen character “的” in the spelling errors is frequently misused as “得” (161 cases), followed by “地” (32 cases). In this case, we create a list of confused/misused characters for the character: “的”.

Shape Similarity. For this error type, we use common and basic vocabulary, e.g., use the word “unusual” rather than the word “arcane”. The [Master Ideographs Seeker](#) (“全字庫” in Chinese) developed by National Development Council, Taiwan provides the components of each Chinese character. For example, a character “腦” consists of components “月”, “𠂔”, “ノ”, “口” and “爻”. We then measure the shape similarity between two characters using the Jaccard similarity coefficient. For example, the similarity of “腦” and “惱” (its components are “忄”, “𠂔”, “ノ”, “口” and “爻”) is 0.66 (=4/6). If a character have more than one unique constructed components such as “行” (there are two constructed components: one is “彳”, “一” and “丁”; the other is “彳”, “二” and “丨”), each combination will be used to measure the similarity respectively and the larger one will be retained.

Pronunciation Similarity. The Master Ideographs Seeker also provides the pronunciation of each character. If more than one pronunciation is used for a character, the most common usage in the CKIP Electronic Dictionary will be adopted. In Chinese phonetics, a character can be pronounced using initial consonants, vowels, and tones. Because its relative complexity, we describe the three phonetic similarity measures individually and their combination with examples as follows.

For the initial consonants of Chinese phonetics, Table 1 shows points and manners of articulation and the corresponding coordinates. The similarity of any two consonants is regarded as distance calculation in the coordinate system. All distances are normalized into values ranging from 0 to 1. The most similar is 1 when identical consonants exist between two characters. For no-consonant cases, its similarity with any one of consonants is 0. Besides, the similarity of two no-consonant cases is also 1.

Table 1. *Places and Manners of Articulation for Consonants and their Corresponding Coordinates*

Manners		Places		bilabial	labiodental	alveolar	velar	palatal	post alveolar	dental
		status	vocal folds							
		airflow								
plosive	voiceless	unaspirated	ㄅ(1,1)		ㄆ(3,1)	ㄎ(4,1)				
		aspirated	ㄆ(1,2)		ㄆ(3,2)	ㄎ(4,2)				
affricate	voiced	unaspirated					ㄑ(5,3)	ㄒ(6,3)	ㄓ(7,3)	
		aspirated					ㄑ(5,4)	ㄒ(6,4)	ㄓ(7,4)	
nasal	voiced		ㄇ(1,5)		ㄋ(3,5)					
lateral approx.	voiced				ㄌ(3,6)					
fricative	voiceless			ㄘ(2,7)		ㄙ(4,7)	ㄨㄥ(5,7)	ㄨㄥ(6,7)	ㄨㄥ(7,7)	
	voiced							ㄨㄥ(6,8)		

For vowels (or called finals for more general cases) of Chinese phonetics, they are divided into five types called: 1) prenuclear glides: 一、ㄨ、ㄛ; 2) simple finals: ㄩ、ㄛ、ㄜ、ㄝ; 3) compound finals: ㄟ、ㄨㄛ、ㄨㄝ、ㄨㄥ; 4) nasal finals: ㄩㄥ、ㄨㄥ、ㄨㄥ、ㄨㄥ; and 5) retroflex final: ㄥ. If two prenuclear glides are the same, the similarity is 1, otherwise 0. For the remaining finals, if two finals are the same, the similarity is 1; if they are the same type, the similarity is 0.5; other cases are 0.

For tones of Chinese phonetics, there are five distinct tones, that is, neutral tone, 1st tone, 2nd tone, 3rd tone, and 4th tone. The largest similarity is 1 for identical tones and the lowest similarity is 0 for the pair tone: neutral vs. 4th. The similarity is 0.75 for the following pairs: neutral vs. 1st, 1st vs. 2nd, 2nd vs. 3rd, and 3rd vs. 4th. The similarity is 0.5 for the pairs: neutral vs. 2nd, 1st vs. 3rd, and 2nd vs. 4th. Finally, the similarity is 0.25 for the pairs: neutral vs. 3rd and 1st vs. 4th.

The above similarities of consonants, vowels, and tones are then averaged to yield the final pronunciation similarity. Take the two characters “生” (尸厶) and “身” (尸匕) for example. 1) Both initial consonants is “尸”; 2) both do not contain prenuclear glides; 3) the vowels “厶” and “匕” belong to the same type (i.e., the nasal finals); and 4) both tones are 1st tone. So the pronunciation similarity of “生” and “身” is 0.875, which is the average of similarities: 1, 1, 0.5 and 1.

Finally, to construct a confusion set, each candidate character is used to measure the confused degree with the seed character. Take a character “買” for example, the confused degree with “賣” is 14.59, in which it is calculated by the sum of real-error frequency 13 times, shape similarity 0.66 (exceeds the threshold of 0.5) and pronunciation similarity 0.93 (exceeds the threshold of 0.8), i.e., $13+0.66+0.93=14.59$. Finally, as an indexed seed character “買”, only a maximum of top 20 confused characters “賣, 乃, 奶, …” ordering by the confused degree are included in our confusion set.

3. Experiments and Evaluation Results

In this section, we evaluate our confusion set, compared with the confusion set provided by Liu et al. (2011) which is the only one we could find. The experimental data came from the SIGHAN 2013-2015 bakeoffs (Tseng et al., 2015; Wu et al., 2013; Yu et al., 2014), including 2,773 Chinese spelling errors. The objective is to measure the effectiveness of confusion sets after erroneous characters have been identified. For this purpose, Mean Reciprocal Rank (MRR), which is the average of reciprocal positions of correct characters in the descending-ordered list of confused characters, is adopted as an evaluation metric. The only previously existing confusion set (Liu et al., 2011) that contains confused characters generating from the combinations of the same/similar in shape/pronunciation/tone/radical/stroke are used for comparisons. Table 2 shows the evaluation results. Our confusion set achieved the best MRR of 0.4184, which significantly outperforms the existing set no matter which similarity measurement is used. This evaluation result supports us to apply the confusion set built by ourselves rather than using the existing one, so that we achieved the second prize in the 2018 Chinese Language Teaching Application Software Competition.

Table 2. Evaluation on Confusion Sets for Chinese Spell Checking

	Our Confusion Set	The Existing Confusion Set (Liu et al., 2011)					
		Similar Shape	Same Pron. Same Tone	Same Pron. Diff. Tone	Similar Pron. Same Tone	Similar Pron. Diff. Tone	Same Radical Same Stroke
<i>MRR</i>	<i>0.4184</i>	<i>0.1277</i>	<i>0.1989</i>	<i>0.0835</i>	<i>0.0078</i>	<i>0.0093</i>	<i>0.0131</i>

4. Conclusions

This study describes a new confusion set constructed for Chinese spell checking. Base on the experimental results, our set performs better than the previous one. Future work may re-rank the confused candidate list based on the contextual information to achieve better MRR performance.

Acknowledgements

This study was partially supported by the Ministry of Science and Technology, under the grant MOST 106-2221-E-003-030-MY2, 107-2221-E-003-014-MY2, 108-2218-E-008-017-MY3 and 108-2634-F-002-022.

References

- Liu, C.-L., Lai, M.-H., Tien, K.-W., Chuang, Y.-H., Wu, S.-H., & Lee, C.-Y. (2011). Visually and phonologically similar characters in incorrect Chinese words: analysis, identification, and applications. *ACM Transactions on Asian Language Information Processing*, 10(2), Article 10.
- Tseng, Y.-H., Lee, L.-H., Chang, L.-P., & Chen, H.-H. (2015). Introduction to SIGHAN 2015 bake-off for Chinese spelling check. *Proceedings of SIGHAN'15* (pp. 32-37). Beijing, China: ACL Anthology.
- Wu, S.-H., Liu, C.-L., & Lee, L.-H. (2013). Chinese spelling check evaluation at SIGHAN bake-off 2013. *Proceedings of SIGHAN'13* (pp. 35-42). Nagoya, Japan: ACL Anthology.
- Yu, L.-C., Lee, L.-H., Tseng, Y.-H., & Chen, H.-H. (2014). Overview of SIGHAN 2014 bake-off for Chinese spelling check. *Proceedings of CLP'14* (pp. 126-132). Wuhan, China: ACL Anthology.

Reading Assistance for EFL Readers with Kit-build Concept Map with Source-connection

Banni Satria ANDOKO^{ab*}, Yusuke HAYASHI^a, Tsukasa HIRASHIMA^a & Atiqah Nurul ASRI^b

^a*Department of Information Engineering, Hiroshima University, Japan*

^b*Department of Information System, State Polytechnic of Malang, Indonesia*

*ando@lel.hiroshima-u.ac.jp

Abstract: Recently, access to information is easier than before, but most of them are written in English. It leads to another new problem for a non-English speaking country. This research proposes an approach to improve English reading comprehension with the help of Kit-build concept map (KB-map) with an additional function named source connection. A university teacher in Indonesia practically used this approach and the other two approaches in several English classes of the second year of undergraduate student to compare the effectiveness of those approaches. The result shows that KB-map with the source-connection function was more effective in English reading comprehension compare to the traditional summarization and the normal Kit-build method.

Keywords: Concept map; Reading assistance; EFL; Kit-build concept map

1. Introduction

Concept mapping is one of the popular strategies used to improve Reading comprehension for English as a Foreign Language (Chang, Sung, & Chen, 2002). As a kind of concept map method, Kit-build concept map (KB-map) proven to have the same efficiency as the original concept map but unlike the original concept map (Funaoi, Ishida, & Hirashima, 2011, Alkhateeb, Hayashi, & Hirashima, 2013) even though KB-map is a kind of “closed-end” approach because it delivering the component of the teacher or expert concept map and student can only reconstruct the provided components without the ability to create or delete the component. This paper proposes an approach to improve English reading comprehension with the help of Kit-build concept map (KB-map) with an additional function named source connection. This approach helps the student to externalize their understanding in the form of the concept map and to confirm the relation between the concept map and the reading material. This study evaluates the effectiveness of the approach with the practical use of this by a university teacher in Indonesia, compared with the traditional summarization and the normal Kit-build concept map.

2. Kit-build Concept map and the additional function called source connection

Kit-build concept map (KB-map) has several phases (See Figures 1), the first phase called goal map building; this phase required a teacher/ expert to create an initial map from the reading material. The initial map will be called a goal map. After the teacher/ expert finished creating the goal map, the system will automatically disassemble it into components called kits consisting of nodes and links. The second phase, called learner’s map building. This phase required students/ learners to reconstruct the kits into a map. They cannot create or modify a component. The third phase is the KB Analyzer. This is the last step where the teacher/ expert can check the comparison result between their map and the students or learners. Since between the teacher or expert and the students using the same component, it will be easier to make the comparison by conducting an exact match comparison method.

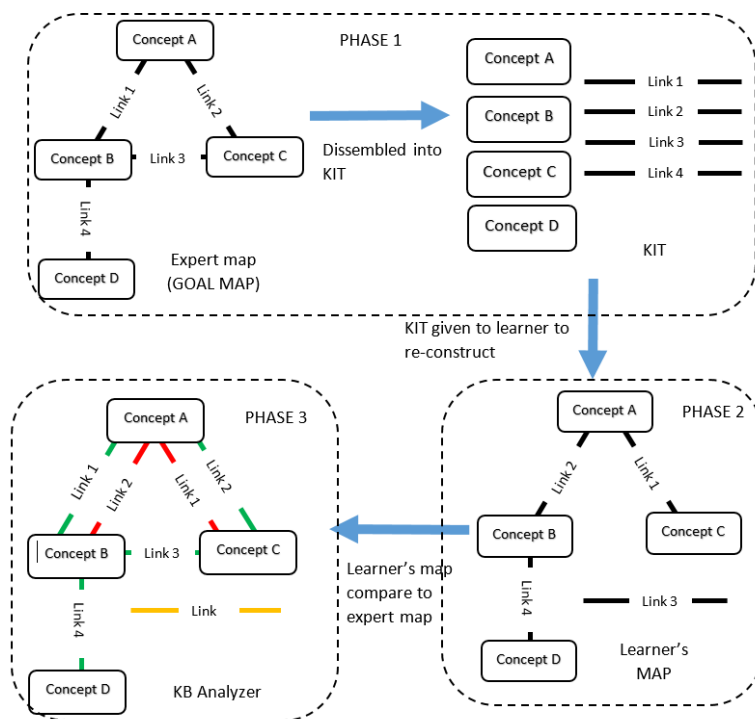


Figure 1. KB Map Phases

Source connection function aims to improve the use of KB map. This function facilitates learners to make confirmation between their reconstructed map to the reading material directly. By making a confirmation, hopefully, will decrease the number of misunderstanding. Source connection function facilitates students/ learners to confirm each proposition (consist of two nodes and one link) of the map to the reading material (see Figure 2) and then modify the proposition if necessary.

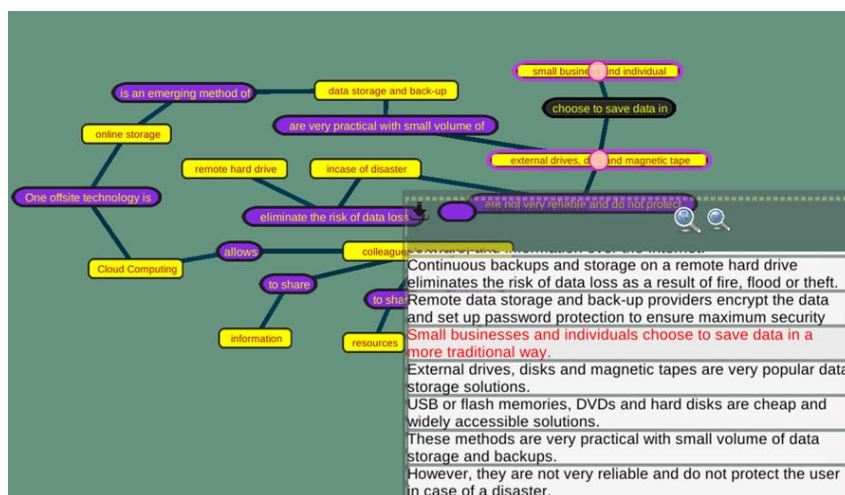


Figure 2. KB map with source connection function

3. Experimental setting and result

We experimented Indonesia. The experiment was involving two hundred and thirty-two undergraduate students. They are all learning English as a Foreign Language as a mandatory subject in the university. We are dividing the students into three groups. We divided them equally by using the English comprehension test result conducted before the experiment. The experiment was conducted for three times by using three different material for almost four weeks. Each experiment has the same activity as described in Figures 3.

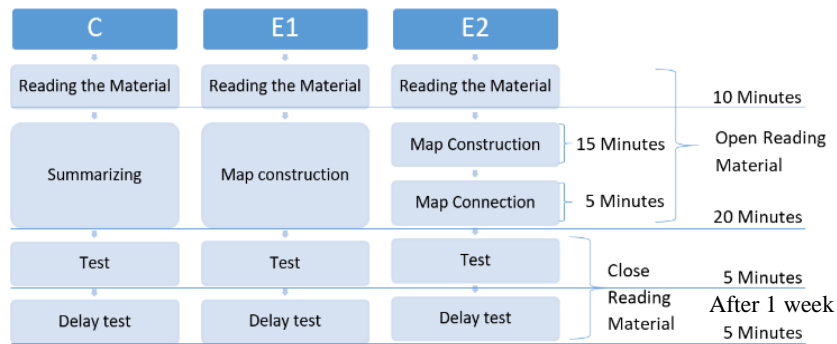


Figure 3. Experimental setting activity

The test result for each material is shown in Figure 4; we also conduct the Holm's Sequentially Rejective Bonferroni Procedure to rank up the groups to see the differentiation among the groups. The result shows the significant difference between the KB-map with source connection group and the control group in all material. While the KB map group only superior in two material to the control groups.

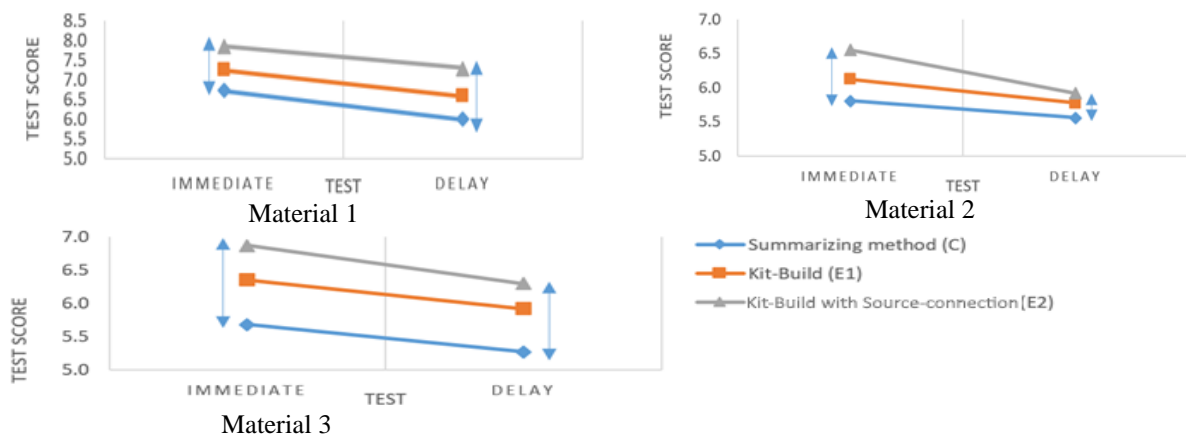


Figure 4. Test result for all material

4. Conclusion

In this paper, we propose the new function of KB map, source connection. The source connection function in KB-map helps students to confirm the relation between the concept map they have made and the reading material as the source. From the result of the practical use of it in Indonesian university, the effectiveness of this function is better than the reading comprehension with the traditional summarization and the normal Kit-build concept map. The future work is the analysis of the reading process with KB-map and source connection to confirm whether the function provides learners with the opportunity for self-reflection.

References

- Alkhateeb, M., Hayashi, Y., Rajab, T. & Hirashima, T.: Experimental Evaluation of the KB-mapping Method to Avoid Sentence-by-Sentence Map-building Style in EFL Reading with Concept Mapping, *INFORMATION AND SYSTEMS IN EDUCATION*, Vol. 15, No. 1, pp.1-14, 2016
- Chang, K. E., Sung, Y. T. & Chen, I. D. (2002). The Effect of Concept Mapping to Enhance Text Comprehension and Summarization. *The Journal of Experimental Education* 71(1), 5–23.
- Funaoui, H., Ishida, K., & Hirashima, T. (2011). Comparison of Kit-Build and Scratch-Build Concept Mapping Methods on Memory Retention. *Proceedings of the 19th International Conference on Computers in*, (pp. 539-546). Chiang Mai.

MAIN CONFERENCE C7

PRACTICE-DRIVEN RESEARCH, TEACHER PROFESSIONAL DEVELOPMENT AND POLICY OF ICT IN EDUCATION (PTP)

- FULL PAPER -

**FACTORS AFFECTING THE BEHAVIORAL INTENTION OF K-12 STUDENTS TO PURSUE AN IT DEGREE IN A
TRANSITIONING EDUCATIONAL SYSTEM..... 710**

RYAN EBARDO, LAIZA LIMPIN & MA. ROWENA CAGUIAT

CO-DESIGNING MULTIMODAL PEDAGOGICAL CONTENT KNOWLEDGE WITH INDONESIAN TEACHERS 720

HYO-JEONG SO, CHING SING CHAI, NUR ARIFAH DRAJATI, UMA NATARAJAN, HYERAN LEE & YUE WANG

TEACHING CONCEPTIONS AND TEACHING CONTEXT AFFECT THE ENACTED PRACTICE OF ONLINE TEACHERS 729

DENIESE COX & SARAH PRESTRIDGE

- SHORT PAPER -

**RELATIONSHIP BETWEEN PARENTS' PERCEPTIONS OF PROGRAMMING EDUCATION AND THEIR EMOTIONAL AND
BEHAVIORAL OUTCOMES 738**

SIU-CHEUNG KONG & YI-QING WANG

SUPPORTING TEACHERS IN GROUP WORK FORMATION AND ANALYTICS FOR IN-CLASS GROUP ACTIVITIES..... 744

LIANG CHANGHAO, IVICA BOTIČKI & HIROAKI OGATA

ONLINE TEACHER PROFESSIONAL DEVELOPMENT IN ICT INTEGRATION IN TANZANIA: AN EXPERIENCE REPORT 750

LUCIAN VUMILIA NGEZE & SRIDHAR IYER

**EDUCATIONAL USE OF SPHERICAL VIDEO-BASED VIRTUAL REALITY: A PRELIMINARY STUDY FROM THE TEACHER
PERSPECTIVE..... 756**

GRACE CHIH-NUO CHAO, MORRIS SIU-YUNG JONG, CHING-SING CHAI, & ERIC TSUN-HIN LUK

**A PRELIMINARY CASE STUDY OF PROMOTING TEACHER PROFESSIONAL DEVELOPMENT VIA POST-LESSON
DEBRIEFING WITH THE SUPPORT OF SOKRATES CLOUD SERVICES 762**

OSKAR KU, JEN-KAI LIANG, SUNG-BIN CHANG, & MICHAEL WU

**EXPLORING K-12 STUDENTS' ACCEPTANCES OF A COMPUTATIONAL MAKING PROGRAMME: A CASE STUDY IN
SINGAPORE..... 768**

LONGKAI WU, CHEE-KIT LOOI & SUJIN HE

- POSTER PAPER -

**IMPACT OF THE STEM PROGRAM ON INFORMATION TECHNOLOGY COLLEGE STUDENTS' GOALS: PERSPECTIVES
FROM THE PHILIPPINES 774**

JYPZIE M. CATEDRILLA, JOSEPHINE M. DE LA CUESTA, MA. ROWENA R. CAGUIAT

**EXPLORING THE EFFECTS OF SOCIO-ECONOMIC STATUS, MOTIVATION AND ICT USE ON SCIENCE ACHIEVEMENT:
FINDINGS FROM PISA 2015 777**

SHENG-HUA HUANG, YI-CHAO JIANG & MORRIS SIU-YUNG JONG

Factors Affecting the Behavioral Intention of K-12 Students to Pursue an IT Degree in a Transitioning Educational System

Ryan EBARDO^{a,c*}, Laiza LIMPIN^{b,c} & Ma. Rowena CAGUIAT^c

^a*Jose Rizal University, Philippines*

^b*Mindanao State University – General Santos City, Philippines*

^c*De La Salle University – Manila, Philippines*

*ryan.ebarido@jru.edu

Abstract: The rising technology sector in the Philippines has called for a holistic structural reform in its educational system through the implementation of the K-12 program. As such, higher educational institutions are in constant quest to look for strategies to increase enrollment in their Information Technology programs. To be effective, educational institutions must understand individual perceptions towards pursuing an IT degree in higher education. This paper presents the results of a study among senior high school student, who are one of the first recipients of the newly implemented system of education. Using additional constructs to the established Theory of Reasoned Action, a survey instrument was deployed in two higher educational institutions and gathered a total of 431 responses. Using Partial Least Squares – Structural Equation Model, results revealed that computer self-efficacy and job availability does not positively influence the attitude towards an IT degree at a significant level. The study also confirmed prior research that job salary and social image are primary factors that positively influence the attitude of students. Aligned with the findings of other research, attitude and social norms are strong predictors of the behavioral intention of students to pursue an IT degree. The paper concludes by suggesting theoretical, educational and social implications of the study and acknowledging its limitations.

Keywords: Theory of reasoned action, computing education, K-12 education, IT education, digital natives

1. Introduction

The Philippines is one among the Asian countries that has exhibited a healthy economic growth in the recent years. A Philippine Economic Outlook report reveals that the country's GDP increased by 6.7% in 2017, a figure that is slated to grow in the next years to come (FocusEconomics, 2018). The technology services industry is one sector that contributes to this development, which in 2014 raked in \$2 billion in revenues, a 20 percent increase from the previous year, and the second-highest growth amount in the information technology-business process management (IT-BPM) subsectors. The positive economic outlook was credited to the diversification of services by global in-house centers (GICs) and the growth of information technology outsourcing (ITO).

An industry road map for 2011-2016, however, reveals a need for the industry to develop more talent and skills and to gain stronger government support, specifically in providing remedial training and structural reforms at all educational levels (Mitra, 2013). The industry, government and academia are further called to establish stronger linkages and to exert more effort to expand the scale and scope of educational and ICT-related training programs. Education is considered as the main foundation of national development and serves as the driver of social and economic development (UNESCO, 2008). Higher education is expected to provide new skills to meet the demands in emerging markets (Hendel & Lewis, 2005). However, the Philippine education system has not been matched by adequate resources to deliver relevant skills (World Bank, 2012). As a result, a Filipino graduate is considered to be underqualified compared to the graduates coming from other countries (Acosta & Acosta, 2017).

The supply of graduates is not adequate to meet the requirements of the industry. In a Korn Ferry Institute report (2018), there are 85.2 million job openings globally, almost half will be

contributed by countries in the Asia Pacific region. Failure to address this deficit will have long-term negative economic impact worldwide. In order to address the growing manpower demand particularly in the technology services sector, the Philippine government embarked on an ambitious education sector reform program. In 2013, the Republic Act No. 10533 known as the Basic Education Act was passed and extended the elementary and secondary education cycle from 10 to 12 years (DepEd, 2013). The Philippines was one of the three countries left to have implemented the K-12 program worldwide, only having the 10-year basic education system and as such, the educational reform was a necessary move towards the improvement of the nation's global competitiveness. The transition to the new educational system is to be completed by 2021.

Prior studies investigating factors that influence the decision-making process of students in pursuing technology-related degrees are based mainly on environments that have adopted and implemented the twelve-year basic educational system for years. The unique situation in the Philippines presents an opportunity for research to understand determining factors of K-12 students in their pursuit of an IT degree while the system is undergoing a shift in its educational structure. As one of the latest to adopt this educational system, investigating the insights of students born in the digital age and how they perceive an IT degree as an option in their future academic plans merits further scholarship inquiry. Lastly, by incorporating additional dimensions to the Theory of Reasoned Action (TRA) such as computer self-efficacy, job availability, job salary and social image, we contribute further to its theoretical precision.

2. Related Studies

Recognizing the economic impact of technology courses, prior research has investigated factors and motivation of students that influence their intention to enroll in such programs in higher education. The study by Joshi and Kuhn (2011) argues that attitudes and social beliefs collectively influence the decision of students to enroll in a tech-related program. The study further investigated the influence of other factors such as self-efficacy, image, work value and congruency in the behavioral intention to pursue an IT career. The students' interests, prospects, curriculum and social influence are also critical factors considered by students (Zhang, 2007). Moreover, in a cross-cultural work by Shin, et al., (2018), career motivation of university-bound students is affected by different socio-cultural factors such as gender, race and educational level. A longitudinal study by Mau (2016) also reveals that gender and racial differences are major factors. These studies further call for an investigation of socio-cultural contexts embedded in each country when examining the factors that contribute to the students' decision in pursuing a degree that will help them land a technology-related career in the future.

In the Philippines, there is a high regard for university graduates and education is embedded in the political, social and cultural ecologies. Since the K-12 program in the country is still in its infancy, its impact remains under-investigated. In the study of Montebon (2014), students are found to be perceptive to the new curriculum of the K-12 program which meets their cognitive requirements. A study by Mohammad (2016) captured insights from students, parents and the community and revealed heterogenous perspectives towards the new educational system. The additional years in senior high school are perceived to be a financial burden to some parents. On the other hand, students expressed the lack of their readiness to embrace K-12. While these studies focused on perceptions, the work by Bonifacio (2013) addresses the challenge of K-12 by calling for standards in the ICT curriculum. Furthermore, this study stressed the importance of different stakeholders to successfully implement K-12 such as administrators, teachers, and the government.

3. Theoretical Background and Hypotheses

The Theory of Reasoned Action or TRA (Ajzen & Fishbein, 1980; Ajzen, 1975) predicts an individual's behavior based on their pre-existing attitudes and behavioral intentions. It posits that a person's behavior is motivated by the intention to perform a certain action and the stronger the intention, the more likely the behavior will be executed (Zhang, 2007). The two components that lead to the development of behavioral intention includes a person's attitude and the subjective norms. Attitude refers to an individual's interpretation of the consequences of performing an action and subjective norms refers to an individual's perception towards a behavior based on the social and environmental pressures. Therefore, in this model (as shown in Figure 1), students' intention to choose an IT degree are

influenced by their attitudes towards pursuing an IT degree and the social pressures exerted on them to choose the course.

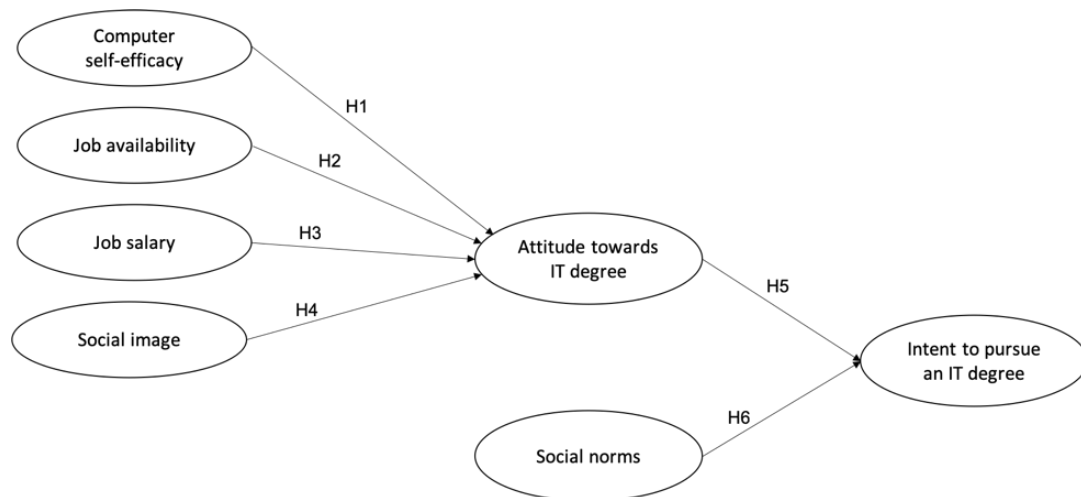


Figure 45 Theoretical Framework

Computer self-efficacy refers to a “judgement of one’s capability to use a computer” (Compeau & Higgins, 1995) through multiple domains which influences an individual’s decision to use a computer to become knowledgeable (Torkzadeh, Chang, & Demirhan, 2006). The relationship of computer self-efficacy and student’s performance has been the object of various studies. For instance, a study reveals that students with a positive attitude towards computers improved their self-efficacy significantly compared to those with negative attitudes (Compeau & Higgins, 1995). In addition, students with a high-level of computer self-efficacy shows better attitudes and academic performance than those with a lower level of computer self-efficacy (Akbulut & Looney, 2007; Peinado & Ramírez, 2014; Smith, 2002). Behavioral and psychological factors also influence a person’s computer self-efficacy and as a result, it positively influences a students’ performance in computational environments (Moos & Azevedo, 2009).

We argue that students’ attitude towards pursuing an IT degree will be partly influenced by their computer self-efficacy beliefs and that a high self-efficacy which corresponds to the IT skills that they possess will lead to a more positive attitude towards an IT course.

H1: Computer self-efficacy positively influences a student’s attitude towards pursuing an IT degree.

Studies have found that job market influences relate to students’ choice of college degrees (Fiorito & Dauffenbach, 1982; Zhang, 2007). Job availability was found to be a crucial factor for students to be confident in choosing a degree to pursue in college (Stanko, Zhirosh, & Krasnikhin, 2015). Moreover, students who pursue an IT degree were found to be aware that job opportunities in the IT field exist and were not worried about its availability (Kuechler, Mcleod, & Simkin, 2009). However, studies have either identified or confirmed job availability as a relevant predictor of a student’s intention to pursue an information systems-related degree (Jackling & Calero, 2006; Kuechler et al., 2009; Kumar & Kumar, 2013; Snyder, 2013; Tantuco, 2017; Zhang, 2007). Moreover, high job availability were related to student accountants’ decision to pursue a degree in the accounting field (Jackling & Calero, 2006) and not surprisingly, a lower job availability discouraged students to get into any IT-related degree (Lee & Lee, 2006). Therefore, we argue the following:

H2: Job availability positively influences a student’s attitude towards pursuing an IT degree.

A satisfying job salary has also been observed as an important factor in pursuing an IS degree (Granger, Dick, Jacobson, & Slyke, 2007; Snyder, 2013) as it is in other fields such as accounting (Cohen & Hanno, 1993) and medical-related majors (Sheikh, Naqvi, Sheikh, Naqvi, & Bandukda, 2012). In May 2013, the US Bureau of Labor Statistics reported an average annual salary of \$81,860 for computer occupation (“Occupational Employment and Wages,” 2013). A Forbes report also revealed the IT professionals have stayed satisfied with their income and their jobs in recent years (Adams, 2015).

We argue that students' outcomes expectations specifically on job salary for IT jobs will influence a positive perception, therefore:

H3: Job salary positively influences a student's attitude towards pursuing an IT degree.

A positive social image of a profession is essential to a successful career (Kumar & Kumar, 2013; Zhang, 2007) and a student's perception of a certain profession provides guidance in their career choice development (Glerean, Hupli, Talman, & Haavisto, 2017). For instance, nursing students have a positive image of the field which influences their attitude and perception towards nursing as a profession (Emeghebo, 2012). Moreover, while IT professionals may be perceived as nerdy and geeky, the IT profession is still described as exciting, youthful and enjoyable (E. Myers & Beise, 2001). In this context, we argue that:

H4. Social image positively influences a student's attitude towards pursuing an IT degree.

In this study, we hypothesized that factors such as computer self-efficacy, job availability, job salary and social image contribute to a student's positive attitude towards pursuing a degree in IT. Attitude, in this context, refers to a student's perception about choosing an IT degree. Moreover, a more positive attitude to IT is associated with a greater intention to pursue a degree in IT, therefore, we hypothesize the following:

H5. A positive attitude towards IT is positively associated with the intent to pursue an IT degree in college.

In TRA, a student's intention to pursue a degree in IT is also influenced by the social norms or the social pressures exerted to them to choose the course. Individuals are influenced by the beliefs of people who are important to them and plays an important role in determining behavior in a wide variety of domains such as in adopting a technology (Venkatesh & Davis, 2000). Students may feel pressured to pursue a degree based on the influence of their family, guidance counselors, high school advisers, friends and fellow students (Kumar & Kumar, 2013; Snyder, 2013; Stanko et al., 2015; Zhang, 2007). Therefore, we argue that

H6. Social norms are positively associated with the intent to pursue an IT degree in college.

4. Methodology

To test the applicability of the TRA in the research context of this study, an empirical approach is adopted. Questions from prior research are incorporated in a survey instrument. To test the validity of the instrument, a pilot test was conducted with selected students. Two institutions, A and B, offering a senior high school program were approached and the validated instrument was distributed online. Institution A is a state university based in the Southern Philippines. Institution B is a private college located in Metro Manila. The responses are recorded and a Partial Least Squares – Structural Equation Model is applied to test the propositions by the study.

4.1 Instrument Development

To account for the constructs in the proposed framework, questions from previous studies were modified and included in an initial instrument. To represent the dimension of computer self-efficacy (CSE), 3 questions from a study by Joshi and Kuhn (Joshi & Kuhn, 2011) were included. Social image (SI) is operationalized in this study through the 3 questions adopted from the research of Zhang (Zhang, 2007). Job availability (JA) and job salary (JS) were investigated in prior studies to have an influence in student's attitude towards pursuing a college degree and are integrated in the instrument with 2 questions each adopted from the study of Kumar and Kumar (Kumar & Kumar, 2013). Attitude (ATT), social norm (SOC) and intention (INT) are the primary constructs of the TRA and are adopted in this study through the questions of Zhang (2007) with 2, 5 and 2 questions respectively. The preliminary survey instrument is shown in Table 1.

Table 1. Instrument

	Indicators	Questions
Computer Self-Efficacy	CSE1	I often find it easy to use a computer.
	CSE2	I do not need assistance in computer-related tasks or assignments.
	CSE3	I usually find computer-related assignments easy.
Social Image	SI1	Businessmen look up to IT professionals
	SI2	If I choose an IT degree in college or university, I will have a respectable career.
	SI3	The business world treats IT professionals with great respect.
Job Availability	JA1	There will be plenty of job opportunities available in IT.
	JA2	There will be job opportunities available in IT when I graduate.
Job Salary	JS1	I can get a high paying job in IT.
	JS2	My starting salary will be satisfying in IT.
Attitude	ATT1	Choosing IT degree seems a good idea to me.
	ATT2	It will be wise for me to choose an IT degree.
Social Norms	SOC1	My friends think I should choose an IT degree.
	SOC2	Other students recommend that I enroll in an IT degree.
	SOC3	My class adviser recommends that I should choose an IT degree.
	SOC4	My teachers believe that choosing an IT degree is best for me.
Intention	INT1	I intend to choose an IT degree in college or university.
	INT2	It is likely that I will choose IT degree.

4.2 Instrument Validation

To test the validity of the instrument, 16 students were invited to answer the online survey as part of the pilot test. A Partial Least Squares through SmartPLS is applied to the results of the pilot sample to identify questions are below the minimum threshold of Cronbach's Alpha (0.70) Average Variance Extracted (0.60) and Composite Reliability (0.70). One question from the social norm construct, SOC5, was dropped from the initial instrument due to the low values in Average Variance Extracted and Composite Reliability. Table 2 shows that the Cronbach's Alpha, Composite Reliability (CR) and Average Variance Extracted (AVE) values are above minimum threshold required to establish reliability and validity of the instrument for this study (Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, 2014).

Table 2. Instrument Validation Result

Constructs	Cronbach's Alpha	Composite Reliability (CR)	Average Variance Extracted (AVE)
Computer Self-efficacy	0.773	0.846	0.651
Job availability	0.915	0.959	0.922
Job salary	0.939	0.970	0.942
Social image	0.862	0.913	0.779
Attitude	0.908	0.956	0.915
Social Norm	0.937	0.954	0.839
Intention	0.792	0.903	0.824

5.3

4.3 Demographics

5.4

A total of 431 respondents from Institution A and Institution B answered the validated research instrument. All students are enrolled in academic year 2018-2019 in the senior high school department

of both institutions. Of the sample population, 184 or 42.69% are females while 247 or 57.31% are males. Majority of the respondents are between 16 to 18 years old and more than half or 50.59% of participants are 17 years old. Of the respondents, 209 or 48.49% are from Year 11 and 222 or 51.51% are on Year 12.

Table 3. Demographics

Number of Respondents		A	137
		B	294
Gender	Male	184	42.69%
	Female	247	57.31%
Age	15 and Below	2	0.46%
	16	70	16.24%
	17	218	50.58%
	18	131	30.39%
	19	7	1.62%
	20	2	0.46%
	21	1	0.23%
Year Level	11	209	48.49%
	12	222	51.51%

5. Results and Discussion

To test the applicability of our proposed model, a Partial Least Squares - Structural Equation Model or PLS-SEM was applied to the results of the survey. The number of the indicators, the sample size and the predictive nature of this investigation are the primary motivators for the adoption of this statistical technique. In addition, PLS-SEM have been applied in prior literature in the education domain with great success (Ifinedo, 2017; Ramírez-Correa, 2017). Using a bootstrapping technique in SmartPLS (Ringle, Wende, & Becker, 2015), results are examined as shown in Table 3 – Summary of Results to assess whether to accept or reject a hypothesis at 0.05 significance level.

Table 4
Summary of Results

Hypothesis	Statement	T-Statistics	Decision
H1	Computer self-efficacy positively influences a student's attitude towards pursuing an IT degree.	0.793	Rejected
H2	Job availability positively influences a student's attitude towards an IT degree.	0.930	Rejected
H3	Job salary positively influences a student's attitude towards an IT degree.	5.068	Accepted
H4	Social image positively influences a student's attitude towards an IT degree.	3.421	Accepted
H5	A positive attitude towards IT is positively associated with the intent to pursue an IT degree in college.	6.590	Accepted
H6	Social norms are positively associated with the intent to pursue an IT degree in college.	10.860	Accepted

While existing literature have supported computer self-efficacy to be highly influential in students' intention to enroll in further studies (Lim, 2001; Soykan & Kanbul, 2018), this cannot be proven in the context of this study as proven by the T-statistics value of 0.793 for H1. Consistent with the findings of a prior research (Joshi & Kuhn, 2011), computer skills does not translate to a positive

attitude towards pursuing a technology-oriented degree. Several explanations are offered by this study. First, a possible explanation is that the respondents are classified as digital natives, a generation described to possess ICT skills at an early age and are inherently technology capable. These are individuals who are exposed to digital technologies and generally have access to computers and the Internet (Margaryan, Littlejohn, & Vojt, 2011; Uwizeyimana, 2018). Since acquisition of technical skills to operate technology devices is considered a normalcy among digital natives, it can be inferred that possession of such will not necessarily convert to a positive attitude towards pursuing an IT degree. Second, prior studies have also explored the affordances of ICT among students. Students have used digital technologies mostly for leisure such as gaming, communication and social media (Šorgo, Bartol, Dolničar, & Boh Podgornik, 2017). Through this luxury and coupled with access to a wealth of information available online, students at this age may conduct preliminary research on their interests in their college degree plans that are not necessarily specializing in IT.

Prior research have identified job availability as an influential factor in the attitude towards pursuing a technology-oriented degree (Kuechler et al., 2009). While the T-statistics score of 0.930 for H2 implies a positive influence, this cannot be supported at a significant level. Several reasons are offered by this study. First, majority of the respondents are coming from Institution B and that majority of IT job opportunities are concentrated in Metro Manila where business processing outsourcing (BPO) companies have established their presence (Price, Francisco, & Caboverde, 2016). Although major multinational companies engaged in the technology sector have slowly ventured into regions outside the capital, the job opportunities brought forth by such movements are yet to be enjoyed by residents in the provinces (Price et al., 2016). Second, job prospects among Filipinos are still largely anchored on opportunities outside the Philippines (Welsh, 2016). The economy is primarily driven by remittances of Overseas Filipino Workers who are working in industries such as nursing and hospitality (Aquino, Tuazon, Yap, & David, 2017; Castro-Palaganas et al., 2017; Joia & dos Santos Vinhais, 2017). The culture of economic success through working overseas is intricately embedded among Filipinos and demonstrates a strong influence in how they perceive job opportunities and consequently their decision to choose their educational path.

Consistent with existing scholarship, job salary (H3) and social image (H4) both positively influence attitude within the proposed research framework with T-Statistics values of 5.068 and 3.421 respectively. The perceived competitive salary rates associated with the ICT industry including the Philippines encourages keen interest towards pursuing an IT degree (Kuechler et al., 2009; Tantuco, 2017). Despite the fact that the social image of IT professionals have been perceived to have a negative image in prior literature (Grim, Harmon, & Gromis, 2006), this may not be true in the Philippine context. IT professionals are regarded as catalysts of innovation who deliver products mostly enjoyed by digital natives.

Lastly, a positive attitude and social norms lead to behavioral intention to pursue an IT degree with values of 6.590 and 10.860 respectively for H5 and H6. Consistent with other studies, a positive attitude is a crucial determinant in the decision making process of students in their academic pathways (Joshi & Kuhn, 2011; Sathapornvajana & Watanapa, 2012). Social norms appear to be the most influential factor in the students' in choosing an IT degree for college. Normative pressures in student life have been research extensively in research. It assumes that peer beliefs and behaviors are influential in the decisions of an individual (Abdullah & Ward, 2016; Camara, Eng-Ziskin, Wimberley, Dabbour, & Lee, 2017). In the context of this study, social norms are represented by people who students primarily interact with such as classmates, friends, advisers and teachers who can significantly influence their decision to attend an IT degree program in higher education.

6. Conclusion and Future Directions

In conclusion, this research aimed to investigate the factors affecting the behavioral intention of K-12 students to pursue a degree in IT. This study has supported the applicability of Theory of Reasoned Action in investigating the factors influencing student's behavior towards pursuing an IT degree. Particularly, it confirms that a high initial salary and a good professional image of working in the IT industry leads to a positive attitude towards an IT degree which further translates to a behavioral intention to pursue a degree in IT. Additionally, the pressures from the student's social environment such as classmates, friends, and teachers relates to their decision as to which academic direction they want to pursue in the future.

However, the result of the study revealed that a student's computer skills do not necessarily translate to a positive attitude towards pursuing a technology-oriented degree. Several explanations were offered by this study including the fact that the respondents are digital natives who possess ICT skills and are inherently technologically savvy. In addition, the respondents may have used their access to digital technologies, the Internet and its vast array of information to research professions that they are interested in, which may not be necessarily specializing in IT.

Additionally, the study does not support that the availability of IT jobs encourages students to pursue an IT course in college. A possible factor leading to this result is the fact that most BPO and tech companies in the country are concentrated in the urban areas and are yet to be distributed into the other regions. Also, many Filipinos still view job opportunities available abroad particularly in the hospitality and healthcare industry which possibly influence their decisions to choose their academic path.

Several implications can be drawn from this study. For theoretical implications, attitude and social norms are confirmed to be influential in behavioral intention to pursue IT. However, other determinants in addition to job salary and social image of the IT profession may be more appropriate to positively influence the attitude towards pursuing an IT degree. For educational implications, regulatory bodies that govern K-12 should emphasize on active industry-academic curricular integration and promote opportunities in the technology sector. For social implications, to narrow the digital divide, stakeholders of the technology industry should develop a blueprint that will distribute the IT opportunities across the country.

This study is not without limitation therefore the findings should be interpreted with caution. First, although this study examined perspectives from private and public higher educational institutions, the limited number of respondents may limit the generalizability of the results of this research. Second, there was no comparison between the results of institutions A and B, therefore this investigation cannot account for the significant differences of the results of the two sets of respondents. In this regard, we encourage future researchers to focus on the significant differences in attitudes and behavioral intention of students in pursuing an IT degree in different contexts such as public versus private schools, rural versus urban areas, and schools with different K-12 track offerings.

References

- Abdullah, F., & Ward, R. (2016). Developing a General Extended Technology Acceptance Model for E-Learning (GETAMEL) by analysing commonly used external factors. *Computers in Human Behavior*, 56, 238–256. <https://doi.org/10.1016/j.chb.2015.11.036>
- Acosta, I. C., & Acosta, A. S. (2017). A Mixed Methods Study on Teachers' Perceptions of Readiness of Higher Education Institutions to the Implementation of the K-12 Curriculum. *Universal Journal of Educational Research*, 5(7), 1215–1232. <https://doi.org/10.13189/ujer.2017.050714>
- Adams, S. (2015). Money Doesn't Buy Happiness At Work, New Study Says. Retrieved December 11, 2018, from <https://www.forbes.com/sites/susanadams/2015/06/18/money-doesnt-buy-happiness-at-work-new-study-says/#7955953f19bc>
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Ajzen, Icek. (1975). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research., 6(2), 244. <https://doi.org/10.2307/2065853>
- Akbulut, A. Y., & Looney, C. (2007). Inspiring students to complete computing degrees. *Communications of the ACM*, pp. 67–71.
- Aquino, R. S., Tuazon, G. P., Yap, T. W., & David, I. B. M. (2017). In search of greener pastures? Investigating Filipino tourism and hospitality management students' willingness to work overseas. *Asia-Pacific Journal of Innovation in Hospitality and Tourism*, 6(2), 81–90.
- Bonifacio, A. L. (2013). Developing Information Communication Technology (ICT) Curriculum Standards for K-12 Schools in the Philippines, 12. <https://doi.org/10.3748/wjg.v19.i38.6458>
- Camara, S. K., Eng-Ziskin, S., Wimberley, L., Dabbour, K. S., & Lee, C. M. (2017). Predicting Students' Intention to Plagiarize: an Ethical Theoretical Framework. *Journal of Academic Ethics*, 15(1), 43–58. <https://doi.org/10.1007/s10805-016-9269-3>
- Castro-Palaganas, E., Spitzer, D. L., Kabamalan, M. M. M., Sanchez, M. C., Caricativo, R., Runnels, V., ... Bourgeault, I. L. (2017). An examination of the causes, consequences, and policy responses to the migration of highly trained health personnel from the Philippines: the high cost of living/leaving—a mixed method study. *Human Resources for Health*, 15(1), 25. <https://doi.org/10.1186/s12960-017-0198-z>

- Cohen, J., & Hanno, D. M. (1993). An analysis of underlying constructs affecting the choice of accounting as a major. *Issues in Accounting Education*, 8(2), 219. Retrieved from <https://search.proquest.com/docview/210925575?accountid=190474>
- Compeau, D. R., & Higgins, C. A. (1995). Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly*, 19(2), 189. <https://doi.org/10.2307/249688>
- DepEd. (2013). Implementing Rules and Regulations of the Enhanced Basic Education Act of 2013. *Official Gazette*. Retrieved from <https://www.officialgazette.gov.ph/2013/09/04/irr-republic-act-no-10533/>
- E. Myers, M., & Beise, C. (2001). *Nerd Work: attractors and barriers perceived by students entering the IT field*. <https://doi.org/10.1145/371209.371236>
- Emeghebo, L. (2012). The image of nursing as perceived by nurses. *Nurse Education Today*, 32, e49–e53. Retrieved from <http://10.0.3.248/j.nedt.2011.10.015>
- Fiorito, J., & Dauffenbach, R. (1982). Market and Nonmarket Influences on Curriculum Choice by College Students Author (s): Jack Fiorito and Robert C . Dauffenbach Published by : Sage Publications , Inc . Stable URL : <https://www.jstor.org/stable/2522295> Market and Nonmarket Influences on CU. *Industrial and Labor Relations Review*, 36(1), 88–101.
- FocusEconomics. (2018). Philippines Economy - GDP, Inflation, CPI and Interest Rate. Retrieved November 15, 2018, from <https://www.focus-economics.com/countries/philippines>
- Glerean, N., Hupli, M., Talman, K., & Haavisto, E. (2017). Young peoples' perceptions of the nursing profession: An integrative review. *Nurse Education Today*, 57, 95–102. <https://doi.org/https://doi.org/10.1016/j.nedt.2017.07.008>
- Granger, M. J., Dick, G., Jacobson, C. M., & Slyke, C. Van. (2007). Information Systems Enrollments: Challenges and Strategies. *Journal of Information Systems Education*, 18(3), 303–312.
- Grim, B. J., Harmon, A. H., & Gromis, J. C. (2006). Focused Group Interviews as an Innovative Quanti-Qualitative Methodology (QQM): Integrating Quantitative Elements into a Qualitative Methodology. *The Qualitative Report*, 11(3), 516–537. <https://doi.org/10.1007/978-3-531-92053-5>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). Partial least squares structural equation modeling (PLS-SEM). *Sage Publisher*. <https://doi.org/10.1108/EBR-10-2013-0128>
- Hendel, D. D., & Lewis, D. R. (2005). Quality assurance of higher education in transition countries: Accreditation - accountability and assessment. *Tertiary Education and Management*, 11(3), 239–258. <https://doi.org/10.1080/13583883.2005.9967149>
- Ifinedo, P. (2017). Examining students' intention to continue using blogs for learning: Perspectives from technology acceptance, motivational, and social-cognitive frameworks. *Computers in Human Behavior*, 72, 189–199. <https://doi.org/10.1016/j.chb.2016.12.049>
- Jackling, B., & Calero, C. (2006). Influences on Undergraduate Students' Intentions to become Qualified Accountants: Evidence from Australia. *Accounting Education*, 15(4), 419–438. <https://doi.org/10.1080/09639280601011115>
- Joia, L. A., & dos Santos Vinhais, J. C. (2017). From closed source to open source software: Analysis of the migration process to Open Office. *Journal of High Technology Management Research*, 28(2), 261–272. <https://doi.org/10.1016/j.hitech.2017.10.008>
- Joshi, K. D., & Kuhn, K. (2011). What Determines Interest in an IS Career? An Application of the Theory of Reasoned Action. *Communications of the Association for Information Systems*, 29(September 2011), 133–159.
- Korn Ferry Institute. (2018). *Future of work The Global Talent Crunch Economy Perspective: Hong Kong*: Retrieved from <https://focus.kornferry.com/wp-content/uploads/2018/06/KF-Talent-Crunch-Country-Report-HK-Digital.pdf>
- Kuechler, W. L., Mcleod, A., & Simkin, M. G. (2009). Why Don't More Students Major in IS? *Decision Sciences Journal of Innovative Education*, 7(2), 463–488.
- Kumar, A., & Kumar, P. (2013). *An Examination of Factors Influencing Students Selection of Business Majors Using TRA Framework*. *Decision Sciences The Journal of Innovative Education* (Vol. 11). <https://doi.org/10.1111/j.1540-4609.2012.00370.x>
- Lee, Y., & Lee, S. J. (2006). *The competitiveness of the information systems major: An analytic hierarchy process*. *Journal of Information Systems Education* (Vol. 17).
- Lim, C. K. (2001). Computer self-efficacy, academic self-concept, and other predictors of satisfaction and future participation of adult distance learners. *American Journal of Distance Education*, 15(2), 41–51. <https://doi.org/10.1080/08923640109527083>
- Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students' use of digital technologies. *Computers and Education*, 56(2), 429–440. <https://doi.org/10.1016/j.compedu.2010.09.004>
- Mau, W.-C. (2016). Characteristics of US Students That Pursued a STEM Major and Factors That Predicted Their Persistence in Degree Completion. *Universal Journal of Educational Research*, 4(6), 1495–1500.

- <https://doi.org/10.13189/ujer.2016.040630>
- Mitra, R. M. (2013). Leveraging service sector growth in the Philippines. *ADB Economics Working Paper Series*, 366(366), 1–31. <https://doi.org/10.2139/ssrn.2321536>
- Mohammad, N. K. (2016). the Perception of the Parents and Students on the Implementation of K – 12 Basic Education Program in the Philippines, *1*(July), 978–602.
- Montebon, D. R. T. (2014). K12 Science Program in the Philippines: Student Perception on its Implementation. *International Journal of Education and Research*, 2(12), 153–164.
- Moos, D. C., & Azevedo, R. (2009). Learning With Computer-Based Learning Environments: A Literature Review of Computer Self-Efficacy. *Review of Educational Research*, 79(2), 576–600. <https://doi.org/10.3102/0034654308326083>
- Occupational Employment and Wages. (2013). <https://doi.org/10.1021/ja011916o>
- Peinado, S., & Ramírez, J. (2014). Effect of Learning Styles and Computational Self-Efficacy in Internet Forum Performance. *Investigación y Postgrado*, 25, 145–168.
- Price, N. A., Francisco, J. P., & Caboverde, C. E. (2016). *IT-BPO in the Philippines: A Driver of Shared Prosperity? SSRN*. <https://doi.org/10.2139/ssrn.2880408>
- Ramírez-Correa, P. E. (2017). Relationship Between Cyber Plagiarism and the Big Five Personality Traits: an Empirical Study in a Chilean University. *Holos*, 5, 125. <https://doi.org/10.15628/holos.2017.5191>
- Ringle, C., Wende, S., & Becker, J. (2015). SmartPLS 3. <https://doi.org/http://www.smartpls.com>
- Sathapornvajana, S., & Watanapa, B. (2012). Factors affecting student's intention to choose IT program. *Procedia Computer Science*, 13, 60–67. <https://doi.org/10.1016/j.procs.2012.09.114>
- Sheikh, A., Naqvi, S. H. A., Sheikh, K., Naqvi, S. H. S., & Bandukda, M. Y. (2012). Physician migration at its roots: A study on the factors contributing towards a career choice abroad among students at a medical school in Pakistan. *Globalization and Health*, 8. <https://doi.org/10.1186/1744-8603-8-43>
- Shin, S., Rachmatullah, A., Roshayanti, F., Ha, M., & Lee, J.-K. (2018). Career motivation of secondary students in STEM: a cross-cultural study between Korea and Indonesia. *International Journal for Educational and Vocational Guidance*, 18(2), 203–231. <https://doi.org/10.1007/s10775-017-9355-0>
- Smith, S. M. (2002). The Role of Social Cognitive Career Theory in Information Technology based Academic Performance. *Information Technology, Learning, and Performance Journal*, 20(2), 1–10.
- Snyder, J. (2013). Exploitation and demeaning choices. *Politics, Philosophy and Economics*, 12(4), 345–360. <https://doi.org/10.1177/1470594X13496067>
- Šorgo, A., Bartol, T., Dolničar, D., & Boh Podgornik, B. (2017). Attributes of digital natives as predictors of information literacy in higher education. *British Journal of Educational Technology*, 48(3), 749–767. <https://doi.org/10.1111/bjet.12451>
- Soykan, F., & Kanbul, S. (2018). Analysing k12 students' self-efficacy regarding coding education. *TEM Journal*, 7(1), 182–187. <https://doi.org/10.18421/TEM71-22>
- Stanko, T., Zhirosh, O., & Krasnikhin, D. (2015). Why girls with an interest in IT in high-school do not choose an IT career. *Proceedings of 2014 International Conference on Interactive Collaborative Learning, ICL 2014*, 131–137. <https://doi.org/10.1109/ICL.2014.7017760>
- Tantuco, V. (2017). Jobstreet 2017 report: Highest paying jobs for fresh grads. Retrieved January 2, 2019, from <https://www.rappler.com/life-and-style/career/164232-jobstreet-2017-fresh-graduates-report-highest-payin-g-jobs>
- Torkzadeh, G., Chang, J. C.-J., & Demirhan, D. (2006). A contingency model of computer and Internet self-efficacy. *Information & Management*, 43(4), 541–550. <https://doi.org/10.1016/J.IM.2006.02.001>
- UNESCO. (2008). Philippine Education For All 2015: Implementation and Challenges. Retrieved December 11, 2018, from http://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/philippines_efa_mda.pdf
- Uwizeyimana, V. (2018). Digital Native (ness), Mobile Technologies and Language Proficiency in Rwanda, *11*(2), 121–138.
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model : Four Longitudinal Field Studies. *Management Science*, (January 2015). <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Welsh, M. (2016). *The Philippine Labor Brokerage State and the Overseas Filipino Workers*. Retrieved from https://digitalcommons.csUMB.edu/cgi/viewcontent.cgi?article=1016&context=caps_thes_all
- World Bank. (2012). *Putting Higher Education to Work Skills and Research for Growth in East Asia*. Retrieved from http://siteresources.worldbank.org/EASTASIAPACIFICEXT/Resources/226300-1279680449418/7267211-1318449387306/EAP_higher_education_fullreport.pdf
- Zhang, W. (2007). Why IS: Understanding Undergraduate Students' Intentions to Choose an Information Systems Major. *Journal of Information Systems Education*, 18(4).

Co-Designing Multimodal Pedagogical Content Knowledge with Indonesian Teachers

Hyo-Jeong SO^{a*}, Ching Sing CHAI^b, Nur Arifah DRAJATI^c, Uma NATARAJAN^d, Hyeran LEE^a & Yue WANG^a

^a*Department of Educational Technology, Ewha Womans University, Korea*

^b*Department of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong*

^c*Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia*

^d*meriSTEM Centre, National Institute of Education, Singapore*

*hyojeongso@ewha.ac.kr

Abstract: The purpose of this study was to engage Indonesian English teachers in designing lessons for multimodal learning. Given that technological pedagogical content knowledge (TPACK) is the emerging framework that accounts for teachers' professional expertise, this study attempts to address how to support the teachers to develop TPACK through the learning by design approach. In this study, 36 Indonesian high school English teachers participated in the teacher professional development (TPD) workshop, and 28 of them completed the pre-and-post surveys that measured teachers' perceived efficacy to apply TPACK for multimodal literacy. Paired-sample t-tests revealed that all seven factors changed significantly after the workshop, indicating the positive efficacy of TPD. Overall, our findings show that when TPD is designed with the co-design principles and relevant scaffolds, it can be effective for raising teachers' competency in designing technology-integrated lessons and producing positive changes in teachers' efficacy.

Keywords: TPACK, multimodal literacy, teacher professional development

1. Introduction

In the digital world, the conception of literacy has undergone fundamental changes and reviews. Literacy as one's ability to read and write, has been broadened with the ease of producing and sharing digital media and artefacts. Given the current advances in digital technology and media, students now have access to multiple modes of communicative resources, such as video, audio, images and even hybrid forms (e.g., blogs, wikis, social media), in daily life. From pedagogical perspectives, such multiple communication sources offer both opportunities and challenges. Multimodal literacy, which is the combined use of multiple communication modes, offers rich opportunities for learners to better construct and deconstruct the meaning of messages. Advocates of multimodality in education argue that it provides learners with more situated and relevant contexts for meaning-making (Kress, 2003). On the other hand, it is also challenging for learners to understand and interpret how the different modes are subtly shaping messages. Further, multimodal literacy has challenged the traditional notion of language teaching, which mainly focuses on the acquisition of linguistic knowledge and communicative skills. The multimodal approach requires substantial design efforts from teachers to re-contextualize language learning in a media-rich world.

Given that technological pedagogical content knowledge (TPACK) is the emerging framework that accounts for teachers' professional expertise (Chai, Koh & Teo, 2018), and that multimodal communication is a desirable capacity for the 21st century school and workplace (Jenkins, 2006) with potential to enhance students' language literacy, this study attempts to engage language teachers in designing lessons for multimodal learning. This study was the first part of a 6-month long professional development process where the instructors (the first and second authors) conducted a workshop for Indonesian teachers. The aim was to equip the participants with basic knowledge and experiential understanding of multimodal meaning-making and the iterative design processes that would allow them to continue the journey when they leave the workshop.

To achieve such aims within two days, a set of design principles based on the past research on TPACK and teacher professional development (TPD) was adopted. A study conducted in Indonesia indicates that while the Indonesian teachers have positive beliefs about the importance of using technology for pedagogical innovation, they do not possess strong efficacies in designing technology-integrated learning environments (Drajati, Tan, Haryati, Rochsantiningsih, & Zainnuri, 2018). Therefore, the specific design features of the workshop included epistemic and cognitive scaffolds that could help the participants improve their beliefs and efficacy to design technology-integrated lessons based on the principles of multimodal literacy. In this paper, the effects of the TPD workshop on teachers' knowledge and skills are presented in self-reported forms of surveys and interviews. The implications of co-designing pedagogical innovations as well as directions for future research are discussed.

2. Theoretical Framework

2.1 Multimodal Literacy in Education

With the digital turn, the range of technologies available for communication has grown rapidly (Mills, 2010). The emergence of various digital media calls for reconceptualization of the meaning of literacy, which has been broadened with the proliferation of multimodal texts. More than two decades ago, the New London Group (1996) published the seminal work on multiliteracy, which foregrounded the changing nature of literacy, and the profound affordances of multiple modes of communication and texts. In essence, multimodal literacy refers to the combination of multiple semiotic modes. Here, a mode is "a socially and culturally shaped resource for meaning making" (Bezemer & Kress, 2008, p.171), and a mode of meaning includes diverse representations such as visual (still and moving), audio, gestural, and spatial, other than linguistics. Multimodal literacy engages students to create multimodal texts that ensemble different modes. Multimodal literacy, however, is not merely combining different modes, but to understand similarities and differences underlying each mode. Ajayi (2009) reported the study of engaging ESL (English as Second Language) students from diverse cultural and language backgrounds to encode and decode multimodal texts and its positive impact on students' critical thinking skills.

2.2 Teacher Learning in Professional Development

Teacher learning is a developmental trajectory spanning from initial teacher education and induction to in-service training and continuous professional development. In-service teachers enter professional development programs with certain beliefs and ideas with regards to what teaching and learning is about, built on their lived experiences. One of success factors in teacher professional development is to recognize such rich existing experiences and tacit knowledge as valuable assets. The existing body of literature highlights some core features that makes TPD effective. For instance, Garet et al. (2009) argue that teachers' knowledge and skills can be enhanced when TPD is designed to focus on content knowledge, active learning (e.g., hands-on work), and coherence. Under reform initiatives, the design of TPD programs needs to consider teachers' readiness for changes, especially when promoting unfamiliar pedagogical approaches (Twining, Raffaghelli, Albion, & Knezek, 2013).

While numerous efforts toward TPD exist, the research still points to the problem of teachers lacking content-specific teaching skills. This becomes even more problematic and complex when teachers are required to integrate new types of technologies and pedagogical approaches that create tensions with their habitus, or set of disposition (Belland, 2009). In particular, with the paradigm shift toward learner-centered pedagogy, teachers are increasingly seen as designers of learning environments, and designing for pedagogical innovation has become an important competency required for teachers (Laurillard, 2013). Despite such a paradigm shift, how to scaffold teachers' developmental trajectory to become designers of learning environments is less understood. The notion of teachers as designers also carries implications to multimodal literacy in education. In a digital age where new media become rich sources for literacy and communication, teachers' role is shifting from a mere actor of implementing

prescribed curricular to a designer of new learning experiences encompassing multimodal learning activities (Kalantzis & Cope, 2010).

2.3 TPACK and Teacher Professional Development

In this study, we adopted TPACK as an overarching framework to scaffold teacher learning about multimodal literacy. TPACK is a theory designed to account for teachers' ability to integrate technology into the curriculum (Mishra & Koehler, 2006). TPACK includes the technological knowledge (TK), pedagogical knowledge (PK) and content knowledge (CK) and the interaction of these knowledge, which include technological pedagogical knowledge (TPK), technological content knowledge (TCK) and pedagogical content knowledge (PCK). Koh, Chai, Wong and Hong (2015) postulate that teachers need to activate the various forms of TPACK to bear on the classroom pedagogical needs to design feasible lesson ideas.

To identify the relevant literature that attempted to teach in-service teachers about multimodal literacy, we used the keywords such as "multimodal", "language learning" and "teacher professional development" in the Scopus database. The search only surfaced 9 studies. Closer examination, however, revealed that none of the studies focused on developing teacher's capacities to design language learning lessons that employ multimodality as learning resources or multimodality as means for students' production of digital artefacts. Two studies that provide some relevant information for this study were conducted by Drajadi et al. (2018) and Bruce and Chiu (2015). Drajadi et al. (2018) surveyed Indonesian pre- and in-service English language teachers about their understanding of multimodal literacy and TPACK and concluded that both ideas were new to the teachers. Bruce and Chiu (2015) surveyed 240 America English language arts teachers who had completed at least one digital videos project with an open-ended questionnaire. The findings indicate that the teachers welcomed the learning experiences and appreciated the pedagogical affordances of digital videos. In sum, both studies seem to indicate that language teachers are keen to undertake multimodal teaching and learning.

However, without adequate design from teacher educators to enrich teachers' understanding of multimodality and TPACK, teachers may not be able to develop sufficient knowledge and skills to undertake such complex changes in teaching practices. Lee and Kim (2017) attempted to support preservice teachers in developing technology integration using the TPACK framework. Their study reveals that there were several gaps in the teachers' competencies. Their first design surfaced the preservice teachers' lack of pedagogical knowledge and their second design surfaced the problem that the actual teaching practice of their pre-service teachers continued to be teacher-centered. It took three iterative design cycles before they could achieve their pedagogical aims.

Learning by Design (LBD) has been suggested as an effective method to help teachers gain situated understanding about the complexity of technology-integrated lesson design (Koehler & Mishra, 2005). Kolodner et al. (1998) define LBD as "students encountering a design challenge and attempting a solution using only prior knowledge - individually and/or in small groups" (p.16). Similarly in TPD, the LBD approach structures teachers to work collaboratively to solve pedagogical problems through authentic design tasks. Koehler and Mishra (2005) argue that the LBD approach is effective for enhancing teachers' TPACK since the approach engages teachers to make the intricate connection among content, pedagogy and technology in every stage of design. LBD has also been applied as a pedagogy of multiliteracies. Cope and Kalantzis (2016) reported the LBD project as a reflexive pedagogy in multiliteracies with the knowledge processes including experiencing, conceptualizing, analyzing and applying. They argue that LBD is a relevant pedagogy in multiliteracies since it provide teachers with opportunities to reflect on how to achieve transformative learning for diverse students.

2.4 Research Purpose and Questions

TPACK emerges through the explicit and intentional connection among content, pedagogy and technology. Despite more than a decade of TPACK research, however, little is understood of how TPACK emerges and become solid contextualized knowledge. Whether TPACK is a new form of teacher knowledge or an integration of different knowledge types is a debatable issue. What is clear is the role of technology, which is not to merely support or extend but to transform existing practices for pedagogical innovations (Twining et al., 2013).

This study aims to address how to support teachers to develop TPACK through the LBD approach. The bulk of the time in the proposed workshop, hence, was devoted to closely work with the teachers to scaffold their design thinking process. The design of the workshop included specific scaffolding strategies in the learning-by-design approach. In view of the research gaps identified, this study examines the following research questions concerning TPACK and TPD.

- a) How do teachers' efficacy and beliefs about multimodal TPACK change after the workshop?
- b) How do the teachers perceive the impact of the TPD activities?

3. Method

3.1 Research Context and Participants

Thirty-six Indonesian high school English teachers participated in the TPD workshop for improving teaching with TPACK for multimodal literacy. The workshop was held in the Surakarta region for two days in May 2018, as the first part of a 6-month long professional development process. Twenty-eight teachers (24 female and 4 male) completed the pre-and-post workshop surveys that measured teachers' perceived efficacy to apply TPACK for multimodal literacy. Half of the participants was over 41 years old (46%), followed by 9 teachers who were below 30 years old, and 6 teachers in their 30's. The average of teaching experiences was 13.7 years. Overall, the majority of the participants can be described as experienced teachers in their 30s-40s with more than 10 years of teaching experiences. All participants joined the Facebook group that was created to facilitate the workshop activities such as resource sharing, presentations, peer feedback, and reflection.

3.2 Design Principles in Teacher Professional Development

Table 1 presents the core design principles underlying the workshop. Employing TPACK as an overarching framework coupled with the design principles of learning-by-design, the researchers activated relevant knowledge that the teachers have about curricula, school contexts and students. Such a co-design approach ensures that a relevant set of TPACK is foregrounded as epistemic resources to contextualize teachers' design (Koh et al., 2015).

Table 1

Design of the Teacher Professional Development

TPD Activities	Duration	Design Principles
Lecture <ul style="list-style-type: none"> • Multimodality & TPACK • Gamification & Social Media 	2 hours	<ul style="list-style-type: none"> • Knowledge inputs to address knowledge gaps (Sweller, 1994) • TPACK design framework to establish a shared framework for design (Chai et al. 2018) • With knowledge gaps and design framework addressed, cognitive overload during design process and activation of relevant knowledge for design can be achieved
Hands-on activities <ul style="list-style-type: none"> • Google site development • Mobile apps for language learning 	1 hour	<ul style="list-style-type: none"> • Just-in-time learning of technological tools to build technological knowledge (TK)
Group-based design	7 hours	<ul style="list-style-type: none"> • Learning-by-design (Mishra & Koehler, 2006) • Supporting reflective practices through scaffolding by worked examples and in-situ instructor scaffolding (Schön, 1987)

	<ul style="list-style-type: none"> Co-design with distributed expertise where both teachers' CK and researchers' TK and TPK are synthesized and transformed into TPACK (Koh, Chai & Lim, 2017). 	
Peer critique and refinement of lessons	4 hours	<ul style="list-style-type: none"> Iterative co-design (Koh et al., 2017)

The bulk of TPD was then devoted to co-design a web-based learning environment in the Google Site platform (see Figure 1) in small groups (5-6 members) with researchers' in-situ scaffolding and worked examples as another form of scaffolding. Learning how to design is essentially a process of being a cognitive apprentice and engaging in reflective practices (Schön, 1983; 1987). From the perspective of cognitive load (Pass, Renkl & Sweller, 2003), working examples and distributed expertise (i.e., working in teams) are both likely to assist in reducing the complexity of learning to design language learning lessons with multimodal expressions.

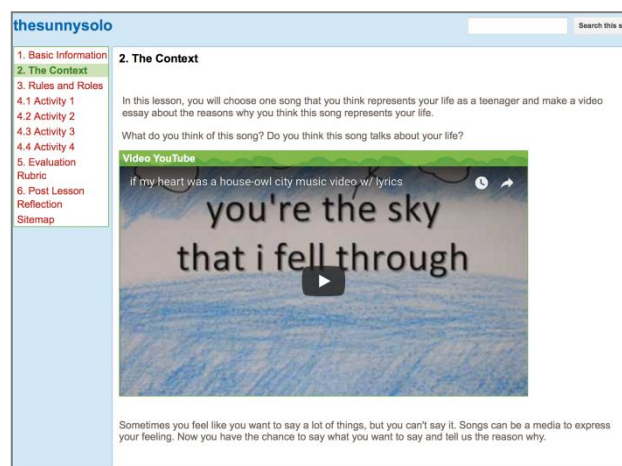


Figure 1. Example of a Google Site lesson created by the participant

In this study, we view a lesson design as a form of concretized TPACK that embodies teachers' integration of content, pedagogy, and technology, coupled with other contingent considerations such as student characteristics and classroom environments. TPACK is a form of contextualized knowledge creation as each lesson is designed for a specific class with unique and collective characteristics of diverse learners. The current TPACK research is moving towards the direction of creating lessons for specific technology, pedagogy and content areas (Chai, Koh, & Tsai, 2016). For the present study, the specific content knowledge (CK) lies in social semiotics (theory of meaning making), and multimodal media with the gamified approach as technological knowledge (TK) and technological pedagogical knowledge (TPK).

3.3 Data Collection and Analysis

To investigate teachers' perceived efficacy of TPACK for multimodal literacy, we administered the survey instrument with 40 items that were adapted from the existing instrument developed and validated by one of the authors for the purpose of this study. Table 2 shows seven factors measured in the survey with sample items and Cronbach's α values. Reflecting this study's focus on multimodal literacy, the factor 'Content Knowledge (CK)' was designed to measure teacher knowledge of social semiotics (theory of meaning making) whereas 'Pedagogical Knowledge (PK)' refers to knowledge of supporting 21st century learning dimensions (e.g., collaborative learning, self-directed learning). Items under 'Technological Knowledge' (TK) were designed to measure knowledge about digital media tools. Other factors such as TPK, TCK, and TPACK were created to measure the intersection of content, pedagogy and technological knowledge. Pedagogical Content Knowledge (PCK) was not measured since the workshop focused on the integrations of technology, and PCK does not include a technological dimension of knowledge. Instead, we included the factor 'Beliefs about the new culture of learning' to measure epistemic beliefs that teachers hold about the changing nature of teaching and

learning. The items were measured on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The reliability measures of each factor in Cronbach's α were in an acceptable range (all above .90). Survey data was analyzed using the SPSS, and the corresponding paired-sample t-tests were conducted to investigate statistically significant differences between the pre- and post-workshop responses.

Table 2

Factors and sample items

Factor	Number of items	Sample items	Cronbach's α
Content Knowledge (CK)	6	I have sufficient knowledge about how semiotic modes (e.g., linguistic, audio, visual, gestural and spatial) work.	.97
Pedagogical Knowledge (PK)	6	I am able to stretch my students thinking by creating challenging tasks for them.	.97
Technological Knowledge (TK)	5	I am able to use social media (e.g., Facebook, Edmodo, Twitter, etc.).	.92
Technological Pedagogical Knowledge (TPK)	5	I am able to use technology to help my students develop diverse perspectives for authentic issues.	.90
Technological Content Knowledge (TCK)	4	I can use specialized software programs created for a multimodal text analysis (e.g., video annotation tools, etc.).	.90
TPACK	7	Using appropriate technology, I can stimulate my students to read and write critically about current affairs represented in media.	.97
Beliefs about the new culture of learning	7	Today's learners should be able to remix relevant resources to publish their ideas.	.93

In addition, we conducted the focus group interview with nine teachers after the workshop to better understand their perceptions and experiences about multimodal pedagogical content knowledge and the co-design activity. The interview was semi-structured with the following guiding questions: (a) can you describe your experience in these two days?; (b) why did you decided to attend this workshop?; and (c) how is your belief about teaching language changed? The interview lasted approximately for 30 minutes and audio-recorded for transcription. The researchers analyzed the content of the transcribed data to identify emerging themes.

4. Results

4.1 Survey Data

To understand the first research question concerning whether the workshop was effective in raising the teachers' perceived efficacy in multimodal lesson design and their beliefs, paired-sample t-tests were conducted between the pre- and post-workshop survey. As documented in Table 3, all seven factors changed significantly after the workshop. In the pre-workshop survey, the mean scores of factors range from 4.51 to 6.04. The mean scores after the workshop ranges from 5.88 to 6.42 where five factors were above 6.00. Relatively large gains were observed in the factors CK (pre-workshop $M = 4.51$; post-workshop $M = 5.88$) and TCK (pre-workshop $M = 4.42$; post-workshop $M = 5.78$). On the other hand, the teachers' beliefs about the new culture of learning were already high in the pre-workshop survey, and hence showed the smallest gain. All six factors except the Belief factor showed the high effect size at .98 and higher. In general, the Cohen's d value is interpreted as a medium effect size at .5 and a larger effect size at .8 or higher (Cohen, 1992).

Table 3

Descriptive analyses and results of paired-sample t-test (N = 28)

Measured factors	Pre-workshop survey		Post- workshop survey		t-test	Cohen's d
	M	SD	M	SD		
CK	4.51	1.19	5.88	.74	6.74*	1.38
PK	5.08	1.23	6.19	.67	4.98*	1.12
TK	4.83	1.40	6.01	.80	5.28*	1.03
TPK	5.18	1.04	6.17	.54	4.96*	1.19
TCK	4.42	1.35	5.78	1.00	5.17*	1.14
TPACK	5.03	1.24	6.07	.84	4.67*	0.98
Belief	6.04	.94	6.42	.56	2.32*	0.49

* $p < .001$

4.2 FGI Data

Overall, the FGI revealed that the teachers see the workshop as relevant and they are beginning to embark on the design developmentary trajectory. They seem to recognize that developing their competencies in designing multimodal language learning is a worthy pursue and a means for personal growth. Most teachers attended the workshop as they were attracted by the novel terms TPACK and multimodal, as the teachers put it “What kinds of thing is this TPACK... it hasn’t been applied yet, so far for English teachers” (MT). “I think... I will be more familiar with... multimodal TPACK” (CT). Some of these teachers travelled far (7-9 hours land transport) to attend the workshop. These topics were attractive because the teachers recognized that their students know more about technology and students are reporting that they are learning English because of games such as “Mobile Legends” (ST). The teachers believe that “we are teaching in the 21st century now so I need to change the way I teach my students.” (ST).

The teachers reported that they knew gamification and some technological tools, but they either use games without computers or Facebook for socialization. In other words, they seem to lack experiences in connecting technological, pedagogical or content knowledge. One teacher reported that “we got our model lesson plan from our government. When you showed us the latest one, I tell myself I have to insist on getting out of my comfort zone. It is going to be very challenging, but that’s fine” (UT). Technology does not seem to be foregrounded. While the teachers are acutely aware of the gaps between them and the students’ in terms of technological skills, their concern is about designing ‘meaningful’ learning experiences and tasks that align with the standards. The changes that the teachers experienced seem to be best captured by the following quote:

“Learning language is not merely doing what the teacher said but is about doing something meaningful. And I am trying to analyze how he (the instructor) develops his task. Several years ago, I took task-based design as my thesis. And I know what a meaningful task is. But it is quite difficult to develop such kinds of task in two days. Actually, I don’t worry about technology because I can ask my students. But what worries in my mind is how I can decide the tasks... something very different from what I always do in my teaching. I mean designing the task that it is really meaningful for the students. And relating the task with the students’ interest. I think the important thing is designing the task” (CT).

Nonetheless, the teachers also reported that they may face challenges in school because of the school infrastructure and policies such as the low bandwidth and banning the use of mobile phones in classrooms. When the researchers suggested that they could do action research to show results, one of them responded “It is not so simple. It is a culture thing”. A teacher representative in the government multiliteracy program also expressed the following concern; “the basic and the fundamental literacy for Indonesia is not really there yet. Literacy is what we need to be concerned about. I just want to be sure that the regulation about multimodality will be suitable for our students. not for creating the jumping programs.”

In sum, while the workshop was welcomed and it started the teachers in connecting relevant knowledge which was manifested in the Google Site lessons they completed at the end of the workshops, they forecasted that they would face other systemic problems.

5. Discussion and Conclusion

This study presents the specific design features of TPD informed by TPACK and teacher learning literature and the impact of the TPD on teachers' knowledge and skills in self-reported forms. It also exemplifies the design principles of TPD that aimed to promote pedagogical approaches that teachers are rather unfamiliar with. From the TPACK perspective, multimodal literacy is the intersection of content, pedagogy and technology. The teachers who participated in the workshop had to understand multimodal literacy and how digital media and technology are related to multimodal literacy in the context of language learning.

This study demonstrates one possible design of TPD, given the condition where teachers are struggling to understand new pedagogical practices, coupled with the lack of relevant internal expertise and support systems. The workshop was successful in promoting efficacies and providing the basis for the developmental trajectory. The lesson design that teacher created in the workshop was complete and workable for the classroom implementation. In particular, the co-design process with multiple scaffolds that aimed to reduce teachers' cognitive load was perceived to be important, coupled with the peer critique that promoted collegial knowledge sharing. Undoubtedly, a workshop, which is a structural approach often scheduled after school or over weekends/school breaks occurring outside of teachers' school context, is the most common format of professional development in the literature. The workshop format, however, has been also criticized as being ineffective in promoting deeper substantial changes in teachers' practices (Garet et al., 2001). While we do not argue that short-term workshops can produce deep changes, our findings show that when TPD is designed with the co-design principles and relevant scaffolds, it can be effective for raising teachers' competency in designing technology-integrated lessons and producing positive changes in teachers' efficacy, which is the first level of change needed to launch the teachers on a developmental trajectory.

Some limitations of this study should be noted. This study mainly examined the perceived beliefs and efficacy of teachers through the self-reported survey and interviews, and did not examine the actual enactment of the designed lessons. Further, the effects of professional development on student achievement were not investigated in this study. As mentioned earlier, this study was a part of the 6-month long professional development process, and we plan to continuously support the teachers' development trajectory through subsequent work, including the support of teachers' lesson enactment and the examination of student achievement, especially their digital artefacts that embody multimodal literacy practices. We also plan to administer the questionnaire again after six months in order to analyze teachers' lesson designs and to observe the enactment of designed lessons.

Acknowledgements

This research was supported by the HEAD foundation, Singapore, and the international collaboration research fund by Ewha Womans University, Korea.

References

- Ajayi, L. (2009). English as a second language learners' exploration of multimodal texts in a junior high school. *Journal of Adolescent & Adult Literacy*, 52(7), 585-595.
- Belland, B. R. (2009). Using the theory of habitus to move beyond the study of barriers to technology integration. *Computers & Education*, 52(2), 353-364.
- Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166-195.
- Bruce, D. L. & Chiu, M. M. (2015). Composing with new technology: Teacher reflections on learning digital video. *Journal of Teacher Education*, 66(3), 272-287.

- Chai, C. S., Hwee Ling Koh, J., & Teo, Y. H. (2019). Enhancing and Modeling Teachers' Design Beliefs and Efficacy of Technological Pedagogical Content Knowledge for 21st Century Quality Learning. *Journal of Educational Computing Research*, 57(2), 360-384.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2016). A review of quantitative measures of Technological Pedagogical Content Knowledge (TPACK). In Mary Herring, Matthew Koehler, & Punya Mishra (Eds.), *TPACK Handbook V2.0: TPACK Research and Approaches* (pp. 97-116). New York: Routledge.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cope, B., & Kalantzis, M. (Eds.). (2016). *A pedagogy of multiliteracies: Learning by design*. Springer.
- Drajati, N. A., Tan, L., Haryati, S., Rochsantiningih, D. & Zainnuri, H. (2018). Investigating English language teachers in developing TPACK and multimodal literacy. *Indonesian Journal of Applied Linguistics*, 7(3), 575-582.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU press.
- Kalantzis, M., & Cope, B. (2010). The teacher as designer: Pedagogy in the new media age. *E-learning and Digital Media*, 7(3), 200-222.
- Koehler, M. J., & Mishra, P. (2005). Teachers learning technology by design. *Journal of Computing in Teacher Education*, 21(3), 94-102.
- Koh, J. H. L., Chai, C. S., Benjamin, W., & Hong, H. Y. (2015). Technological Pedagogical Content Knowledge (TPACK) and design thinking: A framework to support ICT lesson design for 21st century learning. *The Asia-Pacific Education Researcher*, 24(3), 535-543.
- Koh, J. H. L., Chai, C. S., & Lim, W. Y. (2017). Teacher professional development for TPACK-21CL: Effects on teacher ICT integration and student outcomes. *Journal of Educational Computing Research*, 55(2), 172-196.
- Kolodner, J. L., Crismond, D., Gray, J., Holbrook, J., & Puntambekar, S. (1998, December). Learning by design from theory to practice. In *Proceedings of the international conference of the learning sciences* (Vol. 98, pp. 16-22).
- Kress, G. (2003). *Literacy in the new media age*. New York, NY: Routledge.
- Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.
- Lee C.-J. & Kim, C. (2017). A technological pedagogical content knowledge based instructional design model: a third version implementation study in a technology integration course. *Educational Technology: Research and Development*, 65, 1627-1645.
- Mills, K. A. (2010). A review of the "digital turn" in the new literacy studies. *Review of Educational Research*, 80(2), 246-271.
- Mishler, P. & Koehler, M. J. (2006). Technological pedagogical content knowledge: a framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Education Review*, 66(1), 60-92.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1-4.
- Schön, D. (1983). *The reflective practitioner: how professionals think in action*. London: Temple Smith.
- Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312.
- Twining, P., Raffaghelli, J., Albion, P., & Knezek, D. (2013). Moving education into the digital age: the contribution of teachers' professional development. *Journal of Computer Assisted Learning*, 29(5), 426-437.

Teaching conceptions and teaching context affect the enacted practice of online teachers

Deniese COX^{a*} & Sarah PRESTRIDGE

^a*PhD Candidate, Griffith University, Australia*

^b*Senior Lecturer, Griffith University, Australia*

*deniese.cox@griffithuni.edu.au

Abstract: This study examined effects on enacted teaching practice in an exclusively online context within a large vocational education sector in Australia. A wide-ranging exploratory survey instrument was constructed with questions drawn from validated surveys and a review of literature in the field of online pedagogy. The study found that both teaching conceptions and teaching context affect enacted practice in this online vocational education teaching context. Notably, some of the teaching context factors investigated in this study were found to cause enacted practice to shift away from what these teachers perceived as good online teaching. This study contributes to our understanding of online teaching practices and what affects it.

Keywords: Online education, conceptions of teaching, vocational education and training

1. Introduction

Online education is one of the fastest growing fields of education across sectors (Richardson, Besser, Koehler, Lim, & Strait, 2016; Sun & Chen, 2016). However, that growth has brought challenges for institutions as they balance the tension between demand for online education with lower successful completion rates and lower student grades relative to those secured in traditional face-to-face teaching contexts (Capra, 2011; Drouin, Stewart, & Van Gorder, 2015; Gaytan, 2015).

Teachers are at the heart of online education, yet teachers are often unprepared for the online teaching environment and its fundamental differences from teaching face-to-face (Banas & Velez-Solic, 2013). Palloff and Pratt (2013) report that teaching online is more than knowing how to use technology or how to transfer materials to an online platform. Rather, teaching online requires different pedagogy and skills than teaching face-to-face (Terblanché, 2015); it is not something simply added to existing practice. Today's online teachers have little experience of the online learning environment (Niess, 2011), and previous studies have evidenced that knowledge of face-to-face teaching does not directly translate to knowing how to teach online (Seaton & Schwier, 2014; Terblanché, 2015). This means that not all teachers inherently possess the skills to teach online (Yen, Lo, Lee, & Enriquez, 2018) yet, as reported by Banas and Velez-Solic (2013), professional development for online teachers lags behind the growth of online education. Consequently, transitioning to online education can be confronting for experienced teachers as they struggle with becoming novices in the online teaching context (Redmond, 2011) and struggle to communicate their objectives and meanings online (Kearns, 2012). Furthermore, because online education requires a reconceptualisation of the teacher role (Rodrigues, Almeida, Figueiredo, & Lopes, 2019), teachers also go through a process of understanding what good pedagogy is within an online teaching context. Therefore, it is vitally important to understand clearly what a teacher conceptualises as good online pedagogy as this affects their practice, especially if they are coming from a face-to-face setting.

Conceptions of teaching are an indicator of what a teacher thinks they should do, or would prefer to do, in a teaching context (Clark & Peterson, 1986), and teaching conceptions are known to affect enacted practice (Conrad, 2012). Sun and Chen (2016) conceptualise good online pedagogy as characterised by constructivist, student-centred philosophies where facilitation of collaborative learning opportunities are valued. Practices associated with such pedagogy can be understood from Bain's description of '*what the best teachers do*' (2004) which has been updated for online education by Brinthaup, Fisher, Gardner, Raffo, and Woodard (2011). Those works purport that good online

pedagogy is where teachers stimulate intellectual development, foster student engagement, and build rapport with students. Nine practices associated with those principles are presented in Table 1.

Table 1

Bain's 'what the best teachers do' (2004) with teaching practices updated by Brinthaupt et al. (2011) for online education

Category	Nine Online Teaching Practices
Stimulate intellectual development	Facilitate discussion forums where students explore concepts and develop deep knowledge together. Utilise technology for real-time engagement with groups of students. Utilise engaging tasks students will find interesting. Utilise a variety of technologies such as videos or podcasts.
Foster student engagement	Create a friendly, social atmosphere where deep learning is encouraged. Develop group cohesiveness, helping students work together for mutual benefit. Use discussion forums to facilitate social interaction between students.
Build rapport with students	Use introductory videos or other self-disclosure methods to humanise self to students. Consciously build rapport with each individual student.

An additional factor that affects enacted practice is the teaching context. The real or perceived needs, affordances, or limitations within teaching contexts can cause teachers to adapt their practice toward or away from their ideal (Eley, 2006; Samuelowicz & Bain, 1992). This means that different teaching contexts can result in a teacher enacting different practices (Lindblom-Ylänne, Trigwell, Nevgi, & Ashwin, 2006). Example teaching context factors that affect enacted practice within online education include teacher workload (Goldman, 2011), class size (Sorensen, 2014; Taft, Perkowski, & Martin, 2011), the discipline being taught (Arbaugh, Bangert, & Cleveland-Innes, 2010), and course duration (Akyol, Vaughan, & Garrison, 2011).

A final contributor to the enormous complexities within online education is that not all education sectors receive the same research attention. Much of the academic research undertaken to understand and improve online education is set within higher education contexts (Jaggars & Xu, 2013; Nguyen, 2015) and Australia's vocational education sector currently lacks research about how its teachers engage with technology for teaching (Chang, 2016). This is an important gap because vocational education and training (VET) is Australia's largest education sector (Atkinson & Stanwick, 2016). Known as VET, this post-school sector provides work-ready skills for employment within entry-level jobs and highly skilled professions (Productivity Commission, 2019). Future jobs growth in Australia requires 53% of occupation qualifications to be at VET level (Productivity Commission, 2011). As a critical contributor to Australia's labour market, to support widening industry demands VET must continue expanding (Wheelahen & Moodie, 2011). Importantly though, Reeson, Mason, Sanderson, Bratanova, and Hajkowicz (2016) report that a major enabler for VET to meet its current and future challenges is an expansion of online education.

As reported by Bliuc, Casey, Bachfischer, Goodyear, and Ellis (2012), VET is an already complex teaching context and the introduction of online education adds further complexity. VET is bounded by a competency-based-training (CBT) framework making it fundamentally different to other forms of post-secondary education (Fowler, 2017). Furthermore, like other education sectors, VET too has observed lower successful completion rates for students studying online (DET, 2016). Unfortunately, those lower success rates, along with reporting on parliamentary inquiries into VET, has linked unscrupulous VET providers with the inaccurate suggestion that online education is a poor substitute for face-to-face teaching (CEDA, 2016).

This study draws together the complexities of VET and the shift to online teaching by examining what online VET teachers conceive as good practice online and what shapes these practices

as the teachers move from conceptualisation to enactment. In investigating this, an exploratory survey was constructed to examine both teachers' perceived beliefs about online teaching as well as their perceived online teaching practices.

2. Methodology

To investigate the pedagogy of online VET teachers and what affects their enacted practice, a quantitative survey was designed and implemented. Questions for the wide-ranging survey were developed from reviews of online education and VET studies, as well as the incorporation of previously validated survey tool components. As an exploratory study, the focus of this survey was to establish what teachers believe to be good online teaching and how they perceive their enacted practice is affected by factors within their teaching context.

2.1 Survey Instrument

To investigate whether online VET teacher beliefs align with the nine student-centred online teaching practices that were presented previously (see Table 1 above), teachers were asked to rate the importance of these practices within online VET. The validated survey design of Owens (2015) resulted in a profiling of teachers teaching conceptions along a five-point scale from teacher-centred to student-centred. This approach was adopted here. Then, using the same nine teaching practices, teachers were asked in the survey to report their frequency of enacting those practices. This series of nine questions provided data on what practices teachers enacted more frequently than others. In effect, survey data was used to build a teacher profile of both teaching conceptions and (perceived) enacted practice. This data was analysed to identify any differences, similarities and nuances between the two perceptions. Separately within the survey, teachers were asked whether particular teaching context factors affect their practice.

2.2 Survey Participants

The VET teachers voluntarily participating in this survey taught exclusively online with no face-to-face teaching responsibilities within a large public VET provider in Queensland, Australia. They were purposefully drawn from an established team of 66 fully online teachers, teaching more than 40,000 online subjects a year. After survey piloting, and ethical approvals from the researcher's university and the host site, a survey token and weblink was sent to those 66 target participants. Of the 46 teachers who subsequently completed the survey, 76% were female, and 89% were full-time employees, ranging in age from 32 to 66 years (\bar{x} =51). Participants represented a variety of teaching experience ranging from 0 to 40 years face-to-face (\bar{x} =11 years), and 0.5 to 15 years online (\bar{x} =4.5 years). Six teaching disciplines were represented, and the largest group of respondents came from early childhood (33%) followed by business (28%).

2.3 Analysis

Survey data was analysed using the statistical software package SPSS. The survey featured Likert-type and Likert-scale questions that both utilise scales to capture a "belief, preference, judgement, or opinion" (Warmbrod, 2014, p. 31) of participants. Likert-type questions are single questions and were analysed here individually; Likert-scale questions are designed as a series and as such were analysed here through their composite result (Boone & Boone, 2012) achieved by calculating the mean-item score (Warmbrod, 2014). Many of the questions in this survey were categorical and ordinal. For analysis purposes, numbers were assigned to represent the answer options offered, however this did not indicate one category was better than another – all scale items were equally weighted (Warmbrod, 2014). Numerical values for negatively worded questions were reversed prior to analysis (Ary, Jacobs, Sorensen, & Razavieh, 2010; Warmbrod, 2014).

Non-parametric tests were utilised in the analysis because they do not assume the sample is normally distributed (Field, 2013) and are therefore ideal for small sample sizes, especially when

categorical and ordinal data is present (Pallant, 2016). The Wilcoxon Signed Rank Test was utilised to compare participants under two different conditions (Field, 2013; Pallant, 2016), such as when comparing conceptions of teaching and enacted practice responses. The results of this test were reviewed adjacent to Friedman Tests to confirm consistency of result. The magnitude of the effect was interpreted using Cohen (1998) as cited in Pallant (2016): 0.1 = small effect, 0.3 = medium effect, 0.5 = large effect.

Separately, the Kruskal-Wallis H Test is suitable for comparing differences between more than two groups by using ranks (Field, 2013; Pallant, 2016). It was utilised in this analysis when comparing participants by group such as by age, years teaching experience, or discipline being taught. Such analysis identified if, for example, participants with higher level teaching qualifications had similar answers within a question or question set. This paper reports only those results relevant to the topic and that yielded a significant result, seeking <0.05 .

The internal consistency of the survey items were tested using Cronbach's alpha coefficient (Pallant, 2016). Scales reported in this paper were deemed reliable and have alpha coefficients of between 0.67 and 0.91. However, an acknowledged limitation of this study is its small sample size. The survey has internal validity in that it accurately measures what it was intended to measure (Cohen, Manion, & Morrison, 2011; Creswell, 2014). However, with only 46 useable responses, the findings are reported to be informative in nature and not indicated as generalisable or transferable across other online teaching contexts.

3. Results

Teacher demographics were tested to identify relationships with enacted practice, and no relationships were found. In other words, how teachers perceived the way they teach online was not affected by factors such as age, years of service, and teaching qualification. However, as presented next, enacted practice was found to be affected by conceptions of teaching and by factors within the teaching context.

3.1 Enacted Practice Affected by Conceptions of Teaching

The summated responses from teachers ranking the importance of nine student-centred teaching practices developed from literature found 71.7% of teachers aligned with the conception that good online pedagogy is student-centred. Three of the nine teaching practices within this set relate to student-to-student interaction and notably these practices were rated least important by the online VET teachers. This is illustrated in Figure 1, where the nine teaching practices are sorted highest to lowest mean. The three teaching practices at the bottom of the figure are considered least important by teachers, as indicated by the shorter blue line.

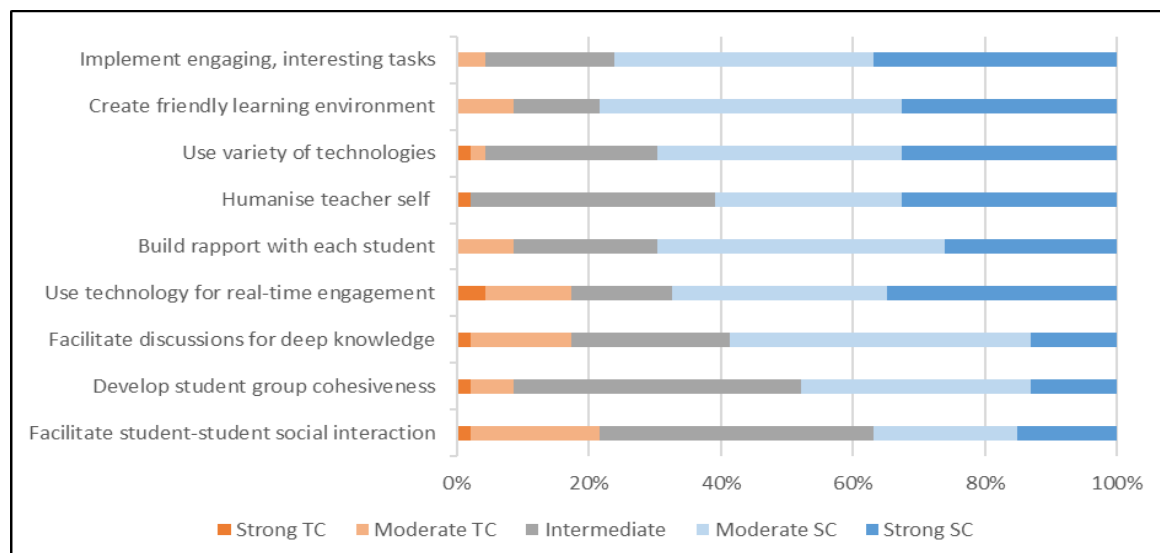


Figure 1. Conception of teaching responses, presented highest to lowest mean.

The summated responses for each teacher enabled them to be profiled as having either teacher-centred, intermediate, or student-centred conceptions of teaching. Differences between these three teacher profiles were then tested for their effect on enacted practice. Teachers with a student-centred conception of teaching reported a higher frequency ($Md=27$) of enacting the practice of using discussion forums to facilitate student-to-student social interaction than teachers with an intermediate ($Md=14$) or teacher-centred ($Md=17$) profile ($H(2) = 7.78, p = 0.02$). Teachers with a student-centred conception of teaching also reported a higher frequency ($Md=26$) of enacting the practice of creating a friendly, social learning environment than teachers with an intermediate ($Md=15$) or teacher-centred ($Md=19$) profile ($H(2) = 6.30, p = 0.04$). Finally, teachers with a student-centred conception of teaching reported a stronger desire ($Md=25$) for more training about how to effectively utilise technology to teach their discipline than teachers with an intermediate ($Md=16$) or teacher-centred ($Md=14$) profile ($H(2) = 7.42, p = 0.02$).

The summated profile of enacted practice was then analysed and compared to the conception of teaching profiles. As illustrated in Figure 2, enacted practice is less-student centred than teaching conceptions, and 45.7% of teachers were found to enact intermediate practice.

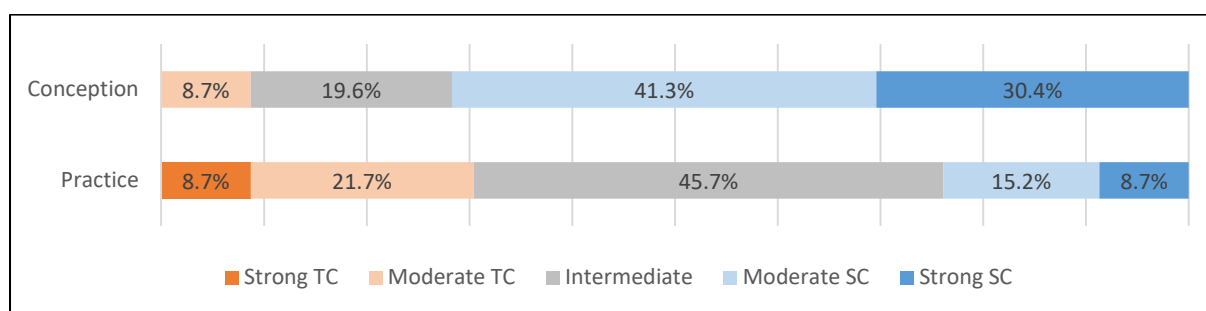


Figure 2. Comparison of conception of teaching profiles with enacted practice profiles.

The difference between perceived conception and practice for four question pairs was found to be statistically significant, with a large effect size ($r=0.5$). These practices are utilising engaging tasks, creating a friendly learning environment, developing group cohesiveness, and using technology for real-time engagement. Two questions resulted in a significant difference between conception and practice with a medium-to-large effect size ($r=0.4$). These are utilising a variety of technologies, facilitating discussions for knowledge development, and humanising self such as with introductory videos. The final two questions, facilitating discussions for student-student social engagement and consciously building rapport with students, reported no statistically significant difference between conception and practice.

Differences between conception and practice indicate that an external factor also affects the enacted practice of this sample of online VET teachers. As proposed in previous literature, factors within the teaching context can affect practice to be enacted that is different from conceptions of teaching, and this was investigated next.

3.2 Enacted Practice Affected by Teaching Context

In this section, eight teaching context factors were identified from literature and investigated in this survey of an online VET teaching context. Of those, five were identified as affecting enacted practice.

Teacher workload was reported by 66.7% of teachers in this study as *often* or *always* preventing them from enacting a teaching practice they believe would be beneficial. Teachers further reported that their three highest workload tasks are assessing student competency (marking assessments), providing direct assessment help, and doing administration. Teachers reported they would most like to spend more time actively facilitating learning and building rapport with students, and they would like to spend less time on technological issues and administration.

Both small and large class sizes were investigated for their effect on practice. 52.6% of teachers reported that small class sizes affect practice, however the reported nature of that effect is a mix of positive (31.6%), negative (7.9%), or both positive and negative (13.2%). In contrast, large class sizes

were reported by 58.8% of teachers as negatively affecting practice and 17.6% of teachers as both positively and negatively affecting practice.

Next teachers reported whether complying with the strictly enforced CBT curriculum affected their practice. Firstly, teachers reported that complying with CBT curriculum *sometimes* (51.2%) or *often* (14.6%) prevented them from enacting a teaching practice they believe would be beneficial. Secondly, teachers reported that complying with CBT curriculum *sometimes* (46.5%) prevented teaching a concept they believe is important for their students to learn.

The number of students per teacher in this study ranged from 10 to 310 students ($\bar{x}=131$) and the number of students per teacher was found to have an effect on one conception of teaching indicator and one enacted practice indicator. Teachers with >175 students were more likely than teachers with less students to agree that with the student-centred conception that students should be helped to learn for themselves rather than be told ($H(2) = 7.70, p = 0.02$). Teachers with <75 students were more likely than teachers with higher student numbers to report more frequently enacting the student-centred practice of using technology to engage with students in real-time ($H(2) = 7.55, p = 0.02$).

Analysis of data identified that the discipline being taught affects enacted practice. I.T. teachers were more likely to report a higher frequency of using discussion forums to facilitate student-student social interaction ($H(5) = 11.26, p = 0.04$), of using technology for real-time student engagement ($H(5) = 22.00, p = 0.00$), and for getting students to participate in online discussions ($H(5) = 12.42, p = 0.03$). Justice and government teachers were more likely to report a higher frequency of facilitating discussion forums for students to explore concepts and develop deep knowledge together ($H(5) = 12.63, p = 0.03$). Accounting teachers were more likely to report a higher frequency of ensuring students are well skilled in the subject competencies ($H(5) = 17.16, p = 0.00$).

Teachers in this study reported being responsible for teaching between two and 57 online classes at any one time ($\bar{x}=12$). Enacted practice was not found to be affected by the number of classes per teacher.

Literature has previously reported that professional development is lacking for VET teachers and is not always relevant for online teachers. These perceptions are not supported by this group of online VET teachers who *agreed* or *strongly agreed* that professional development is regularly available (73.9%), that it is relevant to their online teaching role (54.3%), and that it is high quality (56.5%). Enacted practice was not found to be affected by availability, relevance, and quality of professional development in this context.

Finally, course duration was not found to affect practice in this teaching context and 51.4% confirmed that practice is neither positively or negatively affected by course duration.

3.3 Summary

In summary, data analysed evidences that enacted practice is affected by conceptions of teaching. This is highlighted by teachers with teacher-centred, intermediate, and student-centred profiles reporting different frequencies of enacting some students-centred practices. Importantly, teachers with a student-centred teaching conception reported more frequently using discussion forums and creating a friendly learning environment. Those teachers also reported a stronger desire for more training regarding using technology to teach their discipline. Notably, all 46 teachers in this study reported that their enacted practice is less student-centred than their conception of teaching indicating additional external factors affect practice.

Eight teaching context factors from literature were investigated and five were found to affect practice within this teaching context. Teachers reported that workload, class sizes, and compliance with the CBT curriculum all directly affect their practice. Data analysis found that enacted practice is also affected by the number of students per teacher and the discipline being taught. For example, teachers with the smallest number of students were more likely to report a higher frequency of enacting the student-centred practice of using technology for real-time student engagement. Differences between frequency of enacted practice were also identified among teachers from different teaching discipline groups.

Enacted practice was not found to be affected by teacher demographics, the number of active teaching subjects, course duration, or the availability, relevance, and quality of professional development.

4. Conclusion

There is a deficit in our understanding of how teachers perceive and enact online pedagogy (Prieto-Rodriguez et al., 2016; Sun & Chen, 2016). This deficit extends to our understanding of the tensions that result in online education being perceived by teachers as a challenging teaching environment, and to our understanding of how those challenges affect enacted practice. The under-researched online education context of this study is Australia's VET sector which requires a growth in the quantity and quality of online education to support its continued expansion, which is itself necessary to support jobs growth.

This small (n=46) exploratory study investigated what online VET teachers conceive is good online pedagogy and the alignment of their enacted practice with that conception. Enacted practice was found to be affected by conceptions of teaching, and teachers with student-centred conceptions reported more frequently enacting some student-centred practices than teachers with intermediate or teacher-centred conceptions. Analysis of the data suggested that while the sampled online VET teachers report aligning with conceptions that good online pedagogy is student-centred, their enacted practice is substantially less so. While 71.7% of teachers aligned with student-centred conceptions of teaching, only 23.9% evidenced student-centred enacted practice profiles. Five factors within this online VET teaching context were identified as affecting enacted practice. These are teacher workload, class size, compliance with CBT curriculum, the number of students per teacher, and the discipline being taught. The differences between teaching conceptions and enacted practice found in this study reflects a compromise between conceptions of teaching and the teaching context navigated by teachers (Norton, Richardson, Hartley, Newstead, & Mayes, 2005) and is why teaching context factors are deemed essential to any investigation of teaching (Clark & Peterson, 1986).

For researchers, although the small sample size within this study means the findings are not generalisable across other online teaching contexts, the survey is scalable. Broadening the application of this survey will contribute more to our understanding of what teachers conceive as good practice online and what shapes these practices as teachers in different online teaching contexts move from conceptualisation to enactment. This study is ongoing and the findings reported above have also informed the focus of subsequent data collection strategies. The second stage of data collection includes observation of teaching practices to map the range of online teaching practices enacted and the range of technologies actively utilised by teachers. The third stage of data collection includes interviews with online VET teachers to more deeply understand how they understand the nature and purpose of their teaching role. Findings from these stages will be reported in due course.

The employing organisation who provided access to online VET teachers for this study are navigating both the continued expansion of online education and the increased demand for its enhanced quality and effectiveness. That institution intends to use the results of this study to review their online teacher hiring, training, and management practices, and to better understand online education as more than simply a different mode of delivery. Their intent is that understanding and responding to factors that affect practice will support the enactment of online education that aligns to what is currently established in the literature as good online pedagogy for teachers, thereby subsequently enhancing student learning outcomes.

References

- Akyol, Z., Vaughan, N., & Garrison, D. R. (2011). The impact of course duration on the development of a community of inquiry. *Interactive Learning Environments*, 19(3), 231-246. doi:10.1080/10494820902809147
- Arbaugh, J. B., Bangert, A., & Cleveland-Innes, M. (2010). Subject matter effects and the Community of Inquiry (CoI) framework: An exploratory study. *Internet and Higher Education*, 13, 37-44.
- Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education*. In. Retrieved from <http://www.modares.ac.ir/uploads/Agr.Oth.Lib.12.pdf>
- Atkinson, G., & Stanwick, J. (2016). *Trends in VET: Policy and participation*. Retrieved from https://www.ncver.edu.au/__data/assets/pdf_file/0017/60722/Trends-in-VET.pdf

- Australian Department of Education and Training [DET]. (2016). *Redesigning VET FEE-HELP: Discussion paper*. Australian Government Retrieved from <https://docs.education.gov.au/documents/redesigning-vet-fee-help-discussion-paper>.
- Bain, K. (2004). *What the best college teachers do*. Cambridge, Mass: Harvard University Press.
- Banas, J. R., & Velez-Solic, A. (2013). Designing effective online instructor training and professional development. In *Virtual Mentoring for Teachers: Online Professional Development Practices* (pp. 1-25). Hershey, PA, USA: IGI Global.
- Baran, E. (2011). The transformation of online teaching practice: Tracing successful online teaching in higher education. (3472990 Ph.D.), Iowa State University, Ann Arbor. ProQuest Central; ProQuest Dissertations & Theses Global database.
- Bliuc, A.-M., Casey, G., Bachfischer, A., Goodyear, P., & Ellis, R. A. (2012). Blended learning in vocational education: teachers' conceptions of blended learning and their approaches to teaching and design. *The Australian Educational Researcher*, 39(2), 237-257. doi:10.1007/s13384-012-0053-0
- Boone, H. N. J., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, 50(2).
- Brinthaupt, T. M., Fisher, L. S., Gardner, J. G., Raffo, D. M., & Woodard, J. B. (2011). What the best online teachers should do. *MERLOT Journal of Online Learning and Teaching*, 7(4), 515-524.
- Capra, T. (2011). Online education: Promise and problems. *MERLOT Journal of Online Learning and Teaching*, 7(2), 288-293.
- Chang, W. W. (2016). *Digital competence and professional development of vocational education and training teachers in Queensland*. Queensland University of Technology, Brisbane. Retrieved from http://eprints.qut.edu.au/95088/1/Wei-Wei_Chang_Thesis.pdf VOCEDplus database.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan Library Reference.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Oxon, UK: Routledge.
- Committee for Economic Development of Australia. (2016). *VET: Securing skills for growth* (0858013061). Retrieved from <http://www.ceda.com.au/research-and-policy/research/2016/08/vet-skills-for-growth>
- Conrad, B. (2012). Intentions and beliefs: why they matter and a conceptual framework for understanding them in culturally responsive teachers. *Curriculum and Teaching Dialogue*, 14, 87-99.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). London: SAGE Publications.
- Drouin, M., Stewart, J., & Van Gorder, K. (2015). Using methodological triangulation to examine the effectiveness of a mentoring program for online instructors. *Distance Education*, 36(3), 400-418. doi:10.1080/01587919.2015.1081735
- Eley, M. G. (2006). Teachers' conceptions of teaching, and the making of specific decisions in planning to teach. *Higher Education*, 51(2), 191-214. doi:10.1007/s10734-004-6382-9
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: SAGE Publications.
- Fowler, C. (2017). *The boundaries and connections between the VET and higher education sectors: 'confused, contested and collaborative'*. Retrieved from Adelaide: https://www.ncver.edu.au/__data/assets/pdf_file/0027/158706/Boundaries_-and-_connections_between_VET_-and-_higher_ed_sectors.pdf
- Gaytan, J. (2015). Comparing faculty and student perceptions regarding factors that affect student retention in online education. *American Journal of Distance Education*, 29(1), 56-66. doi:10.1080/08923647.2015.994365
- Goldman, Z. (2011). Balancing quality and workload in asynchronous online discussions: A win-win approach for students and instructors. *MERLOT Journal of Online Learning and Teaching*, 7(2), 313-323.
- Jaggars, S. S., & Xu, D. (2013). Predicting online student outcomes from a measure of course quality. In *CCRC Working Paper No.57: Teachers College Columbia University*.
- Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *MERLOT Journal of Online Learning and Teaching*, 8(3), 198-208.
- Lindblom-Ylänne, S., Trigwell, K., Nevgi, A., & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. *Studies in Higher Education*, 31(3), 285-298. doi:10.1080/03075070600680539
- Nguyen, T. (2015). The effectiveness of online learning: Beyond no significant difference and future horizons. *MERLOT Journal of Online Learning and Teaching*, 11(2), 309-319.
- Niess, M. L. (2011). Investigating TPACK: Knowledge growth in teaching with technology. *Journal of Educational Computing Research*, 44(3), 299-317. doi:doi:10.2190/EC.44.3.c
- Norton, L., Richardson, J., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education*, 50(4), 537-571. doi:10.1007/s10734-004-6363-z

- Owens, T. (2015). Practising what they preach? An investigation into the pedagogical beliefs and online teaching practices of National Teaching Fellows. *International Journal for Academic Development*, 20(1), 76-92. doi:10.1080/1360144X.2014.983112
- Pallant, J. (2016). *SPSS survival manual* (16th ed.). Sydney: Allen & Unwin.
- Palloff, R. M., & Pratt, K. (2013). *Lessons from the virtual classroom: The realities of online teaching* (2nd ed.). San Francisco: Jossey-Bass.
- Prieto-Rodriguez, E., Gore, J., & Holmes, K. (2016). Exploring quality teaching in the online environment using an evidence-based approach. *Australian Journal of Teacher Education*, 41(8), 22-39.
- Productivity Commission. (2011). *Productivity Commission research report: Vocational education and training workforce*. Retrieved from <http://www.pc.gov.au/inquiries/completed/education-workforce-vocational/report/vocational-workforce.pdf>
- Productivity Commission. (2019). *Report on government services 2019: Chapter 5 Vocational education and training*. Retrieved from Canberra: <https://www.pc.gov.au/research/ongoing/report-on-government-services/2019/child-care-education-and-training/vocational-education-and-training/rogs-2019-partb-chapter5.pdf>
- Richardson, J. C., Besser, E., Koehler, A., Lim, J., & Strait, M. (2016). Instructors' perceptions of instructor presence in online learning environments. *International Review of Research in Open and Distributed Learning*, 17(4), 82-104.
- Redmond, P. (2011). *From face-to-face teaching to online teaching: Pedagogical transitions*. Paper presented at the ascilite 2011, Hobart. <http://www.ascilite.org/conferences/hobart11/downloads/papers/Redmond-full.pdf>
- Reeson, A., Mason, C., Sanderson, T., Bratanova, A., & Hajkowitz, S. (2016). *The VET era: equipping Australia's workforce for the future digital economy*. Retrieved from <http://tafeqld.edu.au/resources/pdf/about-us/research-papers/vet-era.pdf>
- Rodrigues, H., Almeida, F., Figueiredo, V., & Lopes, S. L. (2019). Tracking e-learning through published papers: A systematic review. *Computers & Education*, 136, 87-98. doi:https://doi.org/10.1016/j.compedu.2019.03.007
- Samuelowicz, K., & Bain, J. D. (1992). Conceptions of teaching held by academic teachers. *Higher Education*, 24(1), 93-111.
- Seaton, J. X., & Schwier, R. (2014). An exploratory case study of online instructors: Factors associated with instructor engagement. *International Journal of E-Learning & Distance Education*, 29(1), 1-16.
- Sorensen, C. (2014). Classrooms without walls: A comparison of instructor performance in online courses differing in class size. *MERLOT Journal of Online Learning and Teaching*, 10(4), 569-576.
- Sun, A., & Chen, X. (2016). Online education and its effective practice: A research review. *Journal of Information Technology Education: Research*, 15(2016), 157-190.
- Taft, S. H., Perkowski, T., & Martin, L. S. (2011). A framework for evaluating class size in online education. *Quarterly Review of Distance Education*, 12(3), 181-197.
- Terblanché, E. (2015). Deciding to teach online: Communication, opportunities and challenges for educators in distance education. *Communicatio*, 41(4), 543-563. doi:10.1080/02500167.2015.1115416
- Warmbrod, J. R. (2014). Reporting and interpreting scores derived from Likert-type scales. *Journal of Agricultural Education*, 55(5), 30-47.
- Wheelah, L., & Moodie, G. (2011). *The quality of teaching in VET: final report and recommendations*: Australian College of Educators.
- Yen, S.-C., Lo, Y., Lee, A., & Enriquez, J. (2018). Learning online, offline, and in-between: comparing student academic outcomes and course satisfaction in face-to-face, online, and blended teaching modalities. *Education and Information Technologies*, 23(5), 2141-2153. doi:10.1007/s10639-018-9707-5

Relationship between Parents' Perceptions of Programming Education and Their Emotional and Behavioral Outcomes

Siu-Cheung KONG ^{a*} & Yi-Qing WANG ^b

^aDepartment of Mathematics and Information Technology; ^bCentre for Learning, Teaching and Technology, The Education University of Hong Kong, Hong Kong

* sckong@eduhk.hk

Abstract: Parents are assumed as one of the key figures to foster programming learning among children after school. However, little effort has been directed to the examination of parents' perceptions and their consequent emotional and behavioral outcomes regarding programming education to date. This study attempted to explore how parents' intrinsic motivation and positive affect would positively influence the relationship between their perceptions of programming education and involvement for providing a better learning opportunity to children. We collected 695 questionnaires from parents who brought their children to a one-day coding fair where they could experience the fun of programming. Results from Structural Equation Modeling (SEM) analysis indicated that parents' programming perceptions is positively related to their intrinsic motivation and positive affect. Mediation results suggested that parents' perceptions such as understanding, support, and expectation can trigger their intrinsic motivation for encouraging children in programming learning and enhance their positive affect toward learning programming themselves, which consequently lead to increased attempts of parental involvement for the better guidance in technology use for children. Implications of the study were discussed.

Keywords: Intrinsic motivation, parental involvement, parents' perception, positive affect, programming education

1. Introduction

Programming education for younger learners has become a trend around the globe (e.g. Grover & Pea, 2013). Parents' role is important to help children learn programming after school. If parents understand the significance of programming education, they tend to provide greater emotional support and involve in children's coursework (Kong, Li, & Kwok, 2018a). In other words, parents' programming perceptions can lead to their emotional and behavioral changes in programming education. Kong, Li, and Kwok (2018b) developed a scale to measure parents' perceptions of programming education in P-12 schools. However, there is no existing literature further exploring the relationship between parents' perceptions and their consequent emotional and behavioral outcomes regarding programming education. To bridge this gap, this study aimed to investigate parents' dynamic changes in cognitive-behavioral mechanism via the potential factor of emotions towards programming. Specifically, this study comprehensively examined how parents' positive affect (*affect*) and intrinsic motivation (*motivation*) of programming education influence the relationship between their perceptions of programming education (*perceptions*) and involvement in guiding children to use technology (*involvement*). Abbreviations in parentheses for key study variables will be used throughout the paper.

2. Literature Review

2.1 Parents' Motivation of Programming Education as the Mediator

Kong et al. (2018b) proposed that perceptions of programming education include individuals' understanding, support, and expectation. Previous study found that intrinsic motivation is related to people's understanding and expectation. For example, if parents understand that the coursework is useful, they will highly encourage children to engage in it (Katz, Kaplan, & Buzukashvily, 2011). In technology education context, Rozell and Gardner III (2000) also found that perceived task importance will affect one's intrinsic motivation in completing the task. It means that people are more willing to motivate themselves or others to engage in the computer-related activities when they understand the usefulness of computer. Thus, this study expects that parents with positive programming perceptions will have greater intrinsic motivation to encourage children in programming learning.

Hypothesis 1: Parents' perceptions of programming education is positively related to their intrinsic motivation of programming education.

Past studies suggested that one's motivation will affect his or her involvement. For instance, students who have strong intrinsic motivation to study a subject are more likely to dive deeper into the topic (Bergin & Reilly, 2005). Witherspoon, Schunn, Higashi, and Baehr (2016) supported that there is a significant relationship between students' motivation in learning programming and their subsequent involvement. Regarding parents, motivation beliefs like parental role construction and self-efficacy are associated with their involvement in children's education (Walker, Wilkins, Dallaire, Sandler, & Hoover-Dempsey, 2005). Katz et al. (2011) also found that parents' supportive behaviors of helping with children's homework is related to their autonomous motivation. Thus, we assume that parents with greater intrinsic motivation will also have higher involvement in guiding children to use technology.

Hypothesis 2: Parents' intrinsic motivation of programming education is positively related to their involvement in guiding children to use technology.

Hypothesis 3: Parents' intrinsic motivation of programming education will positively mediate the relationship between their perceptions of programming education and involvement in guiding children to use technology.

2.2 Parents' Positive Affect toward Programming Education as the Mediator

Positive affect toward programming education refers to one's positive feeling towards the use of programming. For example, students' liking for school and their perceptions of school subjects are closely related to each other (Ireson & Hallam, 2005). Wang and Holcombe (2010) also found that students' perceptions of the school environment and school engagement will influence their affective reactions. In technology education context, Teo (2009) indicated that the perceived usefulness of technology will bring positive impact to a person's liking of using technology. Although there is a lack of literature directly investigating parents' perspective, this study expects that parents with positive programming perceptions will show more positive affect toward programming education.

Hypothesis 4: Parents' perceptions of programming education is positively related to their positive affect toward programming education.

Fredrickson (2001, p.218) argued that "experience of positive emotions broaden people's momentary thought-action repertoires". For example, Meyer and Turner (2006) argued that students with positive emotions at school will have higher levels of engagement. Besides, Katz et al. (2011) suggested that parents engage in children's homework because they think the work is interesting and enjoyable. In technology education context, it is also believed that people's perceived liking of computer will bring positive impact on the intention of using technology. For example, teachers with positive attitudes towards computers are more likely to integrate computers in their teaching (Zhao & Frank, 2003). Thus, this study expects that parents with higher positive affect toward programming education will have higher involvement in guiding children to use technology.

Hypothesis 5: Parents' positive affect toward programming education is positively related to their involvement in guiding children to use technology.

Hypothesis 6: Parents' positive affect toward programming education will positively mediate the relationship between their perceptions of programming education and involvement in guiding children to use technology.

3. Method

3.1 Procedures and Participants

A one-day coding fair was held in 2018 for parents to bring their children to experience the fun of programming. Some workshops and seminars were organized to introduce computational thinking education. The parents were required to complete the questionnaires before they participated in the fair. There were 695 parents who returned the questionnaires, thus the response rate was 69.5%. In our sample, 43.6% of the participants were male and 51.7% were female. 52.4% were between 41 to 50 years old. 75.5% obtained bachelor's degree or above. 58.4% of the parents did not know programming.

3.2 Measures

Perceptions: This is a 9-item scale developed by Kong et al. (2018b). The sample item is "It is good for my children to learn programming." In our study, the Cronbach's alpha is .96.

Affect: We adopted a 3-item liking subscale developed by Ng (2011). The sample item is "I like to discuss programming-related topics." In our study, the Cronbach's alpha is .90.

Motivation: We adopted 3 items from Gottfried, Fleming, and Gottfried (1994) for this study. The sample item is "I encourage my children to learn more programming." In our study, the Cronbach's alpha is .97.

Involvement: This study adopted a 5-item scale developed by Walker et al. (2005). The sample item is "I will provide feedback when my children are using technological devices." In our study, the Cronbach's alpha is .93. All the scales in the current study anchored from 1 strongly disagree to 5 strongly agree.

Control variables: We included gender, age, education level, and programming experience for investigation, as past studies found that these demographic variables might be closely related to parents' perceptions towards computer-related activities and involvement (e.g., Vekiri & Chronaki, 2008; Overstreet, Devine, Bevans, & Efreom, 2005; Walker et al., 2005).

4. Results

4.1 Descriptive Statistics

The descriptive statistics of mean, standard deviations, and reliability coefficients of all the variables were shown in Table 1. The Cronbach's Alpha coefficient of perceptions scale and its three subcomponents were ranged from .88 to .96, indicating good internal consistency. Besides, affect scale ($\alpha = .90$), motivation scale ($\alpha = .97$), and involvement scale ($\alpha = .93$) all showed satisfactory reliability. In addition, factor loadings for all items in each scale were ranged from .83 to .96, which confirmed convergent validity of the scales. Positive inter-correlations were found among the four study variables—perceptions, affect, motivation, and involvement.

Table 1
Mean, Standard Deviations, and Reliability Coefficients of the Variables

	Mean	SD	1	2	3	4	5	6	7
1. Perceptions	4.00	.80	(.96)						
2. Understanding	4.05	.88	.95**	(.95)					
3. Support	3.88	.83	.93**	.82**	(.88)				
4. Expectation	4.09	.84	.94**	.85**	.82**	(.93)			
5. Affect	3.95	.82	.87**	.79**	.83**	.84**	(.90)		
6. Motivation	4.10	.85	.84**	.82**	.73**	.83**	.77**	(.97)	
7. Involvement	3.94	.86	.71**	.65**	.67**	.68**	.72**	.76**	(.93)

Note. N=673 *p <.05; **p <.01; ***p <.000.

4.2 Structural Equation Model

Hypotheses were tested using structural equation modeling (SEM) in AMOS version 24 (Ackerman, 2003). Several statistics were used to assess model fit, including the chi-square (χ^2) statistic, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and Tucker–Lewis index (TLI). According to Hu and Bentler (1999), CFI and TLI which are greater than .95 suggested an excellent fit. RMSEA in the range of .05 to .08 also indicates acceptable fit (Browne & Cudeck, 1993). In this study, the hypothesized measurement model showed a good fit with the data collected ($\chi^2(95) = 455.695$, $p < .000$, CFI = .97, TLI = .95, and RMSEA = .07). The SEM results showed significant paths from perceptions to motivation ($\beta = .89$, $p < .001$) and from motivation to involvement ($\beta = .56$, $p < .001$). Thus, Hypothesis 1 and 2 were supported. In addition, there were significant paths from perceptions to affect ($\beta = .94$, $p < .000$) and from affect to involvement ($\beta = .31$, $p < .000$). Therefore, Hypothesis 4 and 5 were also supported. Moreover, mediation analyses were conducted using bias-corrected bootstrap method (2000 resamples) in PROCESS (Hayes, 2012) to test the significance of the indirect effect of motivation and affect. Results indicated that the mediating effects are statically significant (95% CI_{motivation} = [.40, .61], 95% CI_{affect} = [.16, .36]). Thus, hypothesis 3 and 6 were fully supported. Figure 1 showed the theoretical model of this study.

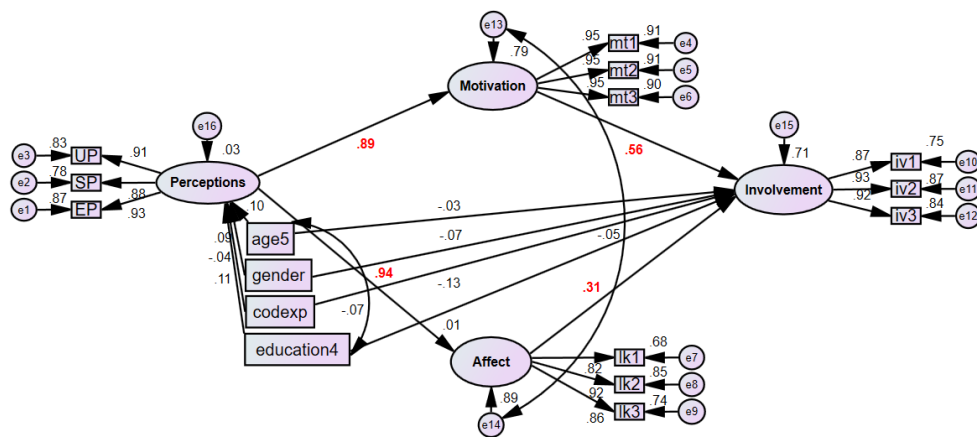


Figure 1. Theoretical Model

5. Discussion

Parents' role in children's education is critical; however, their perceptions and consequent emotional and behavioral outcomes regarding programming education is less known in the literature. Leutner's (2014) study attempted to explore how positive emotions and interest influence the cognitive processing and multimedia learning results. However, he did not provide empirical evidence to support his proposition. Indeed, past studies predominately focused on perceptions (Zainal et al., 2011; Rozell & Gardner III, 2000) and few of them rigorously tested the relationships among cognition, emotion and behavior in a comprehensive way. This study contributes to the current literature by providing empirical results showing the dynamics of cognitive-behavioral mechanism via positive emotions. Specifically, mediation analyses were conducted to examine how parents' affect and motivation influence the relationship between perceptions and involvement. The results found that affect as well as motivation can strengthen the relationship between perceptions and involvement. It yields a significant insight about the importance of emotion and motivation in affecting people's behavioral involvement as a consequence.

This study also found that parents' demographics are associated with perceptions and involvement. Parents' gender is related to perceptions ($\beta = .09$, $p < .05$). Mothers tend to have better perceptions than fathers do. It concurred with Kong et al. (2018a) research which indicated that mothers have higher expectations on what children can learn through programming. Parents' gender is also associated with involvement ($\beta = -.07$, $p < .01$). More fathers tend to involve in children's technology

education, because more of them have programming experience than female counterparts in this study. Knowing the different roles that both parents might play, we advocate the proactive involvement of both parents in children's programming learning, that is, we encourage mothers to provide more emotional support and fathers to provide more instructional support in order to achieve optimal learning experience for their children. In addition, parents' age is related to perceptions ($\beta = .10, p < .05$). Older parents have more positive perceptions than younger parents do. Researchers suggested that older adults can function effectively when the given tasks allow them to draw upon their expertise (Rybash, Roodin, & Hoyer, 1995). Parents' education level also influences perceptions ($\beta = .11, p < .01$). This result is in line with the literatures suggesting that parents' education level is positively related to their educational expectation on children (e.g. Overstreet et al., 2005). However, there is no relationship between parents' education level and involvement. We argue that parents' social ranking like education level cannot predict parental involvement while psychological motivators and life context might be better predictors. Finally, programming experience influences involvement ($\beta = -.13, p < .000$), because parents self-perceived knowledge can contribute to their involvement in children's education.

5.1 Practical Implication

This study found that without perceptions, motivation and affect can hardly be nurtured. Thus, we should increase parents' understanding of programming education to stimulate their motivation of encouraging children to learn programming. Conducting programming workshops is an effective way to enhance perceptions. Kong et al. (2018a) reported that parent's understanding, support, and expectation of programming education increase significantly after workshops. Therefore, educational practitioners are encouraged to design programming workshops that can strengthen parents' understanding and expectation in programming education. It is also crucial to raise parents' motivation in programming through positive interventions. Despite that a single workshop may only generate a situational interest of programming (Teague, 2002), Mitchell (1993) argued that a learning environment with high situational interest will bring positive changes to one's long-term interest over time. In fact, we suggested that a series of workshops should be provided for parents since a single positive experience is insufficient to maintain their interest. More importantly, workshops should be designed in a way that provokes the intrinsic motivation in parents such that they can form habits of supporting children in their learning. Researchers found children can program in a more systematic way with fewer errors if parents are motivated to involve when children are writing programs (Lin & Liu, 2012). Last but not least, findings of this study also provided a general implication that parents' attitudes, emotions, and behaviors are of great importance because their influences are not limited to programming learning, but are profound in all other learning areas, such as science and math, etc.

5.2 Limitations and Future Research Direction

This study has three limitations that we need to address in the future. First, the data might be sample specific. Parents who participated in the coding fair might have greater interest in programming than parents who did not show up. In the future, more general parent samples should be tested for generalizability purposes. Second, this study adopted four validated scales to conduct a mediation analysis with multiple mediators. However, there might be other existing models that can fit our data adequately well. In the future, we should explore more alternatives by comparing our hypothesized model with other potential models. Third, self-reporting data of this study might cause common-method bias, which might inflate relationships between study variables. Ratings from various sources can be used to avoid this problem. For instance, observations from multiple raters of how parents interact with children in the future coding fair can be included for results validation. Moreover, qualitative data such as parent interviews and group discussions are encouraged to comprehend the interpretation of quantitative results.

References

Ackerman, S. (2003). Amos 5: 18–24. *Interpretation*, 57(2), 190-193.

- Bergin, S. & Reilly, R. (2005). The influence of motivation and comfort-level on learning to program. In P. Romero, J. Good, E. A. Chaparro, & S. Bryant (Eds.), *Proceedings of the 17th Workshop on Psychology of Programming* (pp. 293-304). Seattle, WA: ACM.
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1449), 1367-1378.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1994). Role of parental motivational practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology*, 86(1), 104-113.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from <http://www.afhayes.com/>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- Ireson, J., & Hallam, S. (2005). Pupils' liking for school: Ability grouping, self-concept and perceptions of teaching. *British Journal of Educational Psychology*, 75, 297-311.
- Katz, I., Kaplan, A., & Buzukashvily, T. (2011). The role of parents' motivation in students' autonomous motivation for doing homework. *Learning and Individual Differences*, 21, 376-386.
- Kong, S. C., Li, R. K. Y., & Kwok, R. C. W. (2018a). Perception of parents on programming education in P-12 schools. In *Proceedings of the 22nd Global Chinese Conference on Computers in Education* (pp. 1485-1493). Guangzhou, China: Global Chinese Society for Computers in Education.
- Kong, S. C., Li, R. K. Y., & Kwok, R. C. W. (2018b). Measuring parents' perceptions of programming education in P-12 schools: Scale development and validation. *Journal of Educational Computing Research*. Advance online publication. doi: 10.1177/0735633118783182
- Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, 29, 174-175.
- Lin, J. M. C., & Liu, S. F. (2012). An investigation into parent-child collaboration in learning computer programming. *Journal of Educational Technology & Society*, 15(1), 162-173.
- Meyer, D. K., & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18, 377-390.
- Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, 85(3), 424-436.
- Ng, W. L. (2011). Reliability and validity of an information and communications technology attitude scale for teachers. *The Asia-Pacific Education Researcher*, 20(1), 162-170.
- Overstreet, S., Devine, J., Bevans, K., & Efreom, Y. (2005). Predicting parental involvement in children's schooling within an economically disadvantaged African American sample. *Psychology in the Schools*, 42(1), 101-111.
- Rozell, E. J., & Gardner III, W. L. (2000). Cognitive, motivation, and affective processes associated with computer-related performance: A path analysis. *Computers in Human Behavior*, 16, 199-222.
- Rybash, J. M., Roodin, P. B., & Hoyer, W. J. (1995). *Adult development and aging*. New York, NY: McGraw-Hill.
- Teague, J. (2002). Women in computing: What brings them to it, what keeps them in it? *SIGCSE Bulletin*, 34(2), 147-158.
- Teo, T. (2009). The impact of subjective norm and facilitating conditions on pre-service teachers' attitude toward computer use: a structural equation modeling of an extended technology acceptance model. *Journal of Educational Computing Research*, 40(1), 89-109.
- Vekiri, I., & Chronaki, A. (2008). Gender issues in technology use: Perceived social support, computer self-efficacy and value beliefs, and computer use beyond school. *Computers & Education*, 51(3), 1392-1404.
- Walker, J. M. T., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental involvement: Model revision through scale development. *The Elementary School Journal*, 106(2), 85-104.
- Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, 47, 633-662.
- Witherspoon, E. B., Schunn, C. D., Higashi, R. M., & Baehr, E. C. (2016). Gender, interest, and prior experience shape opportunities to learn programming in robotics competitions. *International Journal of STEM Education*, 3(18). doi: 10.1186/s40594-016-0052-1
- Zainal, N. F. A., Shahrani, S., Yatim, N. F. M., Rahman, R. A., Rahmat, M., & Latih, R. (2012). Students' perception and motivation towards programming. *Procedia - Social and Behavioral Sciences*, 59, 277-286.
- Zhao, Y., & Frank. K. A. (2003). Factors affecting technology uses in schools: An ecological perspective. *American Educational Research Journal*, 40(4), 807-840.

Supporting Teachers in Group Work Formation and Analytics for In-class Group Activities

Liang Changhao^{a*}, Ivica Botički^b & Hiroaki Ogata^a

^a *Kyoto University, Japan*

^b *Faculty of Electrical Engineering and Computing, Zagreb, Croatia*

*liang.changhao.84c@st.kyoto-u.ac.jp

Abstract: This paper introduces a system for collaborative learning which is designed to assist teachers in forming and grading groups for in-class group activities. The system is implemented as an extension of a learning analytics dashboard system and uses log data from a learning management system for operation. It consists of a group formation parameter console and the results console where formed groups are visualized and can be graded. The system supports teachers by using algorithms based on reliable learning evidence thereby simplifying the group formation process. All the group formation and grading data is logged thereby cyclically providing an infrastructure for subsequent collaborative learning activities.

Keywords: Collaborative Learning, Group Formation, Learning Analytics, CSCL

1. Introduction

In collaborative learning students work in groups on a task towards a goal (Dillenbourg, 1999). Collaboration is not only important for learning about content, but also for acquiring negotiation skills, learning to provide good arguments to support own ideas and nurturing social sense of belonging to the team (Stahl, Koschmann, & Suthers, 2006). With the advancement of ICT, new tools that include constructivist and collaborative learning are starting to emerge (Paredes & Martin, 2004). In this paper we examine the notion of collaborative learning based on the data in LMS platforms and existing learning analytics services.

According to Vygotsky's Zone of Proximal Development (ZPD), collaboration leverages peer and group resources so that individuals develop their potential and extend their knowledge (Vygotsky, 1978). To achieve this, imbalance between group members and their resources is sometimes designed for in group activities. On other occasions, friendship amongst students can be considered, making the groups homogeneous. For successful in-class collaborative learning, teachers need to envision the lesson, enable collaboration, encourage students, ensure learning, and evaluate achievements (Urhahne, Schanze, Bell, Mansfield, & Holmes, 2010). Amongst these tasks, group formation is the fundamental component since it determines quality of group work (Wessner & Pfister, 2001). When undertaken by the teachers, the task of forming heterogeneous or homogeneous groups is by any means no trivial and teachers might often feel confused since they do not always have an immediate access to the background information and data about the students.

The system presented in this paper builds on an existing learning analytics platform and provides features to support teachers in group work formation and analytics. It also contributes to the learning analytics platform by supplying group work performance data to other learning analytics-based services. By using its visualization support teachers can compare students' performance in group work and make more informed group formation decisions in their subsequent learning designs.

2. Theoretical Background

Since the 1960s, there has been considerable effort invested in research of cooperative learning and small group dynamics (Gillies & Ashman, 2003). Cooperative learning is done by individuals, who then contribute with their individual results to the group and present the collection of individual results as

their group product (Dillenbourg, 1999). In collaborative learning the interaction among group members is stressed and learning occurs socially as collaborative construction of knowledge (Roschelle & Teasley, 1995). Computer-supported collaborative learning (CSCL) is an emerging branch of learning sciences concerned with studying how people learn together with the help of computers (Stahl et al., 2006). In CSCL, teachers are often overwhelmed with the amount of work when setting up and implementing collaborative learning activities. To make in-class group work successful, teachers should direct their energy on orchestrating classroom activities, rather than spending time on technical things like group work set-up and technology adjustments (Austin, Smyth, Rickard, Quirk-Bolt, & Metcalfe, 2010).

Today’s online learning platforms provide massive log data that can be used for learning analytics possibly leading to improved learning designs and outcomes (Siemens, 2013). There exists an abundance of research evidence on how these logs can be used to improve e-book contents and the quality of learning and education (Arnold et al., 2012; Fujimura K; Ogata, H; Okubo, F; Shimada, A; Yamada, M; Yin, C;, 2014; Lu, Huang, Huang, & Yang, 2017). Our study is contextualized around a learning analytics platform connected to a Moodle-based learning management system (LMS) designed in Japan (Flanagan & Ogata, 2018), which is being used at several international locations.

In this paper we examine a variety group formation algorithms (Hoic-Bozic, Mornar, & Boticki, 2008). These algorithms are used in conjunction with user modelling (Boticki, Akçapınar, & Ogata, 2019), where the most common way is to rank students according to the data in the user model variables (one example for such a variable are Moodle quiz scores) and organize them into groups.

3. A Model for Group Formation and Scoring in a Learning Analytics Platform

The group formation module introduced in this paper is an extension of a learning analytics system and its dashboard application which is a central component for visualizing learning evidence (Ogata, Majumdar, Akçapınar, Hasnine, & Flanagan, 2018). As illustrated in Figure 1, learning log data is collected from educational applications such as the BookRoll system (Flanagan & Ogata, 2018), organized into student model variables which characterize students’ features (Brusilovsky et al., 2016) and stored into the analysis database. The group formation module makes use of the user model data and grouping algorithms to generate groups. Once groups are formed and their performance is graded, these data are used to update the student model for further learning analytics use.

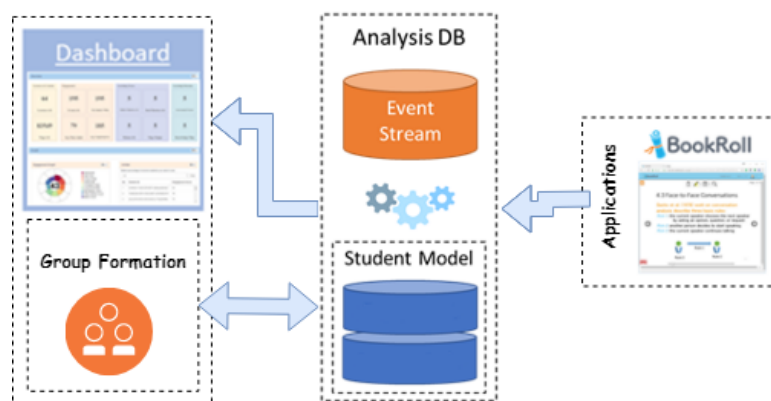


Figure 1. The main learning analytics components including the group formation module.

Figure 2 shows the workflow of group work activities. In the beginning, teachers need to decide which course and students are to be used as part of collaborative work. Following that, the teachers decide on the group formation parameters that best suit the concrete learning activity. During and after group work, teachers grade the performance of group work and give feedback to the students. Using performances obtained from the previous application cyclically, teachers get more informed for the next group formation processes. The group score user model gets more reliable as group activities and grading get frequent.

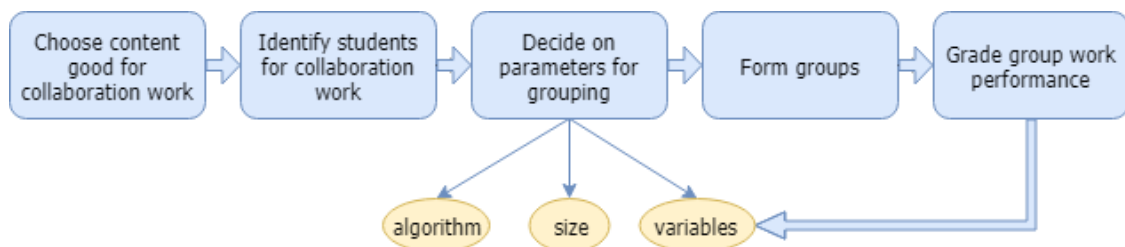


Figure 2. The process of group formation.

The process described in Figure 2 are supported by the group formation module which gets input containing course and student data from the Moodle LMS platform. Teachers can modify the data by excluding inactive students from group formation and use the parameters feature to decide on an adequate algorithm and user model variable (Table 1 and Table 2). Alternatively, the teacher can select automatic grouping and leave the parameter selection to the system (default settings being heterogeneous algorithm, the group score user model variable, and the group size of four). Teachers can view the formed groups and evaluate their group performance in three indicators (Table 3), which are then recorded as group score user model variables.

Table 1

Algorithms used in the group formation process

Algorithm	Algorithm Operation Description
Homogenous	The algorithm groups students with similar values of a variable.
Heterogenous	The algorithm groups students with differing values of variables.
Friendship	The algorithm groups students who are friends as identified by the teacher.
Random	The algorithm groups students randomly.

Table 2

Variables used in the group formation process

Variable	Variable Explanation
Engagement	The variable records the time student spent on using the learning platform.
Reading style	The variable models two reading styles: receptive reading and responsive reading (Pugh, 1979). The former refers to the style of reading page after page sequentially and the latter deals with students engaging in more interaction with the digital material (taking notes and posting comments).
Concept	The variable describes the mastery of each key concept found in academic materials of a course.
Score	The variable records previous assessment scores.
Group score	The variable models students' previous performance of in group, gathered as part of group grading.

Table 3

Metrics of group performance evaluation

Indicator	Metrics for evaluation
Collaboration quality	Interaction and communication occurring during group work, participation of members and rational division of labor
Speed / efficiency	Whether each subtask is finished on time and reasonable time management
Final output	The quality of final outputs and artefacts of group work

4. Teachers' Use of the Group Formation Module

This section will demonstrate two examples of group formation: the heterogeneous group formation based on course grades and group formation using the Friendship algorithm. In a group lesson, teachers may want students with different course grades to work together and learn from each other. In that case, teachers can select heterogeneous algorithm and choose score as the user model group formation variable via the parameters feature shown in Figure 3.

In some learning activity designs teachers may want to group students with good mutual relationships together. After setting mutual relationships between the students and getting an overview of the whole class (Figure 4), the teachers use the Friendship algorithm to form the groups of friends.

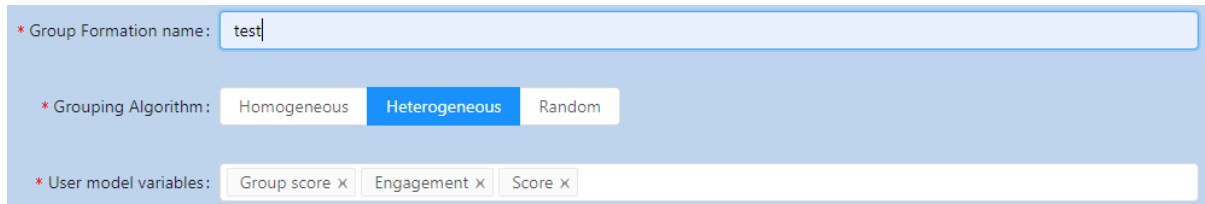


Figure 3. Parameter feature of the group formation module.

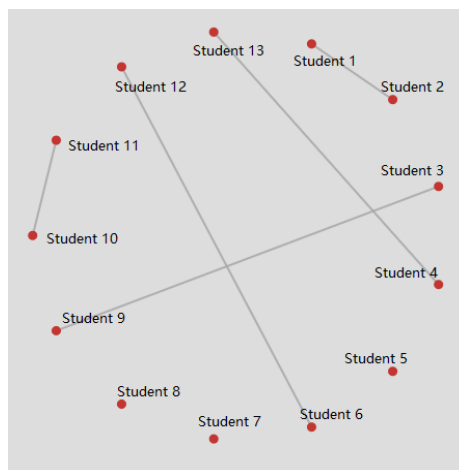


Figure 4. Student relationship graph used by the Friendship group formation algorithm.

Figures 5 and 6 depict the results of the heterogeneous and Friendship grouping algorithms operation, respectively. Traffic-light colors are used to give indication of previous group work performance to the teachers who guide the group formation process (green, yellow and red color denote good, average and low previous group work performance, respectively). Group performance indication is shown both at the group level and at an individual student level. Teachers score the group performance for each indicator and the scores in all three indicators are stored as part of the group user model giving an overall estimation of students' collaboration performance.

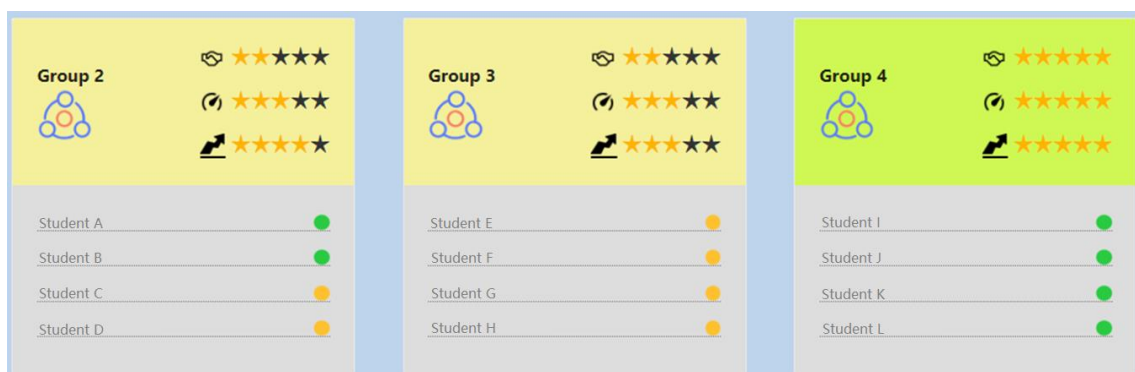


Figure 5. Groups formed (3 groups of 4 students) using the heterogeneous grouping algorithm and the color indication of individual and group collaboration performance.

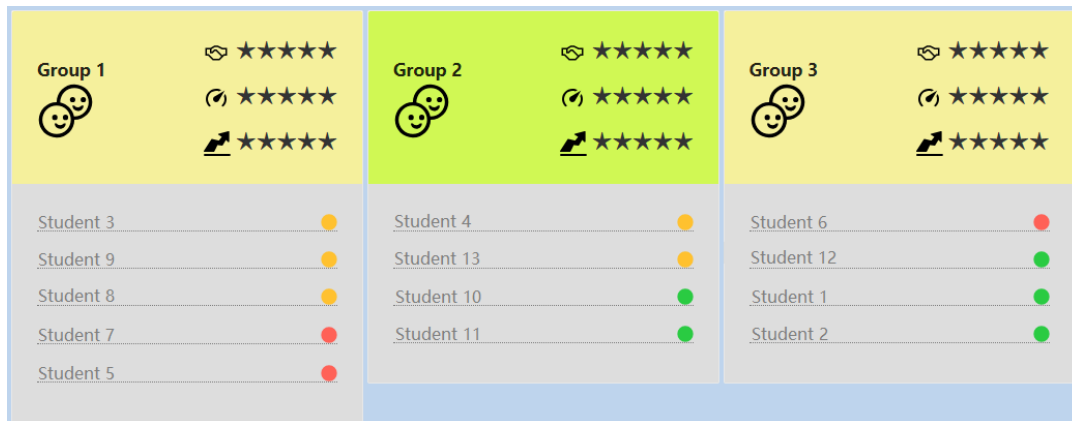


Figure 6. Groups formed (one group of 5 and two groups of 4 students) using the Friendship algorithm and color indication of the individual and group collaboration performance.

5. Conclusions and Future Work

Teachers might get overwhelmed when using CSCL in their teaching activities and need support to be able to execute and manage such activities in a timely and informed manner. The group formation system provides group formation assistance with visuals to the teachers and enriches group work experience with the help of the learning analytics platform. In the system presented in this paper, teachers have an option of choosing an automatic group formation feature which favors imbalance between members of a single group. This is grounded in the research in the area of the Zone of Proximal Development (ZPD) and potentially promotes construction of knowledge and an elevated level of mutual understanding of a topic (Nyikos, 1997).

In terms of group performance, the system allows real time rating of groups by the teachers in the three main indicators. The performance rating component is available directly on the group results panel, for the teachers to take an immediate record of group performance and adjust the grade in real time. Such a feature facilitates the adequate timing of teachers' interventions, which proved to be of importance for in-class collaborative learning (Coll, Rochera, & De Gispert, 2014). Meanwhile, not only summative but also formative indicators such as collaboration quality are stressed in the system (Strijbos, 2011).

The group work module presented in this paper was developed as an extension of an existing learning analytics platform and allows teachers to conduct group formation based on the existing user data (user models) in their in-class activities. Future work will examine the ways of using the system in in-class group activities with a focus on evaluating group performance in realistic collaborative environments with social loafing and free-riding, as the most challenging problems in today's collaborative learning. In parallel, empirical research is necessary to verify whether and to what extent group formation algorithms help teachers with groupwork and in fulfilling the collaboration goals of a lesson. This will lead, as part of the planned future work, to automatized suggestions of group formation algorithms to the teacher, depending on the identified group formation purpose.

Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

- Arnold, K. E., Hall, Y., Street, S. G., Lafayette, W., Pistilli, M. D., Hall, Y., ... Lafayette, W. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. *Lak '12*, (May), 2–5. <https://doi.org/10.1145/2330601.2330666>

- Austin, R., Smyth, J., Rickard, A., Quirk-Bolt, N., & Metcalfe, N. (2010). Collaborative digital learning in schools: Teacher perceptions of purpose and effectiveness. *Technology, Pedagogy and Education*.
<https://doi.org/10.1080/1475939X.2010.513765>
- Boticki, I., Akçapınar, G., & Ogata, H. (2019). E-book user modelling through learning analytics: the case of learner engagement and reading styles. *Interactive Learning Environments*, 0(0), 1–12.
<https://doi.org/10.1080/10494820.2019.1610459>
- Brusilovsky, P., Somyurek, S., Guerra, J., Hosseini, R., Zadorozhny, V., & Durlach, P. J. (2016). Open Social Student Modeling for Personalized Learning. *IEEE Transactions on Emerging Topics in Computing*, 4(3), 450–461. <https://doi.org/10.1109/TETC.2015.2501243>
- Coll, C., Rochera, M. J., & De Gispert, I. (2014). Supporting online collaborative learning in small groups: Teacher feedback on learning content, academic task and social participation. *Computers and Education*.
<https://doi.org/10.1016/j.compedu.2014.01.015>
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative Learning Cognitive and Computational Approaches* (pp. 1–19). <https://doi.org/10.1.1.167.4896>
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management and E-Learning*.
- Fujimura K; Ogata, H; Okubo, F; Shimada, A; Yamada, M; Yin, C.; N. K. (2014). Smart Phone based Data Collecting System for Analyzing Learning Behaviors. *Proceedings of the 22nd International Conference on Computers in Education, ICCE 2014*.
- Gillies, R. M., & Ashman, A. F. (2003). An historical review of the use of groups to promote socialization and learning. In *Cooperative Learning: The Social and Intellectual Outcomes of Learning in Groups*.
<https://doi.org/10.4324/9780203465264>
- Hoic-Bozic, N., Mornar, V., & Botick, I. (2008). Introducing adaptivity and collaborative support into a web-based lms. *Computing and Informatics*.
- Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2017). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments*, 25(2), 220–234. <https://doi.org/10.1080/10494820.2016.1278391>
- Nyikos, M. (1997). Constructivist Theory Applied to Collaborative Learning in Teacher Education : In Search of ZPD. *Modern Language Journal*.
- Ogata, H., Majumdar, R., Akçapınar, G., Hasnine, M. N., & Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning. *ICCE 2018 - 26th International Conference on Computers in Education, Workshop Proceedings*.
- Paredes, P., & Martin, E. (2004). USING LEARNING STYLES FOR DYNAMIC GROUP FORMATION IN ADAPTIVE COLLABORATIVE HYPERMEDIA SYSTEMS. *ICWE Workshops*.
- Pugh, A. K. (1979). Styles and Strategies in Silent Reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language* (pp. 431–443). https://doi.org/10.1007/978-1-4684-0994-9_27
- Roschelle, J., & Teasley, S. (1995). Computer Supported Collaborative Learning (CSCL). *Computers & Education*. [https://doi.org/10.1016/0360-1315\(95\)00059-3](https://doi.org/10.1016/0360-1315(95)00059-3)
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426).
<https://doi.org/10.1145/1124772.1124855>
- Strijbos, J. W. (2011). Assessment of (Computer-Supported) Collaborative Learning. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/TLT.2010.37>
- Urhahne, D., Schanze, S., Bell, T., Mansfield, A., & Holmes, J. (2010). Role of the teacher in computer-supported collaborative inquiry learning. *International Journal of Science Education*.
<https://doi.org/10.1080/09500690802516967>
- Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Cambridge, MA: Harvard University Press.*, pp. 79–91.
- Wessner, M., & Pfister, H.-R. (2001). Group formation in computer-supported collaborative learning. *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work - GROUP '01*. <https://doi.org/10.1145/500286.500293>

Online Teacher Professional Development in ICT Integration in Tanzania: An Experience Report

Lucian Vumilia NGEZE* & Sridhar IYER

IDP in Educational Technology, Indian Institute of Technology Bombay, India

*luciangeze@iitb.ac.in

Abstract: Integration of Information and Communication Technology (ICT) in teaching and learning in schools has been hindered by many barriers, including lack of in-service teacher training, teacher beliefs and lack of infrastructure. In the effort to solve the challenge of teacher training in Tanzania, few in-service teachers, based on different parameters, are selected to participate in the face to face ICT training at the specified training centre. Scaling of such teacher training initiatives has been challenging over time. To reach more school teachers, we developed a ten-day online course, run over Moodle. A total of 134 teachers from all the regions of Tanzania registered and participated. Topics were developed based on the preference of the in-service teachers. Online surveys were used to collect qualitative and quantitative data before and after the course. Teachers were active during the duration of the course and many of them applied the skills in their schools by improving their teaching strategies, use of technology tools and sharing experiences with other teachers in their schools.

Keywords: Online course, ICT integration, teacher professional development, scaling.

1. Introduction

While governments and other stakeholders are merging efforts to integrate technology in schools, there are many barriers that have been highlighted to hinder effective integration. Ertmer (1999) identified two barriers: factors that are extrinsic to the teachers themselves such as lack of equipment, time, and lack of teacher training on ICT, and support from the management; and factors that are intrinsic to teachers such as attitude change.

Training teachers on ICT integration in Tanzania has been done on face to face mode, where selected in-service teachers meet at a training centre to be trained on a number of ICT modules over a period of intensive 5-10 days. In many cases, transfer of learning does not happen. To solve this, Tsai and Chai (2012) suggested designing topics that are relevant for immediate application, with activities that help in designing technology-enhanced lesson planning. However, to have a larger number of teachers, disruptive technologies such as MOOCs can come into the picture.

Despite the growing interest of MOOCs for Teacher Professional Development (TPD), implementation in the context of developing countries such as Tanzania has not taken place. Liyanagunawardena, Williams and Adams (2014) pointed out that varied levels of access to digital tools, ICT facilities; differences of language and accent from one country to the next are barriers to adoption of MOOCs in TPD in developing countries. We developed an online course with topics to help teachers in integrating technology in their teaching and learning. Online courses can be in different formats: synchronous, asynchronous and blended (Elliott, 2017). Advantages of using online learning for teachers include reflection on their practice (Rienties, Brouwer, and Lygo-Baker, 2013); participation in discussions about their practices (Rodesiler, 2017). To solve the challenge of diversity for teachers participating in global MOOCs for TPD, contextualization of the courses is important. This led to a need to develop a local online course to train teachers on ICT integration. This research was guided by the following research questions (RQs):

- i) What are the perceptions of teachers on the usefulness of the online course for their professional development?
- ii) How do teachers intend to apply their learning from the online course?
- iii) To what extent have the teachers applied their learning from the online course in their own practice?

2. Methodology

2.1 Online Course Design

Course design started with a needs analysis survey form for school teachers to suggest topics they would like to be trained on during their June 2018 vacation. This was done to 1) ensure that the selected topics were relevant to most of the school teachers who filled the survey, and 2) allow the different school teachers teaching diverse subjects to participate in the course. A total of 303 school teachers responded to the online survey mentioning up to three topics each. Analysis of the responses gave a total of 19 different topics. The first 8 topics with the highest frequencies were selected. We then developed a 10-day online course named "*Online ICT Course for School Teachers*".

The course consisted of short videos (4-8 minutes) with reflection spots which required participants to respond to questions related to the content. Each topic was followed by a short quiz to test understanding and a discussion forum with a focus question to guide the interactions. The course was run on MOODLE learning management system, from June 15 - 25, 2018. Videos were uploaded on YouTube Channel named *Tehama Shuleni*. The flow of the topics covered is shown in Table 1.

Table 1: *Selected topics and their purpose*

Day	Topic (Session)	Learning Objectives
1	Introduction to ICT	<ul style="list-style-type: none"> To introduce teachers to basic ICT tools, To introduce teachers to ICT integration models
2	Electronic Devices in Teaching and Learning	<ul style="list-style-type: none"> To use basic electronic tools in teaching and learning
3	Internet Connections to Laptops and Mobiles	<ul style="list-style-type: none"> To connect smartphone to a laptop To share internet bundle with others
4	Internet Browsing	<ul style="list-style-type: none"> To identify the different web browsers To browse the Internet using different browsers
5	Live Session 1 using YouTube Live	<ul style="list-style-type: none"> To establish an instructor-participant connect To respond to direct content-related questions from participants
6	Email basic Operations	<ul style="list-style-type: none"> To compose, reply, forward emails To attach different files to an email
7	Use of Visualizations in Teaching and Learning	<ul style="list-style-type: none"> To integrate videos, animations and simulations in teaching their subjects
8	Online Collaboration Tools	<ul style="list-style-type: none"> To collaborate using online tools to create common activities
9	Teacher Professional Development (TPD)	<ul style="list-style-type: none"> To introduce to MOOC platforms (Coursera, edX, Udacity) To enroll for free courses for their TPD
10	Live Session 2 using YouTube Live	<ul style="list-style-type: none"> To establish an instructor-participant connect To respond to direct content-related questions from participants

2.2 Procedure

Information about course starting was shared among teachers' networks. They registered and enrolled into the course. The instructor changes the role to the researcher during analytics and reporting of findings. Before the course started, the instructor developed needs analysis and entry surveys; deigned the course, customized MOODLE platform and uploaded contents. Participants were required to complete course modules, interact with peers in the discussion forum and participate in live

interactions. Impact analysis and learning analytics were done by the researchers after the course duration.

2.3 Sample Characteristics

A total of 134 teachers, majority with 4-10 years of experience; teaching different subjects from sciences, business and arts; majority aged between 31-40 years; from all the 26 regions of Tanzania mainland registered and participated in the course.

2.4 Data Collection

Data were collected using online course entry survey, course end survey and impact evaluation survey. Entry survey consisted of open-ended questions and few scale questions intending to find out subjects taught, teaching experience, and purpose to join the course while exit survey had some questions to measure the usefulness of the course, plan to use the knowledge and confidence to teach the topics. An impact evaluation survey was conducted at the end the two months after the training to collect data on knowledge application in schools.

2.5 Data Analysis

Quantitative analysis of the closed-ended questions was performed to generate frequencies and scales that were needed for determining research constructs such as usefulness and confidence. *interactive Stratified Attribute Tracking (iSAT)* was used to show the proportions of the different strata to one or a combination of the desired attributes of the course (Majumdar, Alse, & Iyer, 2014). iSAT is an interactive visual representation tool developed at the Indian Institute of Technology Bombay is to highlight transitions in a dataset.

Thematic analysis, following Braun & Clarke (2006), was applied to open-ended questions from all the three surveys to identify patterns. Each participant's response was taken as the unit of analysis. We first read all the data from participants to understand their responses. After familiarization with data, categories of responses that had the same meaning were generated. This made it easier for us to create six themes from the categories. These themes were then reviewed to make sure that there are not equal with the same meaning. Finally, we defined the last four themes that are presented in the result section.

3. Results

Results of the three research questions are explained in this section.

RQ1: What are the perceptions of teachers on the usefulness of the online course for their professional development?

At the end of the course, we asked about its usefulness on a scale of 1 (Not Useful at all) to 5 (Most Useful). Most of the participants with different teaching experiences found the course very useful. For example, 95% of all the participants with 4-10 years of experience found the course useful. This could be due to the way topics were selected taking into account diversity and applicability. The variation is shown in Figure 1.

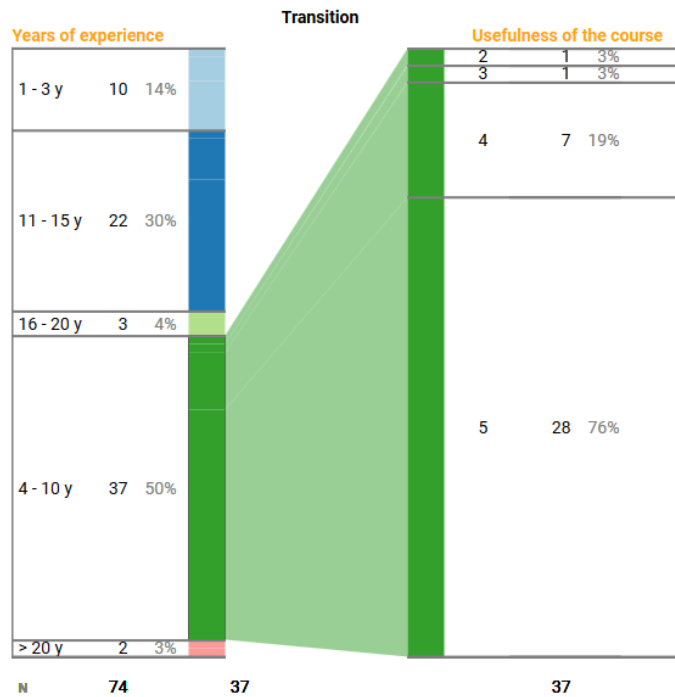


Figure 1: Usefulness of the course

RQ2: How do teachers intend to apply their learning from the online course?

Participants suggested different approaches to use the knowledge from the course. Figure 2 provides a variation in age group, usefulness and plan to use the knowledge gained. 69% of all the participants (in different age groups) who found the course to be very useful to them planned to use the knowledge in many ways. More than half of these (57%) planned to start using some technology in their teaching and learning activities.

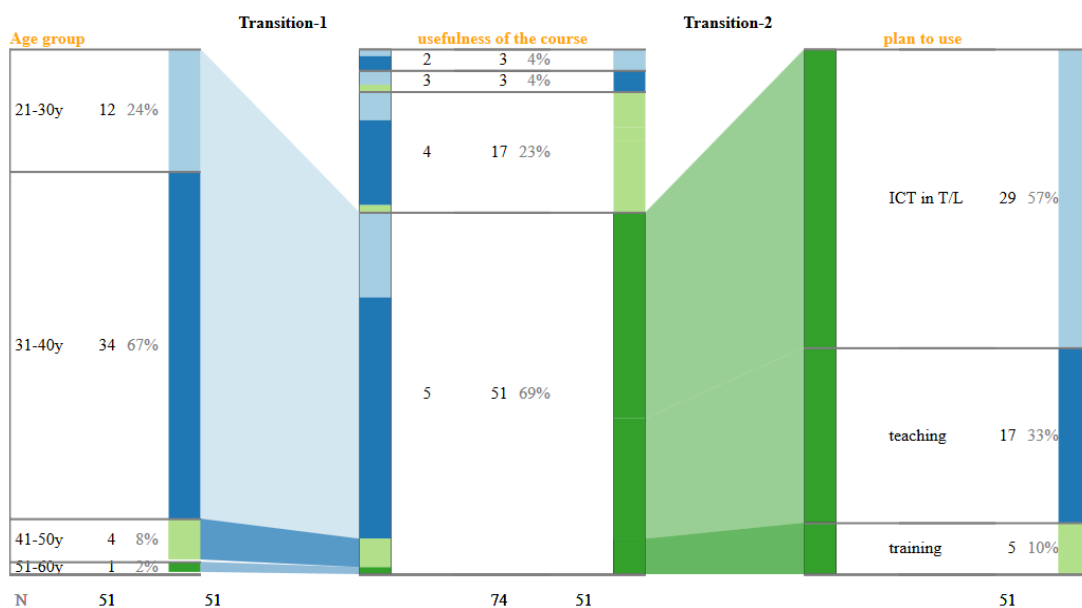


Figure 2: Approaches to use knowledge from the course

RQ3: To what extent have the teachers applied their learning from the online course in their own practice?

Two months after the course, we conducted an online evaluation to find out the impact of the course and how participants might have applied the knowledge from the course in their own settings. Thematic analysis was applied to the open ended question and generated four major themes. Majority of the participants who had access to the Internet used visualizations (animations, videos and simulations) in teaching different abstract concepts. Figure 3 shows the themes and the voices from participants.

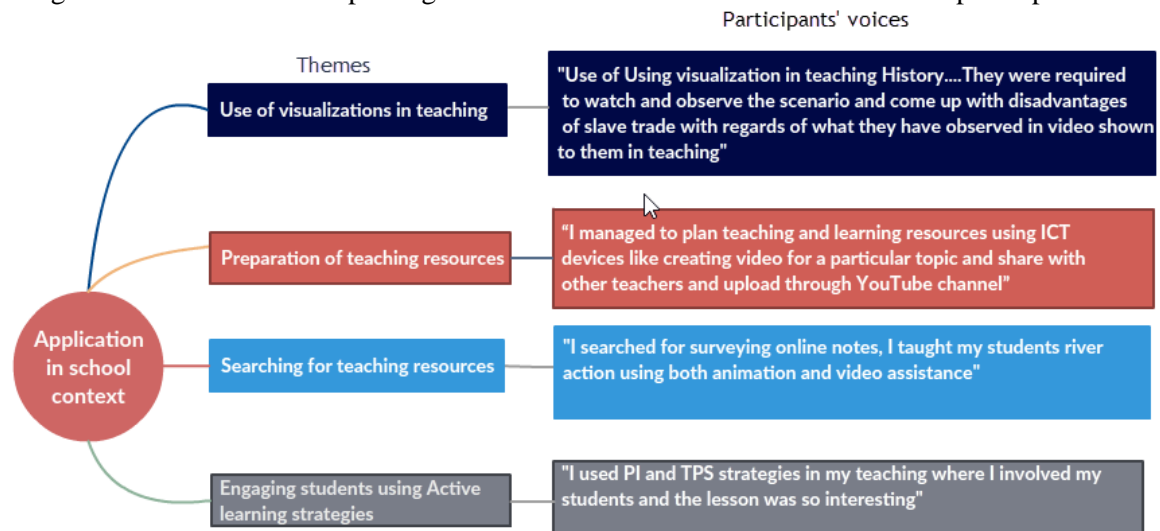


Figure 3: Themes generated from analysis of open-ended application questions

4. Discussion and Conclusion

This work presented an online course that extends the training of teachers to scale the formal face to face teacher training sessions in Tanzania. The design of the course involved teachers in the selection of relevant topics. This was one of the motivating factors to participate in the online course. It also helps to increase the course completion rate. The findings show that Training teachers on ICT integration is a step towards teachers' growth and improved teaching practices.

With 82% of the participants being new to online courses, interactions among the course participants in the discussion forum were during the course were high. The fact that participants knew that they would receive completion certificates kept them working hard. The first research question as shown that 95% of all the participants found the course useful. This level of satisfaction is consistent with the work by Nir and Bogler (2008) which says that teacher satisfaction with professional development courses is an important factor in transforming knowledge. This can be attributed to the fact that the teachers were involved in selecting the topics to be covered (Ketelhut, 2006), and hence making it relevant to the participants.

Previous studies have shown that demographic differences between teachers (age, years of experience) have an influence of teachers' readiness to use technology (Inan, & Lowther, 2010). However, in this case, teachers planned to use the knowledge in many ways: in teaching and training others and improving their teaching practices. This implies that, as long as the course is relevant to the participants, they can be able to transfer the learning in their contexts.

It is always important to evaluate the transfer of knowledge that happens after teachers have completed a professional development activity. Course impact evaluation was done two months later. Milheim (1994) pointed out factors such as structuring expectations and establishing awards as ways to increase learning transfer. In this regard, participants transferred the learning to their own schools in many ways including preparation of teaching resources and improving their own teaching practices. Depending on the course objectives, this study therefore suggests that evaluation of online courses need to be done as long as the number of participant is not massive.

Contextualization was implemented in two ways: topic selection and language. Course topics were suggested by the teachers and this ensured that those with the highest frequency were used in the design and development. On the other hand, participants followed up the lessons with ease since they were familiar with the language accent of the course instructor. In this way, participant diversity in the online course can be addressed and hence leading to effective learning.

Online courses, if well designed, based on the needs of the teachers, can help in improving teaching practices at the same time improving the performance of students. This course has shown that teachers are excited to learn specific topics of their choice in the online environment. They have, however, been able to transfer knowledge from the course in their teaching and learning activities. Another finding is that using the available resources such as smartphones, TPD initiatives can be extended to reach teachers who are in under-resourced schools. These online courses also open doors of sharing experiences among teachers across different subject domains. Through this course, teachers in Tanzania continue sharing experiences through their social groups and YouTube channel. This brings up a professional learning community of teachers making them grow professionally. Overall, this research shows that contextualization reduces the level of diversity and hence learning is enhanced. Future research work would look into how online courses can be deployed to benefit in-service teachers at large scale in the local context.

Acknowledgements

We thank CDEEP and UDOM Studios for video recording and editing; all school teachers in Tanzania who participated in this online course. Special thanks to Dr. Gargi Banerjee of Indian Institute of Technology Bombay, for the initial review of the paper.

References

- Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology*. *Qualitative research in psychology*, 3(2), 77-101.
- Elliott, J. C. (2017). The evolution from traditional to online professional development: A review. *Journal of Digital Learning in Teacher Education*, 33(3), 114-125.
- Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational technology research and development*, 47(4), 47-61.
- Inan, F. A., & Lowther, D. L. (2010). Factors affecting technology integration in K-12 classrooms: A path model. *Educational Technology Research and Development*, 58(2), 137-154.
- Ketelhut, D.J. 2006. "Core tensions in the evolution of online professional development". In *Online professional development for teachers: emerging models and methods*, Edited by: Dede, C. 237–263. Cambridge, MA: Harvard Education Press.
- Liyanagunawardena, T. R., Williams, S., & Adams, A. A. (2014). The impact and reach of MOOCs: a developing countries' perspective. *eLearning Papers*, 38-46.
- Majumdar, R., Alse, K., & Iyer, S. (2014, December). Interactive Stratified Attribute Tracking diagram for learning analytics. In *2014 IEEE Sixth International Conference on Technology for Education (pp. 138-139)*. IEEE.
- Milheim, W. D. (1994). *A comprehensive model for the transfer of training*. *Performance Improvement Quarterly*, 7(2), 95-104.
- Nir, A.E. and Bogler, R., 2008. *The antecedents of teacher satisfaction with professional development programs*. *Teaching and teacher education*, 24 (2), 377–386.
- Rienties, B., Brouwer, N., & Lygo-Baker, S. (2013). The effects of online professional development on higher education teachers' beliefs and intentions towards learning facilitation and technology. *Teaching and teacher education*, 29, 122-131.
- Rodesiler, L. (2017). For teachers, by teachers: An exploration of teacher-generated online professional development. *Journal of Digital Learning in Teacher Education*, 33(4), 138-147
- Tsai, C. C., & Chai, C. S. (2012). The "third"-order barrier for technology-integration instruction: Implications for teacher education. *Australasian Journal of Educational Technology*.

Educational Use of Spherical Video-based Virtual Reality: A Preliminary Study from the Teacher Perspective

Grace Chih-Nuo CHAO^{*}, Morris Siu-Yung JONG, Ching-Sing CHAI, & Eric Tsun-Hin LUK
Centre for Learning Sciences and Technologies
The Chinese University of Hong Kong, Hong Kong
^{*}*gracec@cuhk.edu.hk*

Abstract: Virtual reality (VR), which is regarded as one of the important educational technologies in recent years, has become more popular to be adopted in teaching and learning contexts in Hong Kong. The present study investigated the changes in teachers' pedagogical and technological knowledge, attitude and satisfaction after attending a teacher training course on the educational use of VR. Participants were 29 in-service teachers in Hong Kong taking the course. Data were collected from the pre- and post-participation in the course by using four scales: (i) *Pedagogical and Technological Knowledge*, (ii) *Technological Knowledge*, (iii) *Positive Attitude towards Harnessing VR*, and (iv) *Teacher Satisfaction*. The results showed good reliability of the scales and significant differences between the pre-test and post-test on using VR in their teaching. Participants had significant improvement in the knowledge of the VR technology and the ICT integration after the training course. They also showed a positive attitude towards harnessing VR in their teaching. Since the factor of content knowledge was not significant in the present study, we will reinforce the teacher training course based on the subject content via school-based development. Thus, the teaching training package will be more comprehensive and suitable for training teachers to adopt VR in practice.

Keywords: Virtual reality technology, EduVenture[®] VR, teacher training

1. Introduction

Knowledge construction is based on active experience (Jong, 2019). Piaget and Dewey indicated that educators' role was to shape a learners' experience and the learning environment to promote meaningful learning experience (Ornstein & Hunkin, 1998 cited in Huang, Rauch, & Liaw, 2010). Virtual reality (VR) can create immersive environment for students to experience what they are learning. It can improve students' interaction with the learning environment and make learning more vivid (Chang, Hsu, Chen, & Jong, 2018).

2. Literature Review

2.1 Definition of Virtual Reality

Virtual reality (VR) is broadly defined as "an interactive computer system that is so fast and intuitive that the computer disappears from the mind of the user, leaving the computer-generated environment as reality" technological (Kendrick, 1996, p.145). The characteristic of VR comprises the notion of collaboration and interaction among users in an intuitive computer-generated environment that appears real, with full integration of artificial intelligence products and information tools (Schwienhorst, 2002). In education, visuals are often used for receptive purposes, serving as scaffolds that help students to comprehend complex information (Cappello & Lafferty, 2015) because the use of images helps make ideas more concrete for students.

Nowadays, visualization technologies include anything from still images, infographics, and 3D printing, to VR tools (New Media Consortium, 2016). VR tools largely expand what a single image can

convey by creating immersive, sensational and interactive experiences without the travel. For instance, spherical video-based VR uses 360° images and videos to create virtual scenes. With a mobile device and a pair of VR goggles, users can turn their heads around in a full 360° circle and observe the environment in all angles. This VR tool combines technologies from computer graphics and vision and create virtual environments from either photographs or video images to immerse users in the VR environment (Lorenzo, Lledó, Pomares, & Roig, 2016; Passig et al., 2016; Chang et al., 2018). Faced with the increasing importance of experiential education, the present study uses VR technology to build an educational VR platform, allowing learners to access richer and more lively learning materials specifically accommodated in pedagogical needs of teachers, so as to enhance the effectiveness of experiential learning.

2.2 Application of VR in Experiential Learning

Experiential learning theory defines learning as “the process whereby knowledge is created through the transformation of experience” (Kolb, 2014, p. 41). The four steps of the experiential learning cycle are concrete experience, reflective observation, abstract conceptualization, and active experimentation (Baker et al., 2002). The immersive nature of educational use of VR can be a useful tool to promote and enhance experiential learning (L.M. Pilgrim & J. Pilgrim, 2016). In the virtual tour, students can observe the people, places, and the environment intimately and repeatedly. This may create an enriched experience which students can make sense. For example, students can be immersed in a cave environment where bats live and generate much questions and consequently inquiry about the living environment and the creatures. Thus, VR could be an essential teaching tool for enhancing teaching and learning experience and effectiveness in the classroom. The immersive learning environment provides individual opportunities for experience and reflection (J.M. Pilgrim & J. Pilgrim, 2016). For example, Google Expeditions is one popular mobile application for teachers to virtually bring students to every corner of the world without setting foot outside the classroom.

2.3 Significance of VR in Teacher Education

VR in education is becoming a prominent feature at schools in recent years. However, very few studies examined how teachers develop their competencies in designing VR for classroom use teachers can provide opportunities for students to engage in the environment with VR tools only if they are able to understand the affordances of VR, what are students’ gaps in learning a topic, and how VR can help to bridge understanding gaps. In other words, teacher need VR-based pedagogical content knowledge. This is a form of Technological Pedagogical Content Knowledge (TPACK) framework which is about how teacher integrate information and communication technology (ICT) in their teaching (Chai, Koh, Tsai, & Tan, 2011). The TPACK framework provides new directions for teacher educators in solving the problems associated with integrating ICT into classroom teaching and learning (Hewitt, 2008). In the present study, we emphasis how to teach effectively through educational VR combined with pedagogical approaches to address language learning problems, i.e., we attempt to build the teachers’ VR-PCK. Thus, the technological knowledge such as techniques of making education VR resources, design of VR-based lesson plan is essential in the teacher training course. From the teachers’ perspective of educational VR, it is significant to enhance the participating teachers’ satisfaction and have a positive attitude towards harnessing VR.

2.4 EduVenture[®] VR

EduVenture[®] VR (EVVR) is an educational VR platform developed by the technical team of Centre for Learning Sciences and Technologies of The Chinese University of Hong Kong (Geng, Jong, Luk, & Jiang, 2018; Jong, Luk, Leung, & Poon, 2018). It can design SVVR learning resources by combining 360° images and videos of real scenes and learning content. In contrast with most of the VR entertainment platforms in the industry, EVVR provides an easy-to-use authoring environment for teachers to tailor-make interactive VR field trips that suit their subject content and students’ learning needs. The interactive possibilities on EVVR include voice answering, object searching, etc. which are unique to other VR platforms and greatly accommodate the pedagogical needs of teachers. Students can

then experience the VR scenes by using their mobile phone and low-cost VR equipment (e.g. Google Cardboard, iQiyi, etc.).

The assessment module of EVVR (Retriever) lets teachers view and compare the feedback of students and keep tracking the learning progress of the whole class. The database module contains databases of VR learning materials, learning portfolios, student information, and quiz results. With mobile devices and VR goggles, students can experience outdoor learning in VR, which makes teaching and learning more interesting and flexible.

2.5 Teacher Training of Using VR in Teaching

The training course in the present study focuses on how teachers may enhance learning and teaching effectiveness by adopting VR technology and relevant strategies in the classroom. First, the course provides the definition and global trend of VR technology in education. It provides techniques and essential skills of using VR equipment, designing VR learning resources (i.e. including 360 photo and movie capture in the University campus and 360 movie clip transfer, editing and publishing on different VR platforms) and implementing learning and teaching strategies in the classroom. Composing VR Learning and Teaching in EVVR and school management in EVVR were also introduced at the end of the course. Three research questions framed the present study:

Q1: Will teachers' pedagogical and technological knowledge be enhanced significantly after participating in a teaching training course?

Q2: Will teachers' attitudes towards harnessing VR in education improve significantly after participating in a teaching training course?

Q3: Will teachers' satisfaction be enhanced significantly after participating in a teaching training course?

3. Method

3.1 Participants

The participants were 29 in-service primary and secondary school teachers who attended a 3-hour teacher training course on VR for enhancing students' Chinese language literacy held by CLST of CUHK. Of the total number, 27.6% were male, 69% were female. Their mean of age range was 31-35 (27%). Fifty-nine per cent were teachers and 28% were senior teachers, 4% were teaching assistants, 9% were unknown. Half of them have been used e-learning platform such as Kahoot, Nearpot and VR 1-3 years, 31% of them have been used 4-10 years and a few of them have been used for over 10 years.

3.2 Data Collection Procedure and Measures

Self-administered questionnaires were collected at the beginning and the end of the teacher training course. The instruments used in this study was adapted from Schmidt et al. (2009), Koh et al. (2010) and Chai et al. (2013). The initial instrument consists of 30 items. Four factors were included in this questionnaire, namely (i) Pedagogical and Technological Knowledge, (ii) Technological Knowledge, (iii) Positive Attitude towards Harnessing VR, and (iv) Teacher Satisfaction. Teachers responded using 1 = strongly disagree to 7 = strongly agree. A total of 28 valid responses were matched (i.e. pre-test and post-test) and adopted as sample for subsequent analyses. The preliminary analyses were conducted, including descriptive statistics, reliability of instruments, paired sample t-test and effect size of the variables, using SPSS 25.

4. Results

4.1 Descriptive Statistics

Table 1 displays a summary of the reliability coefficients, mean (*M*) and standard deviations (*SD*), and effect size (*d*) of the variables. The reliability of the four scales was satisfactory, i.e. the range of the

scales was from ($M = 6.05$, $SD = 0.42$, Cronbach's $\alpha = 0.95$) (Factor name: Pedagogical and technological knowledge – content knowledge) to ($M = 5.09$, $SD = 1.10$, Cronbach's $\alpha = 0.97$) (Factor name: Technological knowledge).

Table 1 Mean, standard deviation and reliability of the Teachers on VR training (post-test) ($N = 29$)

Variables	<i>N</i>	No. of item	<i>M</i>	<i>SD</i>	α	Sample Item
Pedagogical and technological knowledge						
Content knowledge	29	4	6.05	0.42	0.95	I have enough professional knowledge in Chinese language.
Knowledge about using technology to teach	29	4	5.39	0.87	0.89	I am able to guide the students to use technology (VR) to examine the hidden details in VR.
ICT integration knowledge	29	6	5.58	0.51	0.88	I am able to design student-centred courses, and appropriately integrate teaching content, teaching techniques and teaching methods.
Technological knowledge	29	3	5.09	1.10	0.97	I am able to independently use (operate) VR technology in the Chinese writing class.
Positive attitude towards harnessing VR	29	4	5.38	1.02	0.996	Using VR in teaching Chinese writing class can help my students learn efficiently.
Teacher satisfaction	29	5	5.30	0.88	0.98	I like teaching this Chinese writing class

Note: 1= strongly disagree to 7= strongly agree

4.2 Significance Test

Table 2 displays the results of the paired sample t-test and effect size of the four scales. Five out of six pairs of t-test showed significant differences between pre-test and post-test. For example, there was a significant difference in the scores of Technological Knowledge (pre-test: $M = 3.54$, $SD = 1.31$) and (post-test: $M = 5.13$, $SD = 1.10$), $t(27) = -6.43$, $p = .000$). In addition, the range of the effect size was from ($d = 0.06$ to $d = 1.31$) of the scales.

Table 2 Paired sample t-test of the teachers on VR training ($N = 29$)

Variables name	<i>M</i>	<i>N</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>P-value</i>	Cohen's <i>d</i>
Pedagogical and technological knowledge							
Pre-content knowledge	6.02	28	.50				
Post-Content knowledge	6.05	28	.43	-.89	27	.381	0.06
Pre- Knowledge about using technology to teach	4.29	28	1.32				
Post- Knowledge about using technology to teach	5.40	28	.88	- 4.98	27	.000	0.99
Pre-ICT integration knowledge	4.83	28	.83				
Post-ICT integration knowledge	5.63	28	.42	-5.03	27	.000	1.22
Pre- Technological knowledge	3.54	28	1.31				
Post- Technological knowledge	5.13	28	1.10	-6.43	27	.000	1.31
Pre- Positive attitude towards harnessing VR	4.69	28	1.18				
Post- Positive attitude towards harnessing VR	5.43	28	1.00	-3.86	27	.001	0.68
Pre-Teacher satisfaction	4.63	28	.71				
Post-Teacher satisfaction	5.34	28	.86	-4.21	27	.000	0.90

Note: 1= strongly disagree to 7= strongly agree

5. Discussion and Conclusion

The results showed a good reliability of the scales and significant differences between the pre-test and post-test on using VR in the teacher training course. Participants showed significant improvement in the knowledge of the VR technology and the ICT integration after the teacher training course. They also showed positive attitude towards harnessing VR in their class. Since the factor of content knowledge was not significant in the present study, we will strengthen the teacher training course based on the subject content via school-based development.

Regarding the teacher training program, it is significant to facilitate teacher development for better integration of ICT. The technological pedagogical content knowledge (TPCK or TPACK) framework (Mishra & Koehler, 2006) has reflected a paradigm shift in the recent studies (e.g., Jong et al., 2018). Therefore, teacher professional development requires cycles of iterative lesson design, implementation, and reflective refinement to resolve. This is especially for emerging technologies where information about how they could be employed in the classroom.

Teacher satisfaction is another focus in the present study. Teachers showed significant improvement after the intervention. Teachers feel satisfied after completing this teacher training course because it can improve students' learning achievement with the implement of VR in a pedagogical design. Teachers' level of satisfaction on the instruction programs will be related to what they perceive regarding various aspects of the teacher training programs (e.g., school-based or non-school-based, whether teachers' actual needs and concerns are being addressed, etc.) (Jong & Tsai, 2016).

Huang, Chen, & Chou (2016) indicated that experiential learning differs from teacher-centered instructive learning which focus on self-regulating judgment, free-thinking, and personal experience. Through interactive learning processes, students gain personal experience from which they derive an understanding of the core aspects of learning tasks and explore the correlation between activity concepts and implications. Learners integrate their learning experience into their lives, that transforms their attitudes and prompts further reflection on their behaviors (Jong, 2017). Experiential learning theory emphasizes the relationship between concrete experience and learning and teaching. In the present study, it showed how the VR technology enriched the teacher development for better integration of ICT, and how they gained more technological knowledge. The participants also have more positive attitude towards harnessing VR in the classroom after the intervention.

It is planned to collect more data to further examine the scale validation. Also, more advanced statistics will be conducted in the future, such as factor analysis and multiple regression analysis. Since the factor of content knowledge was not significant in the present study, we will reinforce the teacher training course based on the subject content via school-based development. Thus, the teaching training package will be more comprehensive and suitable for training teachers to adopt VR in practice.

Acknowledgement

We would like to thank The Hong Kong Jockey Club Charities Trust for supporting the project of Jockey Club VR Project for Enhancing Chinese Language Literacy.

References

- Baker, A. C., Jensen, P. J., & Kolb, D. A. (2002). *Conversational learning: An experiential approach to knowledge creation*. Westport, Connecticut: Quorum Books.
- Cappello, M., & Lafferty, K. E. (2015). The roles of photography for developing literacy across the disciplines. *The Reading Teacher*, 69(3), 287–295.
- Chai, C. S., Chin, C. K., Koh, J. H. L., & Tan, C. L. (2013). Exploring Singaporean Chinese language teachers' technological pedagogical content knowledge and its relationship to the teachers' pedagogical beliefs. *The Asia-Pacific Education Researcher*, 22(4), 657–666.
- Chai, C. S., Koh, J. H. L., Tsai, C. C., & Tan, L. L. W. (2011). Modeling primary school pre-service teachers' Technological Pedagogical Content Knowledge (TPACK) for meaningful learning with information and communication technology (ICT). *Computers & Education*, 57(1), 1184–1193.
- Chang, S. C., Hsu, T. C., Chen, Y. N., & Jong, M. S. Y. (2018). The effects of spherical video-based virtual reality implementation on students' natural science learning effectiveness. *Interactive Learning Environments*, 1-15.

- Dunlap, J., Dobrovolny, J., & Young, D. (2008). Preparing eLearning designers using Kolb's model of experiential learning. *Innovate: Journal of Online Education*, 4(4), Article 3.
- Geng, J., Jong, M. S. Y., Luk, E. T. H., & Jiang, Y. (2018). Comparative study on the pedagogical use of interactive spherical video-based virtual reality: The EduVenture-VR experience. *Proceedings of the 4th International Symposium on Educational Technology (ISET 2018)* (pp. 261–263). Osaka, Japan.
- Huang, T. C., Chen, C. C., & Chou, Y. W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, 96, 72–82.
- Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3), 1171–1182.
- Hewitt, S. (2008). *The British war on terror: terrorism and counter-terrorism on the home front since 9/11*. London: Continuum International Publication Group.**
- Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3), 1171–1182.**
- Kendrick, M. (1996). *Cyberspace and the technological real*. In R. Markley (Ed.), *Virtual realities and their discontents*. Baltimore, London: Johns Hopkins University Press.**
- Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.
- Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2010). Examining the technological pedagogical content knowledge of Singapore pre-service teachers with a large-scale survey. *Journal of Computer Assisted Learning*, 26(6), 563–573.**
- Jong, M. S. Y. (2017). Empowering students in the process of social inquiry learning through flipping the classroom. *Educational Technology & Society*, 20(1), 306–322.**
- Jong, M. S. Y. (2019). Sustaining the adoption of gamified outdoor social enquiry learning in high schools through addressing teachers' emerging concerns: A three-year study. *British Journal of Educational Technology*, 50(3), 1275–1293.**
- Jong, M. S. Y., Chan, T., Hue, M. T., & Tam, V. (2018). Gamifying and mobilising social enquiry-based learning in authentic outdoor environments. *Educational Technology & Society*, 21(4), 277–292.**
- Jong, M. S. Y., Luk, E. T. H., Leung, J. K. P., & Poon, S. S. K. (2018). *Eduventure VR* [Computer platform]. Hong Kong: Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong: Hong Kong. Retrieved August 29, 2019 from <http://vr.ev-cuhk.net>.
- Jong, M. S. Y., & Tsai, C. C. (2016). Understanding the concerns of teachers about leveraging mobile technology to facilitate outdoor social inquiry learning: The EduVenture experience. *Interactive Learning Environments*, 24(2), 328–344.
- Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers & Education*, 98, 192–205.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Ornstein, A. C., & Hunkins, F. P. (1998). *Curriculum: Foundations, principles, and issues* (3rd ed.). Needham Heights: Allyn and Bacon.
- Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers & Education*, 98, 192–205.
- Passig, D., Tzuriel, D., & Eshel-Kedmi, G. (2016). Improving children's cognitive modifiability by dynamic assessment in 3D immersive virtual reality environments. *Computers & Education*, 95, 296–308.
- Pilgrim, J. M., & Pilgrim, J. (2016). The Use of Virtual Reality Tools in the Reading-Language Arts Classroom. *Texas Journal of Literacy Education*, 4(2), 90–97.
- Schwienhorst, K. (2002). Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. *Simulation & Gaming*, 33(2), 196–209.
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, 42(2), 123–149.
- Sherman, W. R., & Craig, A. B. (2003). *Understanding virtual reality*. San Francisco, CA: Morgan Kauffman.

A Preliminary Case Study of Promoting Teacher Professional Development via Post-Lesson Debriefing with the Support of Sokrates Cloud Services

Oskar KU*, Jen-Kai LIANG, Sung-Bin CHANG, & Michael WU
HABOOK Group, Taiwan
*oskar@habook.com.tw

Abstract: During the process of teacher professional development, debriefing serves as a critical element, which not only helps teachers reflect on their own teaching practice but also helps all peer teachers promote their professional development. A successful debriefing session requires teachers to recollect their experience of a taught lesson. The more details are recollecting, the more issues will be identified, which, in turn, may enhance the effects of debriefing. In this vein, the authors develop a system, Sokrates Video Player (Sokradeo), which incorporates the abilities of automatic data collection and analysis, enhanced video player, and analytic dashboards to help teachers reconstruct their experience and identify critical issues of their teaching. A preliminary study was conducted to evaluate the effects of using Sokradeo to support post-lesson debriefing. Although the results indicated that Sokradeo might enhance the efficiency of debriefing and helped novice teachers master the use of smart classroom tools within a relatively short time, the study was not a well-controlled experiment due to the fact that the school schedule did not allow such an experiment. Hence, rigorous studies should be further conducted to investigate how Sokradeo contributes to the process of post-lesson debriefing.

Keywords: teacher professional development, Sokrates, Sokradeo, debriefing, AI engine

1. Introduction

Debriefing, a postexperience analytic process (Lederman, 1984; 1992) originated from army (Gardner, 2013), plays an imperative role in the process of professional development in various fields. Due to its simplicity and effectiveness in improving performance (Tannenbaum & Cerasoli, 2013), the debriefing technique was then applied to other career fields, such as nursing and healthcare (Dreifuerst, 2015), human resource management (Reyes, Tannenbaum, & Salas, 2018), and software development (Collier, DeMarco, & Fearey, 1996).

In addition to the above-mentioned fields, debriefing is also regarded as an effective technique for improving teachers' teaching capabilities. Such a technique provides teachers with opportunities to gain a deep insight of their teaching practice, which, in turn, may not only help teachers enhance their teaching and research skills but also facilitate students' learning thanks to the improvement of teachers' instructional plans and implementations (Hail, Hurst, & Camp, 2011). In schools, (post-lesson) debriefing is usually regarded as a method of inquiry aiming at helping inquirers explore and look into the critical issues that are not easy to be identified by the inquirers themselves via disclosing personal experiences to peers (Lincoln & Guba, 1985).

A debriefing session consists of seven elements: (1) the debriefer (guide), (2) the participants to be debriefed, (3) participants' experience, (4) the impact of participants' experience, (5) the recollection of participants' experience, (6) reporting mechanisms, and (7) time for debriefing (Lederman, 1991). More specifically, there would be a debriefer who coordinates the entire debriefing session within a specific duration. The debriefer facilitates the process of experience sharing and manages the foci of the experience to be debriefed. For a post-lesson debriefing session, the participants to be debriefed are the teachers who have taught a planned lesson. A debriefing session starts with experience sharing. The

participants are requested to recount their personal experience of the planned lesson. The debriefer may ask some questions to facilitate the debriefing process. In addition to the debriefer–participant conversation, the debriefer would also invite peer teachers to express their opinions about the strengths and the weaknesses of the participant’s lesson. Peer teachers are also encouraged to interact with the participants after the participants finish their experience sharing (Lewis, 2002). Such interaction among peers may not only help the participants recall more details but also provide the participant with different perspectives for self-reflection (Hail, Hurst, & Camp, 2011).

Table 1

The Three Phases of Debriefing Framework

Phase 1	Review and reflection	The participants share their experience, including facts, feeling, and personal opinions towards specific events. Peers are encouraged to interact with participants to help them elicit more detail and identify the pros and cons of their teaching.
Phase 2	Understanding and analysis	The identified pros and cons are discussed to explore the underlying causes. The debriefer may raise questions to stimulate the participants and peer teachers to think deeply and broadly and link up with their past experiences so that they may gain a deeper understanding about the topics being discussed and explore the impact of their experience by which the participants’ performance on future tasks might be influenced.
Phase 3	Summary and generalization	This phase reviews what has learned from the experience and discusses how to apply the learned lesson in the future.

Past studies from different disciplines involve various forms of debriefing, which implies there is no a standard format for debriefing (Spillett, 2003). Even so, there are three phases (Table 1) which forms the main debriefing framework (Gardner, 2013; Lederman, 1992; Rudolph, Simon, Raemer, & Eppich, 2008). In this framework, experience sharing constitutes the first step of debriefing. The latter steps rely on the participants’ experience as the source for discussion. In other words, the quality of recollected experience may decide the result of a debriefing session – success or failure. Hence, it is critical to help participants recollect their experience. Traditionally, the debriefer may prepare some tools, such as a survey or questionnaire, to facilitate the recollection of participants’ experience, (Gardner, 2013). Although these tools may help participants recall more details, they still rely on participants’ memory. Alternatively, we started to consider if there is a tool which provides objective and detailed information to facilitate debriefing. Actually, there is a tool which may fulfill the requirements: videos.

Flanagan (2008) suggested that videos may be served as a tool for facilitating experience recollection. Past studies indicated that teachers behaved differently on recalling their teaching experience if they were provided with the video of their teaching. Without the support of videos, the recalled experience mainly focused on classroom management and personal performance. However, teachers would put greater emphasis on reviewing how they deliver their instruction and students’ responses when they were provided with the video clips of their lecturing (Rosaen, Lundeberg, Cooper, Fritzen, & Terpstra, 2008). Although videos provide detailed information, teachers may have to interrupt debriefing and spend time on seeking and playing the videos, which may become another problem.

2. The Design of Sokradeo: A Tool for Facilitating the Process of Debriefing

To address the above issues, an interactive teaching analytic system, the Sokrates Video Player (Sokradeo) (Fig. 2), was developed. Sokradeo is an interactive player which integrates videotaped lectures, smart tags, analytic dashboards, and a comment collector. These features are designed to reconstruct the whole lecture and foster the recollection of teachers’ teaching experience during the process of debriefing. The following subsections describe the design details of the Sokradeo.

2.1 The Design Principles of Sokradeo

Three principles are applied to develop the system; the system should be (1) automatic, (2) intelligent, and (3) interactive. First, it is a labor-intensive task to record all the events occurred in classrooms with traditional methods (e.g. Gall & Acheson, 1980; Flanders, 1966). Thanks to the advancement of technology, we are now able to operate various devices to record and process necessary data of a lesson automatically and simultaneously. Second, the data should be processed intelligently. The raw data is processed with AI algorithms tuned by human experts. These algorithms imitate how human experts evaluate teachers' performance and then visualize the results with analytical dashboards. Third and the last, instead of merely serving as a one-way information deliverer, the system should provide users with an interactive interface so that users may get additional information through interacting with the system. As shown in Figure 1, the interface of Sokradeo is composed of four different panels; they are (1) a video player, (2) smart tags of teaching behavior, (3) smart tags of comments, and (4) analytic dashboards.

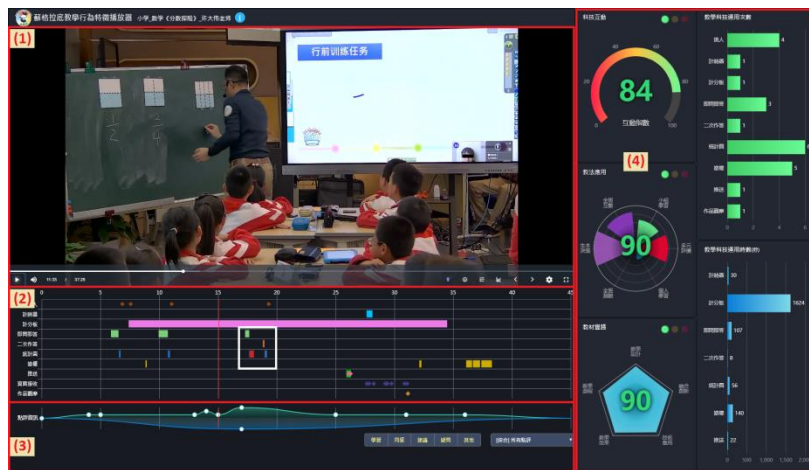


Figure 1. The Sokrates Video Player.

2.2 Video Player with Smart Tags of Teaching Behavior

The left three panels can be operated jointly to provide users with an extended controlling interface for playing lecture videos. The smart tags in panel 2 are generated automatically after the Sokrates Cloud Services analyzed teachers' teaching behavior data. This panel serves as a navigator guiding users what interaction tools the teacher executed at a specific time. These "smart tags" can be used to interact with the video directly. Users can click on the smart tags to trigger the video jump to the corresponding time point. For example, if a teacher wants to share the experience of carrying out a pedagogical pattern (the framed area in panel 2 of Figure 1), the teacher can click on the smart tags in the frame and describe her/his experience with the support of corresponding video.

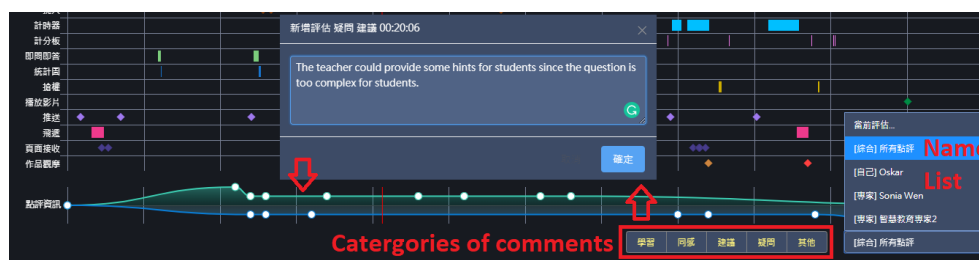


Figure 2. The comment collector.

2.3 Comment Collector

In addition to the automatically generated smart tags, users can manually add annotations to a video. These annotations then become the other type of smart tags (smart tags of comments) that interact with

videos. Experts and peer teachers who are invited to evaluate a taught lesson can leave public comments. Other users can add private comments for personal reference. For example, teachers who taught a planned lesson can add some annotations to the video so that they could use the annotations to remind themselves what to share during a debriefing session. Figure 2 illustrates the interface of the comment collector of Sokradeo.

2.4 Analytic Dashboards

The panel 4 shown in Figure 1 demonstrates the analytic dashboards which summarize how teachers delivered their lectures in smart classroom environment. These dashboards with the orders from top to bottom, left to right are (1) technological interaction index, (2) pedagogical application index, (3) content implementation index, (4) the frequencies of interactive technologies, and (5) the accumulated time on using interactive technologies. Unlike the other four automatically generated dashboards, the third dashboard, content implementation index, is designed to complement the other two automatically rated indices since the Sokrates AI engine still has its own limitations. Hence, we implemented this index rated by human experts to evaluate how teachers prepare, implement, and deliver their lectures.

3. A Preliminary Study of Using the Sokradeo

In order to evaluate the effects of using Sokradeo to support debriefing sessions and find out the shortcomings of system design, a preliminary study was conducted in 2018 in an elementary school. The study lasted for three months (from September to November). There are totally 83 teachers, of which 39 teachers have never taught in a smart classroom before participating this study while the other 44 teachers have one to three years' experience of using a smart classroom, participated in this preliminary study (N = 83). The details of this evaluation are presented in the following subsections.

3.1 Method

Teachers' technological interaction and pedagogical application indices (T index & P index) of their lessons are used as the indicators to observe the change of teachers' performance on their teaching. T-tests are applied to compare the performance between teachers with no and at least one year's experience on teaching in a smart classroom. Due to the fact that this preliminary study is conducted in a regular semester, a rigorous experiment was not able to be conducted since we could not alter school's schedule as well as interfere how teachers carried out their teaching plans. Hence, a qualitative report from the teacher participants, which explained how teachers utilized Sokradeo to conduct post-lesson debriefing to promote teacher professional development was adopted to serve as a supplemental material to discuss the possible advantages of debriefing with the support of Sokradeo.

3.2 Quantitative Results

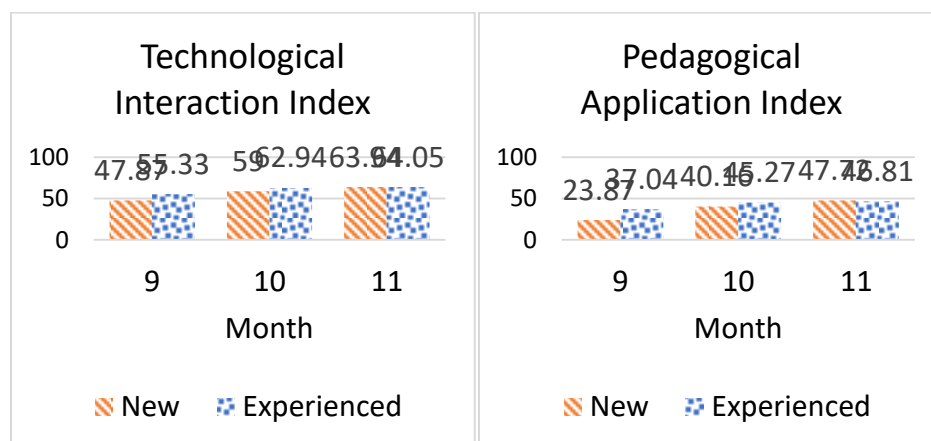


Figure 3. The Mean Scores of the Technological Interaction of Pedagogical Application Indices.

Teachers' performance on the two indices is shown in Figure 3. All teachers gained improvement on both indices from September to November. In addition, the teachers without any experience in teaching in smart classrooms seemed to perform almost equally on the two indices when comparing to their experienced peers in the third month.

In order to confirm whether the observation is correct, t-tests are applied to compare the differences of the T index and P index between the novice and experienced teachers. The results of independent-samples t-tests are shown in Table 2. The results indicate that there are significant differences in both indices in September and October between the novice and experienced teachers with $ps < .001$ while no significant differences are found in the two indices in November between the novice and experienced teachers ($ps > .05$).

Table 2

The Result of T-tests between Teachers with Different Levels of Experience on using Smart Classrooms

Experience	N	September, 2018		October, 2018		November, 2018		
		T index	P index	T index	P index	T index	P index	
0	39	M	47.87	23.87	59.00	40.16	63.94	47.72
		SD	19.25	27.17	19.47	24.79	18.32	24.93
≥ 1 year	44	M	60.45	46.07	66.27	49.61	64.18	45.73
		SD	19.24	27.78	17.17	22.05	19.00	24.89
		t	7.98***	9.84***	6.14***	6.24***	0.10	0.64

* $p < .05$; ** $p < .01$; *** $p < .001$

3.3 Qualitative Response from Teachers

Three of the teacher participants who coordinated the teacher professional development group created a report to describe how they conducted debrief sessions in their school with the support of Sokradeo. Before this study, teachers in this school had already collected thousands of videotaped lectures. Among these videos, several videos or video clips with model examples were identified and saved as example lists. When this study started, those model clips are used as an additional resource for enhancing the effects of debriefing. Teachers conducted debriefing sessions regularly with the use of Sokradeo. The teachers indicated that there was usually only one case could be debriefed during a traditional debriefing session. However, they might able to debrief at least two or three cases thanks to the features of Sokradeo that help teachers locate critical episodes quickly and precisely.

4. Discussion

The quantitative results indicate the novice teachers might have grasped the skills of teaching in smart classrooms within two months. More specifically, the T index demonstrates the level of a teacher's familiarity with lecturing in smart classrooms while the P index represents the level of a teacher's knowledge of applying the smart classroom tools to carry out various pedagogical model in his/her lesson. In fact, it may take longer time to master the pedagogical skills since teachers might spend time on mastering the skill of using smart classroom tools first. However, the novice teachers performed almost equally to their experienced peers in both aspect after two months. This may be due to the effect of debriefing since debriefing may help teachers not only master the use of tools but also a gain deeper understanding on integrating the tools into the lesson design optimally (Groth, 2011). The qualitative responses from teachers might explained why the novice teachers could demonstrate comparable levels of performance to the experienced teachers in both technological and pedagogical aspects in two months since the efficiency of debriefing was improved.

5. Conclusion

Past studies provided a lot of evidence of using debriefing to promote personal performance in various domain of professions, including teaching. Lesson study is one of the common methods adopted by teachers to promote their professional development, and debriefing constitutes the final step of lesson study, which help teachers reflect on if their teaching fulfilled their plans. More importantly, the results of debriefing create the foundation, which may induce the next lesson study cycle (Groth, 2011).

This study developed a system, Sokradeo, of which an important application is supporting the process of debriefing. Since the teachers being debriefed have to reconstruct their experience of a taught lesson, Sokradeo provides a feature to help teachers locate critical episodes efficiently. A preliminary study was conducted to examine the effects of using Sokradeo to support the process of debriefing to promote teacher professional development. The results indicated that novice teachers achieved similar levels of performance when comparing to their experienced peers after two months. Such a result might be contributed by the integration of debriefing and Sokradeo. Since a controlled experiment was not able to be conducted, the true effect of Sokradeo could not be confirmed yet. We only got a message which implied the Sokradeo might improve the efficiency of debriefing from teachers' response. Hence, a controlled experiment should be conducted to verify the effects of Sokradeo on supporting post-lesson debriefing. A rigorous qualitative study should also be conducted to find out the deep meaning of how teachers interact with Sokradeo to improve their professional development.

Acknowledgements

We would like to thank the reviewers for their insightful comments and all the people who participated in this project.

References

- Collier, B., DeMarco, T., & Fearey, P. (1996). A defined process for project post mortem review. *IEEE software*, 13(4), 65-72.
- Dreifuerst, K. T. (2015). Getting started with debriefing for meaningful learning. *Clinical Simulation in Nursing*, 11(5), 268-275.
- Flanagan, B. (2008). Debriefing: theory and techniques. In R. H. Riley (Ed.), *Manual of simulation in healthcare* (pp. 155-170). New York, NY: Oxford University Press.
- Flanders, N. A. (1966). *Interaction Analysis in the Classroom: A Manual for Observers*. Ann Arbor, MI: University of Michigan
- Gall, M. D., & Acheson, K. A. (1980). *Clinical Supervision and Teacher Development: Preservice and Inservice Applications*. New York, NY: Longman.
- Gardner, R. (2013). Introduction to debriefing. *Seminars in Perinatology*, 37(3), 166-174.
- Groth, R. E. (2011). Improving teaching through lesson study debriefing. *Mathematics Teacher*, 104(6), 446-451.
- Hail, C., Hurst, B., & Camp, D. (2011). Peer Debriefing: Teachers' Reflective Practices for Professional Growth. *Critical Questions in Education*, 2(2), 74-83.
- Lederman, L. C. (1984). Debriefing: A critical reexamination of the postexperience analytic process with implications for its effective use. *Simulation & Games*, 15(4), 415-431.
- Lederman, L. C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming*, 23(2), 145-160.
- Lewis, C. C. (2002). *Lesson study: A Handbook of Teacher-Led Instructional Change*. Philadelphia, PA: Research for Better Schools, Inc.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Reyes, D. L., Tannenbaum, S. I., & Salas, E. (2018). Team development: The power of debriefing. *People & Strategy*, 41(2), 46-51.
- Rosaen, C. L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2008). Noticing noticing: How does investigation of video records change how teachers reflect on their experiences? *Journal of Teacher Education*, 59(4), 347-360.
- Rudolph, J. R., Simon, R., Raemer, D. B., & Eppich, W. J. (2008). Debriefing as formative assessment: closing performance gaps in medical education. *Academic Emergency Medicine*, 15(11), 1010-1016.
- Spillett, M. A. (2003). Peer debriefing: Who, what, when, why, how. *Academic Exchange Quarterly*, 7(3), 36-41.
- Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors*, 55(1), 231-245.

Exploring K-12 Students' Acceptances of a Computational Making Programme: A Case Study in Singapore

Longkai WU*, Chee-Kit LOOI & Sujin HE

National Instituted of Education, Nanyang Technological University, Singapore

*longkai.wu@nie.edu.sg

Abstract: In this paper, we describe research conducted around a 180 hour programme designed to introduce secondary school students to computational making through app development. The quantitative data analysis indicated that most of the participating students, whether male or female, seem to have developed a high level of interest and acceptance of computational making activities. The correlational and SEM analysis indicated that although these students, from various schools, have benefited in different dimensions. Design thinking readiness and coding readiness should be highlighted as the prominent factors derived from this Computational Making Programme for students to develop their interest and make recommendations to others

Keywords: Computational Making, Computational Thinking, Interdisciplinary Readiness

1. Introduction

Singapore launched a “Digital Readiness Blueprint” in 2018 to ensure that every citizen is equipped to live in a Smart Nation and have access to the tools and knowledge to benefit from technology. This push towards digital technology aims to deepen its citizens’ technical capabilities, especially in key areas such as data science, artificial intelligence and cyber security. The value of computational capabilities is highlighted in this vision and to meet this vision, the government has started introducing computational making related activities in schools. Computational making is an important skill to have in today’s digital economy, especially so to meet the ever-changing global challenges and work demands (Riddle, 2015). In education, it is regarded as a key component of 21st century learning. Marenko (2015) describes computational making as an activity where computation and making come together in the digital medium through sensory participation and an understanding of materiality. Making has its origins in the works of Papert, Dewey, Piaget, and Montessori (Martinez & Stager, 2013) and its emphasis on active learning through the creation of artefacts is gaining traction in recent years (Rode, Weibert, Marshall, Aal, von Rekowski, Mimouni, & Booker, 2015). It has been suggested that making is apt for linking the digital and the physical mediums, especially in the computing field (Rosner, 2010; Buechley, Rosner, Paulos, & Williams, 2009; Mellis, Follmer, Hartmann, Buechley, & Gross, 2013). Papert, the pioneering force behind the maker movement, strongly encouraged children to use computers to invent and build various artefacts (Gershenfeld, 2007). Indeed, making should allow learners to understand the complexities and workings of technology, rather than be contented as a passive consumer of technology (Kafai et al., 2014a).

As schools start introducing computational making in the classroom, it is important to take note of some of the difficulties faced by learners. Computational making can be cognitively challenging as it is technically complex and demands both declarative and procedural knowledge (Álvarez & Larrañaga, 2016; Renumol, Jayaprakash, & Janakiram, 2009). Computational making also encompasses more than just coding – students learn skills related to problem-solving via computer science concepts like abstraction and decomposition (Lye & Koh, 2014). A review of the literature indicates that students often consider programming or computational making boring as they find the theoretical concepts and techniques too tedious and abstract (Bennedsen, Caspersen, & Kölling, 2008). Traditional didactic

ways of teaching are often critiqued for their limitations in engaging the diverse learning needs of students (Benda, Bruckman, & Guzdial, 2012; Boy, 2013). It is especially difficult for students who are hands-on, less linear learners, as a didactic style tends to emphasize an analytical teaching approach that dictates that there is a right (and only) solution to a prescribed problem (Buechley, Eisenberg, Catchen, & Crockett, 2008; Grove, Jorgenson, Brummel, Sen, & Gamble, 2011). Other findings have revealed that students' difficulties stem from poor teaching methodologies and low interactivity (Barker, McDowell, & Kalahar, 2009; Coull & Duncan, 2011). The ways lessons are designed do not promote active learning and fail to offer contextual information of how the particular functions and commands are to be used (Kim & Ko, 2017).

In reviewing on the teaching and learning of computational thinking through computational making, Lye and Koh (2014) found that an effective learning environment would be one that is contextualised and most relatable to the students. For lessons to be more intellectually stimulating, students should work on an authentic problem applicable to them and create corresponding artefacts (Jonassen, 2011; Kafai & Resnick, 1996). A study by Marshall and colleagues (2010) proposed the idea of using tangible interfaces to support teaching and learning computing. In a similar vein, Kafai and Burke (2014) suggested applying the maker culture to introduce key computing concepts. Their research in demonstrated that teaching computing using e-textiles can broaden students' participation (Kafai, Lee, Searle, Fields, Kaplan, & Lui, 2014b) Similarly, Rode and colleagues (2015) discussed the benefits of computational making and examined how it could how account for a more effective way to support teaching programming in to a diverse range of students.

In this study, we describe a study conducted around a 180-hour out-of-school programme designed to introduce students to computational making through App development. The students were to develop apps based on challenge-based learning, storytelling and design thinking. The purpose of this study is to explore the effectiveness of the programme and the students' interdisciplinary perceptions and readiness through attending the programme.

2. Learning Through Computational Making

The learning theory that is widely used in learning-by-making is constructionism. Constructionism posits that the efficacy of learning is increased when learners participate in learner-led inquiry which is driven by creativity and making (Stevenson, M., Falloon, G., Forbes, A., & Hatzigianni, M, 2018). It is as Papert (1986) expounded, that a learner experiences meaningful learning by reconstructing information through the building of an artefact rather than as a passive receiver of knowledge. In that context, learning then becomes more relevant to the learner as they gain information through "learning-by-demand" rather than the traditional "just-in-case" curriculum that dispenses a syllabus that is unrelatable to the learners, just so that it might hopefully be of use later (Gershensfeld, 2007).

As a social activity, makers share resources, utilise tools that are accessible to all, working together to achieve a common objective (Tenenber, 2018). Kafai (2016) considered the notion of participation in the social context of making as a crucial and distinct feature of computing. When learners collaborate, they learn from one another, and the exposure to an understanding of diverse perspectives leads them to solve problems in a creative manner (Kafai et al., 2014b; Kafai et al., 2014a; Lewis, 2009; Pepler, Glosson, Kafai, Fields, & Searle, 2011; Xie, Antle, & Motamedi, 2008). It takes a process of iterative design, the harnessing of a technological tool and the collaborative exchange of information for making to be meaningful (Tanenbaum, Williams, Desjardins, & Tanenbaum, 2013).

3. Method

3.1 Research Questions

Specifically, this study is guided by the following research question:

RQ: What are students' acceptances (in terms of recommendation and enjoyment) of the Computational Making Programme?

3.2 Research Instrument

To address the research questions, we design a survey instrument that is to explore the effectiveness of the Programme on students using a framework that incorporates eight inter-related constructs that we have identified from the literature (Hughes, Nzekwe, & Molyneaux, 2013; Kong, Chiu, & Lai, 2018; Lim, Hosack, & Vogt, 2012; Pierce, Stacey, & Barkatsas, 2007). Two constructs are about students' acceptances of the programme including Recommendation and Enjoyment. The other six constructs are about students' interdisciplinary readiness: Interest Development, Science & Math Readiness, Design Thinking Readiness, Problem Solving Readiness, Communication & Collaboration Readiness, and Coding Readiness. To maximize measurement reliability, students were asked to indicate their level of agreement with each item on a 7-point Likert scale (1 = Strongly disagree; 7 = Strongly agree).

3.3 Research Design and Participants

We conducted a study in an after-school computational making programme where 50 secondary school students (between the ages of 13 and 16) worked on developing an app. This programme was a collaboration between a government body, an education company and a tech company. Our research took place in a training room in the tech company. This was to introduce the students to the corporate world of technology. As part of the programme, the students got to interact with other student developers and encountered first-hand experience of developing their own apps from scratch. The programme provided students with insights and skills on commercial app development, as shown in Figure 1. They had to hone the skills needed to pitch their app and on were required to navigate the different levels of brainstorming, planning, prototyping and evaluating their final product. Beyond coding, students also spent time developing skills required to pitch the app, mainly in story-telling and marketing.

During the 180-hour programme, students met every Saturday morning for 3hours each time during the school term and again for several days in June and during the September break. This programme was part of government initiative to nurture secondary school students who demonstrated an interest in computational thinking but who could not find the support or guidance they needed at school. The students were nominated by the school as they had shown interest in learning about App Development. Students did not have any background in coding. More than 70 students applied for the programme and were asked to complete a series of questions on logic and math. Out of all the applications received, 50 students (between the ages of 13 and 16) from 12 schools were selected to be part of the inaugural programme.

We have 28 responses in the survey, with 71.4% male and 28.6 % female. They are from 12 secondary school, aged from 13 to 16.

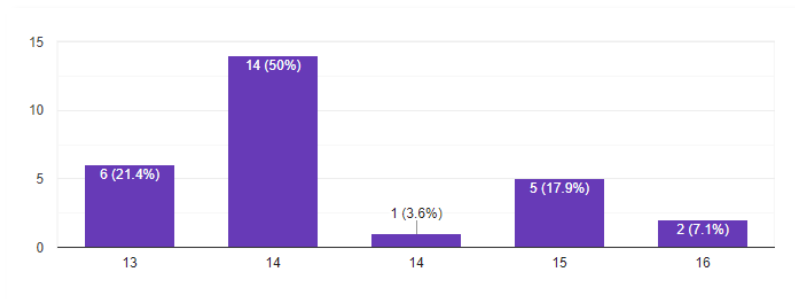


Figure 1. Age Distribution of the Participants

4. Results

4.1 What are students' acceptances (in terms of recommendation and enjoyment) of the Computational Making Programme?

Overall, all respondents would recommend this Computational Making Programme to other students, with 75% strong recommendation. All respondents enjoyed participating in the training sessions in the programme, with 67.9% strong agreement. 85.7% have started learning more about their own coding-related interests after attending the programme. Moreover, 85.7% want to do coding-related work in the future. 71.5% want to do design-related work in the future (such as design of products, apps, human-computer interfaces, etc)

Table 1
Pearson correlation analysis

	Recommendation	Enjoyment	Interest Development	Science & Math Readiness	Design Thinking Readiness	Problem Solving Readiness	Communication & Collaboration Readiness	Coding Readiness
Recommendation	1							
Enjoyment	.718**	1						
Interest Development	.776**	.749**	1					
Science & Math Readiness	.341	.396*	.475*	1				
Design Thinking Readiness	.676**	.621*	.771**	.286	1			
Problem Solving Readiness	.640**	.590**	.652**	.266	.764**	1		
Communication & Collaboration Readiness	.625**	.686**	.608**	.368	.595**	.855**	1	
Coding Readiness	.442*	.465*	.670**	.606**	.479**	.582**	.579**	1

***. Correlation is a significant at the .01 level (2-tailed, $p < .01$)*

**. Correlation is a significant at the .05 level (2-tailed, $p < .05$)*

5. Discussions

In qualitative feedback, one student responded: “Personally, it has helped me to improve my coding skills , and allowed me to now be able to code better, and find bugs more quickly e.t.c. This helps in finding errors in everyday life, and being more meticulous and careful in what I do, while teaching values such as perseverance and resilience as you feel the satisfaction after solving a bug”. Another student felt in another way: “It helped me gain more confidence in presenting things to other people and helped me cultivate more responsibility for my own projects and actions. In terms of academics, it has helped me think of solutions more effectively using methods like design thinking.” Some highlight agree that “I have improved in my math and computing classes, more others think “It has really inspired me to explore more into the field of mobile app development.”

The Pearson correlation analysis of the eight constructs of students’ perceptions of the Computational Making Programme and found they are highly correlated. With an SEM path analysis, it is found that design thinking Readiness and Coding Readiness as the two most common acknowledged factors in the programme for students to develop their interest and make recommendations to others. An ANOVA test shows that both boys and girls share similar perceptions and views on the programme.

6. Conclusion

As the literature suggested, a key strategy of that an effective learning environment of learning computational thinking through computational making would be one that is contextualised and most relatable to the students. This study, with K12 students learning in an out-of-school Computational Making Programme in Singapore, enables us to better understand students’ acceptances and perceived interdisciplinary readiness on six dimensions: on all the six dimensions: interest development, coding, design-thinking, math & science, problem solving, communication & collaboration. The quantitative data analysis indicated that most of the participating students, whether male or female, seem to have developed a high level of interest and acceptance of computational making activities. The correlational

and SEM analysis indicated that although these students, from various schools, have benefited in different dimensions. Design thinking readiness and coding readiness should be highlighted as the prominent factors derived from this Computational Making Programme for students to develop their interest and make recommendations to others.

References

- Álvarez, A., & Larrañaga, M. (2016). Experiences incorporating lego mindstorms robots in the basic programming syllabus: lessons learned. *Journal of Intelligent & Robotic Systems*, 81(1), 117-129.
- Barker, L. J., McDowell, C., & Kalahar, K. (2009). Exploring factors that influence computer science introductory course students to persist in the major. In *The Proceedings of the 40th ACM Technical Symposium on Computer Science Education - SIGCSE'09*. (pp. 153–157) doi:10.1145/1508865.1508923
- Benda, K., Bruckman, A., & Guzdial, M. (2012). When life and learning do not fit: Challenges of workload and communication in introductory computer science online. *ACM Transactions on Computing Education (TOCE)*, 12(4), 15.
- Bennedsen, J., Caspersen, M. E., & Kölling, M. (Eds.). (2008). *Reflections on the Teaching of Programming: Methods and Implementations* (Vol. 4821). Springer.
- Boy, G. A. (2013, August). From STEM to STEAM: toward a human-centred education, creativity & learning thinking. In *Proceedings of the 31st European conference on cognitive ergonomics* (p. 3). ACM.
- Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008, April). The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 423-432). ACM.
- Buechley, L., Rosner, D. K., Paulos, E., & Williams, A. (2009, April). DIY for CHI: methods, communities, and values of reuse and customization. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4823-4826). ACM.
- Gershenfeld, N. (2007). *Fab: The coming revolution on your desktop— from personal computers to personal fabrication*. New York: Basic Books, Inc..
- Grove, F., Jorgenson, N., Brummel, B., Sen, S., & Gamble, R. (2011). Adapting rewards to encourage creativity. In *Multi-Agent Systems for Education and Interactive Entertainment: Design, Use and Experience* (pp. 51-69). IGI Global.
- Hughes, R. M., Nzekwe, B., & Molyneaux, K. J. (2013). The Single Sex Debate for Girls in Science: a Comparison Between Two Informal Science Programs on Middle School Students' STEM Identity Formation. *Res Sci Educ*, 43(5), 1979–2007. doi:10.1007/s11165-012-9345-7
- Jonassen, D., Howland, J., Marra, R.M., & Crismond, D. (2008). *Meaningful learning with technology* (3rd ed.): Pearson/Merrill Prentice Hall.
- Kafai, Y. B. (2016). From computational thinking to computational participation in K-12 education. *Communications of the ACM*, 59(8), 26-27.
- Kafai, Y. B., & Burke, Q. (2014). *Connected code: Why children need to learn programming*. MIT Press.
- Kafai, Y. B., Fields, D. A., & Searle, K. A. (2014a). Electronic textiles as disruptive designs: Supporting and challenging maker activities in schools. *Harvard Educational Review*, 84(4), 532–556.
- Kafai, Y. B., Lee, E., Searle, K., Fields, D., Kaplan, E., & Lui, D. (2014b). A crafts-oriented approach to computing in high school: Introducing computational concepts, practices, and perspectives with electronic textiles. *ACM Transactions on Computing Education (TOCE)*, 14(1), 1.
- Kafai, Y., & Resnick, M. (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. In Lawrence Erlbaum
- Kim, A. S., & Ko, A. J. (2017) A pedagogical analysis of online coding tutorials. In *The proceedings of the 2017 ACM SIGCSE technical symposium on computer science education* (pp. 321– 326). ACM.
- Kong, S.-C., Chiu, M. M., & Lai, M. (2018). A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education*, 127, 178–189. doi:10.1016/j.compedu.2018.08.026
- Lewis, T., Creativity in technology education: Providing children with glimpses of their inventive potential. *International Journal of Technology and Design Education*, 2009. 19(3): p. 255-268.
- Lim, B., Hosack, B., & Vogt, P. (2012). A Framework for Measuring Student Learning Gains and Engagement in an Introductory Computing Course: A Preliminary Report of Findings. *Electronic Journal of e-Learning*, 10(4), 428-440.
- Lye, & Koh. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, 41, 51–61.
- Marenko, B. (2015). When making becomes divination: Uncertainty and contingency in computational glitch-events. *Design Studies*, 41, 110–125.

- Marshall, P., Cheng, P. C. H., & Luckin, R. (2010, January). Tangibles in the balance: a discovery learning task with physical or graphical materials. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction* (pp. 153-160). ACM.
- Mellis, D., Follmer, S., Hartmann, B., Buechley, L., & Gross, M. D. (2013, April). FAB at CHI: digital fabrication tools, design, and community. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 3307-3310). ACM.
- Papert, S. (1986). *Constructionism: A new opportunity for elementary science education*. Massachusetts Institute of Technology, Media Laboratory, Epistemology and Learning Group.
- Peppler, K., D. Glosson, Y. Kafai, D. Fields, and K. Searle. Articulating creativity in a new domain: expert insights from the field of e-textiles. in *Proceedings of the 8th ACM conference on Creativity and cognition*. 2011. ACM.
- Pierce, R., Stacey, K., & Barkatsas, A. (2007). A scale for monitoring students attitudes to learning mathematics with technology. *Computers & Education*, 48, 285–300. doi:10.1016/j.compedu.2005.01.006
- Renumul, V., Jayaprakash, S., & Janakiram, D. (2009). Classification of cognitive difficulties of students to learn computer programming. *Indian Institute of Technology, India*, 12.
- Riddle, S. (2015). The robots are coming for your job! Why digital literacy is so important for the jobs of the future. *The Conversation*, 27, 1-4.
- Rode, J. A., Weibert, A., Marshall, A., Aal, K., von Rekowski, T., El Mimouni, H., & Booker, J. (2015). From computational thinking to computational making. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. doi:10.1145/2750858.2804261
- Rosner, D. K. (2010, April). Mediated crafts: digital practices around creative handwork. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 2955-2958). ACM.
- Stevenson, M., Falloon, G., Forbes, A., & Hatzigianni, M. (2018). High Impact Learning and Teaching using 3D Design and 3D Printing in Primary School Makerspaces. *ACCE 2018*, 75.
- Tanenbaum, J.G., A.M. Williams, A. Desjardins, and K. Tanenbaum. Democratizing technology: pleasure, utility and expressiveness in DIY and maker practice. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2013*. ACM
- Tenenberg, J. (2018). Computational making. *Computing Education Research*, 9(1), 22-23.
- Xie, L., A.N. Antle, and N. Motamedi. Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. in *Proceedings of the 2nd international conference on Tangible and embedded interaction*. 2008. ACM.

Impact of the STEM Program on Information Technology College Students' Goals: Perspectives from the Philippines

Jypzie M. CATEDRILLA^{a,c}, Josephine M. DE LA CUESTA^{b,c*}, Ma. Rowena R. CAGUIAT^c

^a*Mindanao State University General Santos, South Cotabato, Philippines*

^b*Asia Pacific College, Makati City, Philippines*

^c*De La Salle University, Manila, Philippines*

*joannec@apc.edu.ph

Abstract: The implementation of the K to 12 Basic Education program in the Philippines included the addition of senior high school (SHS), grades 11 and 12, which aims to equip students with skills that would help prepare them for the workplace, higher education or entrepreneurship. The Department of Education (DepEd) states that from the 1.2 million SHS graduates as of the end of SY 2017-2018, around half are expected to proceed to college. This study aims to look at the effectiveness of the SHS program, in terms of preparing the students for higher education. Specifically, the study focused on students enrolled in an information technology education (ITE) degree program and how their SHS program contributed to readiness for college. The survey method was used to the first-year college students from both private and state universities. The results of the survey show that past performance and computer self-efficacy positively influence academic goals, which in turn positively influences the technology outcome expectations. However, the results show that technology outcome expectations do not significantly influence the achievement of academic goals.

Keywords: K to 12 education, social-cognitive career theory, past performance, self-efficacy, outcome expectations, academic goal

1. Introduction

The Philippines implemented in SY 2012-2013 the K to 12 Basic Education Program, in response to the need to address the mismatch of competencies and job requirements of industries (“DepEd Order No. 36, s. 2012.pdf,” 2012). The K to 12 program includes kindergarten and twelve years of primary (six years) and secondary (four years junior high school and two years senior high school) education. The two years of senior high school (grades 11 and 12) aim to prepare students for entry to higher education and at the same time equip them with relevant and globally competitive skills that will promote employability and entrepreneurship or *technopreneurship* (“DepEd Order No. 36, s. 2012,” 2012). With a curriculum that is aligned to the 21st Century Core Skills and College Readiness Standards, the Senior High School (SHS) offers different tracks: Academic, Technical- Vocational-Livelihood (TVL), Sports, and Arts & Design (DepED, 2014). Under the Academic track, the following strands are offered: General Academic (GA), Accountancy, Business and Management (ABM), Science, Technology, Engineering and Mathematics (STEM), and Humanities and Social Sciences (HUMSS) (DepED, 2014). The full implementation of SHS was in SY 2016-2017, and the first set of Grade 12 students that graduated at the end of SY 2017-2018 reached over 1.2 million (Mateo, 2018).

For the graduates who decide to continue to college, thereby having the opportunity to learn more and be ready for the workplace after four to five years, it is important that they have the appropriate preparation as this is one of the goals of the SHS program (“The K to 12 Basic Education Program | Official Gazette of the Republic of the Philippines,” 2012). This study aims to evaluate the students’ readiness for higher education, specifically, the SHS-STEM graduates who are currently enrolled in IT education degree programs. The SHS-STEM strand focuses on the several branches of sciences, pure and applied mathematics, engineering, and technology.

2. Research Design

This study will be conducted using a framework based on the social cognitive career theory (SCCT) (Smith, 2002) which states that past performance, computer self-efficacy, outcome expectations, and academic goal positively influences academic performance. This study will focus on the impact of past performance on academic goals only. Specifically, the focus is on academic goals which refer to the student's competence or the expectation of getting good grades and performing better compared to others (Moeller et al., 2012). In light of data privacy concerns and the availability of data, the study will exclude the evaluation of students' grades in relation to academic performance. We hypothesize that past performance predicts computer self-efficacy (H1) and outcome expectations (H3). On the other hand, computer self-efficacy has a positive influence on the academic goal (H2) and outcome expectations (H5). Lastly, we further posit that outcome expectations predict academic goal (H4).

3. Research Methodology

The participants were 162 1st year Information Technology students (65 women, 97 men) from two (2) private universities located in the capital and a state university located in the province. The survey was administered via Google Forms, where responses were received from mid-November to mid-December 2018. Participants were enrolled in a preparatory information technology course for the 1st semester of SY 2018-2019 and came from different secondary education institutions and track specializations: STEM, n = 126, 77.78%; HUMSS, n = 9, 5.56%; ABM, n = 10, 6.17%; and GA, n = 17, 10.49%.

The assessment of past performance was adapted from the 9-item Information Technology Literacy Measurement of Fraillon, et. al. (2014) to assess the students' knowledge of fundamental IT concepts. The study adapted the 30-item Computer Self-Efficacy scale (CSE) of Torkzadeh & Koufteros (1994) to assess students' self-perception of computer-related skills and knowledge. Further, the instrument for the Technology Outcomes Expectation was adapted from the Usefulness of Mathematics scale of Fennema & Sherman (1976) to assess current students' insights of the importance of information technology in regard to their future academic goals and career plans. The academic goal variable was assessed by asking the students what grade they expected to receive at the end of an ITE course. The grade distribution is: 1.0-1.9, Excellent (n = 74, 45.68%); 2.0-2.9, Good (n = 79, 48.77%); 3.0-3.5, Satisfactory (n = 9, 5.56%); 5.0, Failed; NC, No Credit; and WD, Withdraw.

The research model was tested by performing Structural Equation Modeling (SEM) for data analysis of the structural models, consistent with other studies that used PLS-SEM (Ebarido, 2018, 2017). The result of the pilot test shows *Cronbach alpha* of greater than 0.70 values and the *Average Variance Extracted (AVE)* of greater than 0.50, showing an acceptable degree of internal consistency reliability.

4. Results and Discussion

The student's past performance in SHS has a significant effect on the student's self-efficacy in using information and communications technology (ICT)-related technologies in college, which supports H1. Computer self-efficacy also influences the students to achieve their academic goals in college; therefore, H2 is supported. Also, both past performance and computer self-efficacy have a strong influence on technology outcome expectations, which supports H3 and H5. However, the technology outcome expectations have a less significant effect on the students' achievement of their academic goals. Thus, the evidence for H4 is not enough to support the hypothesis, which is consistent with the previous study (Smith, 2002). Therefore, further investigations may need to consider clarifying and refining the outcome expectations measurement.

The research paper did not consider the actual grade to refer to academic performance since the study was conducted prior to the end of the semester and grades were not available yet. Academic goals defined by students in terms of grade expectations served as the reflection and perception of their performance in school. Past studies show that students' previous performance in school mediated by self-efficacy could be a determining factor to shape academic grade goals (OECD, 2012). Thus, the higher the expectation or academic goals of the student, the more they are prepared for higher education.

A practical application of this study may include designing the curriculum and educational process in college that allows assessment of IT students' previous experience using technology and efficacy concepts. With these, educators will have a better understanding of the competency and confidence of the students and will facilitate the revision of teaching methods and policies in response to the current needs of the students. Additionally, the SHS curriculum must also be updated to provide students with an understanding of the impact of different influences on their self-development and to explore career options for learning and work.

5. Limitations and Conclusion

The newly implemented K to 12 Basic Education program in the Philippines provides opportunities for expansion of this study and other future research endeavors. First, there is a need to enrich the data collected to ensure equal representation of all the universities in the Philippines. Second, a comparative analysis is recommended between private and public schools in determining the effectiveness of the implementation of the said program. Also, future research should apply other methodologies such as qualitative interviews, content analysis or social media analytics to discover other socio-cultural factors that may also influence students on their academic perception and performance in the field of ICT.

The study contributes to the body of knowledge by using the social-cognitive career theory to explain what influences students to achieve their academic college goals. The results of this study support previous studies that students' past performance influence computer self-efficacy, encouraging them to achieve their academic goals (Smith, 2002). Therefore, it is important that educators and institutions investigate the data acquired from past performance, computer self-efficacy, and outcome expectations and incorporate it into the curriculum and ICT education to further enhance students' motivation in terms of academic achievement.

References

- Council, N., & Education, M. (2018). Fennema-Sherman Mathematics Attitudes Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males Author: Elizabeth Fennema and Julia A. Sherman Source: *Journal for Research in Mathematics Education*, 7(5), 324–326.
- DepED. (2014). Senior High School (SHS) Curriculum and Program Requirements, 1–27.
- DepEd Order No. 36, s. 2012.pdf. (2012).
- Ebardo, R. (2017). The use of activity trackers for health empowerment and commitment : the Philippine cycling perspective commitment : the Philippine cycling perspective ". In *7th International Workshop on Computer Science and Engineering WCSE 2017*, pp. 25–27.
- Ebardo, R. A. (2018). Visibility and Training in Open Source Software Adoption : A Case in Philippine Higher Education. In *8th International Workshop on Computer Science and Engineering WCSE 2018*, pp. 368–373.
- Fennema & Sherman. (1976). Mathematics Attitudes Scale: Instruments Designed to Measure Attitudes toward the Learning of Mathematics, 7(5), 324–326.
- Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for Life in a Digital Age*.
- Mateo, J. (2018). Senior high school results exceeded expectations — DepEd | Philstar.com. Retrieved December 13, 2018, from <https://www.philstar.com/headlines/2018/05/11/1814173/senior-high-school-results-exceeded-expectations-deped>
- Moeller, A. J., Theiler, J. M., & Wu, C. (2012). Goal Setting and Student Achievement: A Longitudinal Study. *The Modern Language Journal*, 96(2), 153–169.
- OECD. (2012). *Grade Expectations: How Marks and Education Policies Shape Students' Ambitions*. PISA, OECD Publishing.
- Smith, S. M. (2002). The Role of Social Cognitive Career Theory in Information Technology based Academic Performance. *Information Technology, Learning, and Performance Journal*, 20(2), 1–10.
- The K to 12 Basic Education Program | Official Gazette of the Republic of the Philippines. (2012). Retrieved December 13, 2018, from <https://www.officialgazette.gov.ph/k-12/>
- Torkzadeh, G., & Koufteros, X. (1994). Factorial validity of a computer self-efficacy scale and the impact of computer training. *Educational and Psychological Measurement*, 54(3), 813–821.
- Villarreal, B. J., Heckhausen, J., Lessard, J., Greenberger, E., & Chen, C. (2015). High-school seniors' college enrollment goals: Costs and benefits of ambitious expectations. *Journal of Adolescence*, 45, 327–340.

Exploring the Effects of Socio-Economic Status, Motivation and ICT Use on Science Achievement: Findings from PISA 2015

Sheng-hua HUANG, Yi-chao JIANG* & Morris Siu-yung JONG

Faculty of Education, The Chinese University of Hong Kong, Hong Kong S.A.R., China

*michael.jiang@link.cuhk.edu.hk

Abstract: The present study used the PISA 2015 dataset ($n = 39,297$) of six East Asian economies to investigate the relationship between students' ICT-based learning and their science proficiency level. Using Mplus 7, CFA and SEM were conducted to obtain the three major findings. Firstly, students' socio-economic status is highly related to students' science proficiency level but is weakly related to ICT-based learning. Secondly, although both learning enjoyment and instrumental motivation are weakly, positively related to ICT-based learning, the association between learning enjoyment and students' science proficiency level is much stronger than that between instrumental motivation and science proficiency level. Lastly, ICT efficacy is positively related to ICT-based learning, which in turn, is significantly and negatively related to science proficiency level.

Keywords: socio-economic status, motivation, ICT use, science achievement, PISA

1. Introduction

Students' science literacy and performance are considered closely related to 21st century skills and their future life in modern society (Millar, 2006; Tsai, 2015). With the mushrooming of internationally large-scale surveys and related research, contradictory findings have been obtained regarding how students' science achievement has been influenced. Since PISA 2015 is the most up-to-date measurement with a focus on students' science achievement, the present study targets at exploring the relationships between students' science achievement and its major influencing factors.

2. Research Hypotheses

H1a: Students' SES is positively related to their science proficiency level.

H1b: Students' SES is positively related to their ICT-based learning.

H2: Both science learning enjoyment and instrumental motivation in science learning are positively related to science proficiency level.

H3: Both science learning enjoyment and instrumental motivation in science learning are positively related to ICT-based learning.

H4a: ICT efficacy is positively related to ICT-based learning.

H4b: ICT-based learning is positively related to students' science proficiency level.

3. Method

3.1 Participants

The participants of this study thus were 15-year-old students from Japan (16.8%), Korea (14.1%), Mainland China (Beijing, Shanghai, Jiangsu and Guangdong Provinces) (24.8%), Chinese Taipei

(19.5%), Hong Kong (13.5%) and Macao (11.3%). After removing missing data, a final sample of 39,297 students was used in our study.

3.2 Measures

3.2.1 Scales

The present study adopted four scales from PISA 2015 survey, measuring science learning enjoyment, instrumental motivation for learning science, ICT efficacy and ICT-based learning. See brief information about the scales below.

Table 1

Scales involved in the present study

Scales		Likert scale	No. of items	Cronbach's α
Science learning enjoyment		4-point (1 = <i>strongly disagree</i> to '4 = <i>strongly agree</i> ')	5	0.95
Instrumental motivation for learning science		4-point (1 = <i>strongly disagree</i> to '4 = <i>strongly agree</i> ')	4	0.93
ICT efficacy		4-point (1 = <i>strongly disagree</i> to '4 = <i>strongly agree</i> ')	5	0.85
ICT-based learning		5-point (1 = <i>never or hardly ever</i> to '5 = <i>every day</i>)	12 (6 parcels)	0.91

3.2.2 ESCS

The index of economic, social and cultural status (ESCS) for each student was calculated and provided by OECD, and it is operationalized as a comprehensive measure of students' socio-economic status in the present study.

3.2.3 Science Proficiency Level

Science proficiency levels were derived from the plausible values provided in PISA dataset and taken as a measure of students' science achievement. Using cut-off scores for proficiency level provided by OECD (2015), the 10 plausible values were recoded into 10 proficiency levels.

3.3 Data Analysis

Using Mplus 7, confirmatory factor analysis (CFA) was conducted to ensure the validity of the measurement model, and structural equation modeling (SEM) was then conducted to estimate all path coefficients. SPSS 21 was also used to generate descriptive data and synthesize replicate estimates.

4. Results and Conclusion

4.1 Reliability, Descriptive Statistics, and Correlations

Table 2

Reliability, descriptive statistics, and correlations

	M	SD	1	2	3	4	5	6
1. sci-enj	2.62	.79	(.95)					
2. sci-mot	2.85	.77	.47**	(.93)				
3. ICT-eff	2.50	.65	.18**	.09**	(.85)			
4. ICT-L	1.70	.70	.20**	.16**	.31**	(.91)		

5. ESCS	-.55	.99	.01**	-.06**	.09**	.06**	-
6. spl 1	3.09	1.29	.26**	.09**	.10**	-.06**	.36**

Notes: the Cronbach's α coefficients were in parentheses; ** $p < 0.01$.

4.2 CFA Results

A four-factor model was tested using CFA. The four factors were science learning enjoyment, instrumental motivation for learning science, ICT efficacy, and ICT-based learning, respectively. When conducting CFA with the final weight, the results generally confirmed the construct validity of the measurement with an acceptable model fit ($\chi^2 = 6388.203$, $df = 164$, $p < .001$, CFI = .97, TLI = .96, RMSEA = .031, SRMR = .032).

4.3 SEM Results

The final results of SEM were obtained by synthesizing 90 replicate estimates. To be specific, for the ten models estimated with 10 science proficiency levels by the final weight, $\chi^2 = 8846.50\sim 8923.15$, $df = 200$, $p < .001$, CFI = .96, TLI = .95, RMSEA = .033, SRMR = .037~.038; for the model estimated with the first science proficiency level by the 80 replicate weight, RMSEA = .052, SRMR = .038.

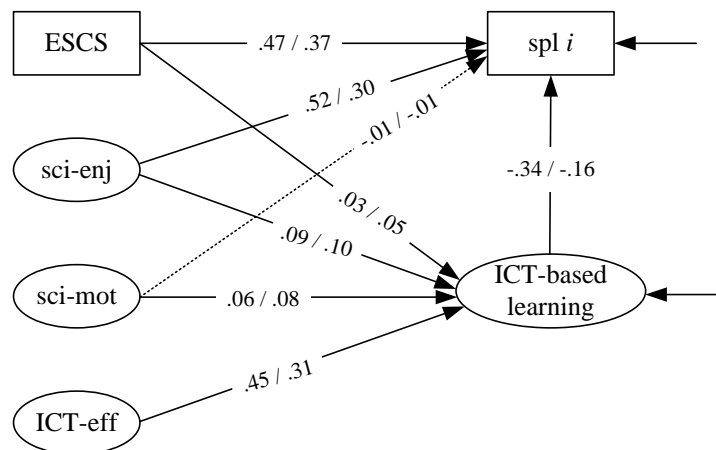


Figure 1. The SEM results of the hypothesized model.

Notes: all significant paths (significant at .001 level) were presented as solid lines, and a non-significant path presented as a dotted line; numbers before the slashes were unstandardized estimates of the path coefficients, and numbers after the slashes were standardized estimates.

As is shown in Figure 1, the findings were threefold. First, ESCS was highly related to students' science proficiency level ($\beta = .37$, $p < .001$), but was weakly, though significantly, related to ICT-based learning ($\beta = .05$, $p < .001$). Second, although both learning enjoyment and instrumental motivation were weakly and positively related to ICT-based learning (sci-enj: $\beta = .10$, $p < .001$; sci-mot: $\beta = .08$, $p < .001$), the association between learning enjoyment and students' science proficiency level ($\beta = .30$, $p < .001$) was much stronger than that between instrumental motivation and science proficiency level ($\beta = -.01$, n.s.). Third, ICT efficacy was positively related to ICT-based learning ($\beta = .31$, $p < .001$), which in turn, was significantly and negatively related to science proficiency level ($\beta = -.16$, $p < .001$). Thus, hypotheses H1, H3 and H4 were supported but hypotheses H2 was rejected due to a non-significant negative association between instrumental motivation and science proficiency level.

References

- Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, 28(13), 1499-1521.
- OECD. (2015). *Students, Computers and Learning*. OECD Publishing.

AUTHOR INDEX

A

ABADIA	RHODORA	495
ABANES	DON ROMIELITO N.	447
ABDULLAH	TIEN TIEN	429
ABERIN	MARIA ALVA Q.	141
ABIDIN	SIEW WEI	429
AGAPITO	JENILYN L.	537
AKCAPINAR	GOKHAN	300,312,318,351,360,698
AKYEA	THELMA	185
ALEJANDRO	ROMERO-HERNANDEZ	567
ANDOKO	BANNI SATRIA	706
ARAKAWA	YUTAKA	466
ASHIDA	ATSUSHI	22
ASRI	ATIQA NURUL	706
ATIENZA	FRANCIS ARLANDO L.	423

B

BAHAROM	FARIDAH LISA	429
BAI	SHURUI	515
BERGE	ELIAS	42,90
BHAGAT	KAUSHAL KUMAR	600
BHAT	JAYADEVA	402
BIYIK	ÇAĞRI ÖZKÖSE	634
BODEMER	DANIEL	121,131,158,412
BONIFACIO	KENT LEVI A.	96
BORJA	MANERO	567
BOTICKI	IVICA	312,744
BUCHHOLZ	ALEXANDER	412
BUENVIAJE	WINNA MIA VICTORIA D.	549
BULUT	IBRAHIM HAKKI	152

C

CABAUATAN	DOMINIC C.	423
CAGUIAT	MA. ROWENA	710,774
CAI	SU	685
CALVERT	JAMES	495
CASANO	JONATHAN DL.	447
CATEDRILLA	JYPZIE M.	774
CHAI	WASAN NA	336
CHAI	CHING SING	720,756
CHAN	TAK-WAI	5
CHANG	BEN	5
CHANG	CHIH-KAI	555
CHANG	MAIGA	561
CHANG	SUNG-BIN	762
CHANGHAO	LIANG	744

CHAO	GRACE CHIH-NUO	756
CHAVAN	PANKAJ	63
CHEN	WENLI	5
CHEN	BODONG	6
CHEN	GAOWEI	6
CHEN	ZHI-HONG	93
CHEN	WENLI	165
CHEN	JUN-MING	203,441
CHEN	NIAN-SHING	561
CHEN	MING-PUU	617
CHEN	CHENG-HUAN	667
CHEN	NIAN-SHING	691
CHEN	MEI-RONG ALICE	698
CHENG	KUN-HUNG	191
CHENG	QIAN-QIAN	608
CHENG	YA-WEN	691
CHEW	CHIOU SHENG	662
CHIH	WEN-CHIEH	93
CHOU	CHIH-YUEH	93
COOK	MARGARET	585
COX	DENIESE	729
CRUZ	MA. ANNIELA B. DELA	549
CUESTA	JOSEPHINE M. DE LA	774
D		
DAI	XIN-YU	555
DASIKA	RATNA	495
DEEP	ANURAG	402
DELIALIOGLU	OMER	152,634
DENDEN	MOUNA	561
DING	XUE-FEI	579
DITCHAROEN	NADH	209,591
DIY	WALFRIDO DAVID	525
DONG	YAN	644
DRAJATI	NUR ARIFAH	720
DU	A-MEI	453
E		
EBARDO	RYAN	710
ERKENS	MELANIE	158
ESSELMY	FATHI	561
F		
FADRIQUELA	INGRID MARIE THERESE P.	549
FAJARDO	ROLANDO R.	423
FELLMAN	DANIEL	90
FENG	XUANQI	505
FENG	MENG-NI	579

FLANAGAN	BRENDAN	300,312,318,351,360,698
FUNG	KA-YAN	673
FUNG	KWONG-CHIU	673
FURTADO	PEDRO GABRIEL FONTELES	101,543
G		
GABRIEL	FONTELES FURTADO PEDRO	624
GAO	MING	324
GAO	MENG-YA	608
GARCIA	ERROL MATTHEW C.	141
GONEDA	MAMERTO T.	423
GOTTIPATI	SWAPNA	232
GUAN	JUEQI	386
GUO	MEIYI	224
H		
HASEGAWA	SHINOBU	75
HASIBUAN	ZAINAL A.	487
HASNINE	NEHAL	312
HASNINE	MOHAMMAD NEHAL	318
HAYASHI	YUGO	172
HAYASHI	YUSUKE	101,273,369,543,624,706
HE	SUJIN	768
HECHANOVA	MA. REGINA M.	525
HEIMBUCH	SVEN	121,158
HERRAS	INGRID YVONNE	447
HEW	KHE FOON	515
HIRAMATSU	YUTA	354
HIRASHIMA	TSUKASA	101,273,369,543,624,706
HSIEH	JUN SCOTT CHEN	667
HSU	TING-CHIA	2
HSU	TIEN-YU	203
HSU	YU-CHING	394
HSU	TIEN-YU	441
HU	YILING	146,357
HUANG	IWEN	490
HUANG	BIYUN	515
HUANG	RONGHUI	561
HUANG	YAOHUA	685
HUANG	SHENG-HUA	777
I		
IDRIS	NORISMA	662
IHARA	DAIKI	306
INOUE	KEI	573
ITOH	YUKIHIRO	306,348,354
IWAI	KENGO	624

IYER	SRIDHAR	750
J		
JANSEN	MARC	42,90
JEMNI	MOHAMED	561
JIANG	FANGTING	377
JIANG	HUI-YU	490
JIANG	YI-CHAO	777
JONG	MORRIS SIU-YUNG	756,777
JONSSON	BERT	90
JUNUS	KASIYAH	101
K		
KAO	YU-TING	617
KASHIHARA	AKIHIRO	75
KAWASHIMA	HIROAKI	81
KHANTHUSAENG	ACHARAPORN	591
KINSHUK		561
KITAOKA	ASAKA	573
KOGURE	SATORU	306,348,354
KOH	ELIZABETH	6
KOJIMA	KAZUAKI	51
KOJIRI	TOMOKO	22,87
KONG	SIU CHEUNG	2,5,738
KONISHI	TATSUHIRO	306,348,354
KRIZANOVIC	CHRISTOPHER	178
KRÜGER	JULE M.	412
KU	OSKAR	762
KUMAR	AMAN	63
KUO	PO-CHUN	531
KUROMIYA	HIROYUKI	351
L		
LAI	SONG	69
LASALA	JOSE RAFAEL	525
LE	NGUYEN-THINH	178
LECAILLIEZ	LOUIS	698
LEE	ALWYN VWEN YEN	242
LEE	YUAN-HSUAN	394
LEE	NURUL SYAFIQAH YAP	429
LEE	CHENG EAN (CATHERINE)	484
LEE	LUNG-HAO	703
LEE	HYERAN	720
LI	HUAN	224
LI	ZHIZHUANG	282
LI	LINGYU	288
LI	HUIYONG	300
LI	BAOPING	330

LI	YONG-BO	453
LI	JING-WEI	490
LI	YIN	644
LI	JIAN-HONG	703
LIAN	JINGJING	650
LIANG	HSIN-YI	203,441
LIANG	MING-RONG	691
LIANG	JEN-KAI	762
LIAO	YI-CHENG	531
LIMPIN	LAIZA	710
LIN	HANYUNING	200
LIN	HAO-CHIANG KOONG	531
LIN	KUAN-YU	531
LIN	SHU-YUAN	617
LIN	YU-CHI	703
LINCKE	ALISA	42,90
LIU	CHU	377,386
LIU	MIAN	460
LIUA	JIAQI	69
LO	SIAW LING	252
LO	FANG-YING RIVA	667
LOH	ER FU	662
LOOI	CHEE KIT	2,5,768
LUAN	LIN	644
LUK	ERIC TSUN-HIN	756
M		
MA	LEANNE	185
MAJUMDAR	RWITAJIT	312,318,351,698,300
MANAHAN	DOMINIQUE MARIE ANTOINETTE B.	537
MANUEL	GONZALEZ-RIOJO	567
MATSUDA	YUKI	466
MCLAUGHLIN	TIMOTHY	585
MEINEL	CHRISTOPH	263
MENG	LINGLING	597
MERIEM	EL-YAMRI	567
MILRAD	MARCELO	42,90
MIRANDA	JOHN PAUL P.	423
MITRA	RITAYAN	63
MITROVIC	ANTONIJA	57
MIWA	KAZUHISA	51
MIZUTANI	KOZO	435
MOURI	KOUSUKE	478
MURTHY	SAHANA	111,402
N		
NARAYANAN	SOUMYA	111

NATARAJAN	UMA	720
NGEZE	LUCIAN VUMILIA	750
NIAN	MEI-WEN	394
NICHOLAS	ALLAN	679
NITIN	GOKARN MALLIKA	232
NIU	BINGBING	69
NOGUCHI	YASUHIRO	306,348,354
NOMURA	TOSHIHIRO	273
O		
OCHEJA	PATRICK	698
OGASAWARA	KENTO	348
OGATA	HIROAKI	288,300,312,318,351,360,478, 698,744
OGATA	YU	624
OKUBO	FUMIYA	342
OLLESCH	LISA	121
ONG	BENSON	197
OTA	KOICHI	75
OTAGAKI	TOYA	87
OUH	ENG LIEH	252
OZA	UTSAV	472
P		
PAN	JING-WEN	460
PANG	CHRIS	197
PAYNE	AUSTIN	585
PENG	XIAOLING	294,357
PILLAI	JAYESH	472
PINE	MICHELLE	585
PINKWART	NIELS	178
PRASETYA	DIDIK DWI	369
PRESTRIDGE	SARAH	729
PROMJAN	PHIRAPHONG	209
PRYOR	MATHEW R	200
PUTRA	LUTHFI K.	206
Q		
QIAN	DONG-MING	579
R		
RAHMAH	AMALIA	487
REN	YOUQUN	386
RODRIGO	MA. MERCEDES T.	12,525,537
ROSARIO	NICO B. DEL	447
ROY	RODRIGO EMMANUEL	525
RUAN	JIAHUI	357,597
RUANGRAJITPAKORN	TANETH	336

S

SADITA	LIA	101
SAKAMOTO	YOKO	366
SAKATA	NOBUHIRO	366
SAKHARDANDE	PRABODH	472
SALLY	DR.	673
SANTOSO	HARRY BUDI	101, 206,487
SARKAR	PRATITI	472
SATO	YOSHIKI	75
SCHLUSCHE	CHRISTIAN	131
SCHNAUBERT	LENKA	131,158
SEO	JINSIL HWARYOUNG	585
SHAMS	AHMED	263
SHANG	CHENCHEN	224
SHANKARARAMAN	VENKY	232
SHARMA	VANDIT	600
SHIH	JU-LING	2,5,197
SHIMADA	ATSUSHI	342
SHIMOJO	SHIGEN	172
SHIMONISHI	KEI	81
SIEW	EDWIN THENG FUNG	484
SO	HYO-JEONG	720
SONEDA	YUSUKE	466
SONGÜL	BEHICE CEYDA	634
SOTARDI	VALERIE	57
SUMI	KAORU	32
SUN	DI	324
SUPIAN	SADIAH	429
SUPNITHI	THEPCHAI	336

T

TAHIR	FAIZA	57
TAKAOKA	RYO	75
TALANDRON-FELIPE	MAY MARIE P.	12,96,194
TALUKDAR	JEEVANKUR	600
TAN	JESMINE S H	165
TAN	SENG CHEE	214
TAN	AIK LING	242
TAN	SENG CHEE	242
TAN	KAR WAY	252
TAN	DIVINE-KIA T.	525
TEO	CHEWLEE	6
TEO	TANG WEE	214
THO	NURUL AINA SYAKIRAH ZAINAL	429
TIAM-LEE	THOMAS JAMES	32
TLILI	AHMED	561
TOKUMARU	MASATAKA	22
TOMIEK	KARLA	312

TONG	FANGYUAN	597
TSENG	SHU-FEN	93
TSENG	YUEN-HSIEN	703
U		
UEKI	KOUSUKE	81
UOSAKI	NORIKO	288,478
W		
WADA	TAKUYA	87
WAN	WAI-YAN	673
WANG	GE	146
WANG	JINGYUN	342
WANG	CHUAN	453
WANG	XIAO-DONG	453
WANG	DI-FEI	460
WANG	LI-LI	608
WANG	LI-CHUN	617
WANG	SHASHA	644
WANG	DANTONG	685
WANG	YUPING	691
WANG	YUE	720
WANG	YI-QING	738
WARRIEM	JAYAKRISHNAN M.	351
WIDIYANINGTYAS	TRIYANNA	369
WONG	LUNG HSIANG	5,6
WU	FATI	69
WU	BIAN	146
WU	DI	224
WU	BIAN	294
WU	JIUN-YU	394
WU	WEN-CHI VIVIAN	662,667
WU	SHUMIN	685
WU	WUN-SYUAN	703
WU	MICHAEL	762
WU	LONGKAI	768
X		
XIE	HAORAN	617
XU	LING-YU	608
Y		
YAMADA	MASANORI	505
YAMAGUCHI	KOUDAI	288
YAMAMOTO	TOSH	197
YAMAMOTO	RAIYA	573
YAMASHITA	KOICHI	306,348,354

YAN	TENG	282
YANG	YUANYUAN	300
YANG	BO	330
YANG	CHRISTOPHER C.Y.	360
YANG	CHANG-HSUAN	394
YANG	YU-FENG (DIANA)	603
YANG	ZI-TONG	608
YANG	YANG	685
YANG	YA-FEI	691
YAP	ANTON GABRIEL	525
YASUMOTO	KEIICHI	466
YIN	CHENGJIU	288,478
YONEKAWA	TAKAHIRO	478
YUAN	PEI-YAN	453
Z		
ZHANG	JINGJING	324
ZHANG	JIANG	324
ZHANG	JIANGXIANG	377
ZHAO	XIAO-YAN	453
ZHENG	YONG	363
ZHENG	CHUN-PING	608
ZHOU	PENG	224
ZHOU	JUAN	288
ZHOU	YU-JIE	656
ZHOU	YONG	656
ZHU	ZHENGZHOU	282
ZOU	DI	617



2019.12.2-12.6, Kenting, Taiwan

27th International Conference on Computers in Education

is organized by



The Asia-Pacific Society for
Computers in Education (APSCE)



National Cheng Kung University



National Central University



National University of Tainan

is sponsored by



National Cheng Kung University

