So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Word Cloud, Pareto and Fishbone: Towards less computation-intensive data-driven project management

Chien-Sing LEE* & Fatima GACHI AHMED
School of Science and Technology, Sunway University, Malaysia.
*chiensingl@sunway.edu.my

Abstract: This study aims to explore how to enable more efficient strategizing in a less computation-intensive yet data-driven manner. The methodology would serve as a preliminary means to identify key factors prior to further data mining. With data set for Quality Education from 2013-2018 obtained from existing literature and meta-analysis literature on disruptive technologies, we use three techniques to discover key important factors leading to the evolution/development of disruptive technologies from 2013-2018. Text processing is used to generate word clouds. Subsequently, word cloud data is fed to project management tools (Pareto chart and Ishikawa diagram) to discover in greater detail, associative key influencing factors. Significance lies in the less computation-intensive yet data-driven methodology, which falls under hybrid semi-automated mining; acknowledging the contribution of human heuristics. We hope this hybrid semi-automated method would provide a preliminary means to gauge what would be interesting for further mining.

Keywords: Word Cloud, Pareto, Fishbone, text mining, analogical heuristic approach to less computation-intensive semi-automated inferences

1. Introduction

The United Nations' Development Program (UNDP) is one of the leading organizations working to fulfil the Sustainable Development Goals (SDGs) by the year 2030. SDGs have been established in 170 countries and territories focusing on poverty alleviation, democratic governance and peacebuilding, climate change, disaster risk, and economic inequality.

Corresponding with the SDGs, technology is constantly changing, becoming more pervasive and ubiquitous. Hence implementing new and existing technology in various aspects of life as well as discarding technology that are unadaptable over time is critical in order to strategize and manage socio-technological design and development more effectively.

1.1 Research aims

Friedewald and Raabe (2011) define disruptive technology as "a newly developed innovation that creates more incentive and value than the existing technology." It drastically changes day-to-day activities as well as business and economic output, creates new players and new markets while marginalizing old ones, and delivers dramatic value to stakeholders who successfully implement and adapt to the innovation. Therefore, disruptive technologies provide dramatic improvements to current product market paradigms and produces physical products and services that initiate new industries".

We will analyze trends across 2013-2015 and 2016-2018 for SDG 4 Quality Education. The aim is to identify key causes and effects and thus generate insights to enable appropriate management of technological innovation. By identifying key factors, we improve scope and cost management and are more likely to reduce risk. Consequently, organizations would be better prepared to meet the challenges of current and future trends across the industry. This applies to both academic and corporate institutions.

1.2 Research Questions

Thus far, there are many data analytics tools in the market. However, these can be quite expensive for the small and medium enterprises. The analogical approach we adopt for this study is aimed at hybrid semi-automated mining. It is popular due to its flexibility to tweak parameters at different checkpoints e.g. in multi-level data mining, to result in better outcomes. Examples of earlier studies are Lee and Singh (2004) studies on multi-level self-organizing map-principal component analysis for adaptive learning and Kiu and Lee's (2007) studies on self-organizing map-k-means and for ontology mapping and merging.

For this study, the inspiration to combine Word Cloud with Pareto chart and Fishbone diagram is from Pinto's (2018) textbook on project management in 2018. Our research questions for this study concerns the evolution of Disruptive Technologies from 2013-2015 and 2016-2018. We are interested in:

- What are the major factors driving the evolution and development of disruptive technologies for SDG4 Quality Education using text extraction, Pareto analysis and Ishikawa diagram? (This first research question is quite broad but analyses are possible as some of our references are meta-analyses).
- What insights can we gain from these tools/instruments in comparison with existing literature? Will they be the same or can the process be simplified?

2. Related work

2.1 Project management tools (Pareto chart and fishbone diagram)

The Pareto Principle is a simplified version of the Mathematics behind Pareto distribution. Pareto Principle uses 80-20 as a rule of thumb which states that for many phenomena, about 80% of the consequences are produced by 20% of the causes (Dunford, Su, Tamang, & Wintour, 2014). In other words, the "vital few" items occupy a cumulative percentage of 80% while 20% is occupied by "useful many". The total frequency results in 100%. This is often used in Management, Economics, Business, Computer Science and Human activity to enhance productivity and decision making. Pareto Analysis is an application of the Pareto Principle. Classified as a quality control, cause and effect technique, it ranks data classifications in descending order from the highest frequency of occurrences to the lowest frequency of occurrences.

The Ishikawa Diagram (Coccia, 2017) also known as Fishbone Diagram, is a technique used to identify the problem, the major factors involved, possible causes and the root cause on issues of quality. The bevel line segments in the Ishikawa Diagram represent the distribution of the multiple causes and sub-causes which produce them. The root cause is partly determined through group participation and group knowledge of the process. Such discussions help determine areas where data should be further collected.

2.2 Data analytics, text extraction tools

In this study, we look at text mining. Defined as "the process of finding useful or interesting patterns, models, directions, trends, or rules from unstructured text," text mining is a multidisciplinary field. It involves information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning, and data mining. However, text mining is more complicated due to its unstructured nature.

3. Methodology

We aim to gain insights on the evolution of disruptive technologies between 2013-2018 and to compare these with existing literature. We surmise that by comparing technology-enhanced human heuristics with actual literature, we would be able to confirm the potential of our heuristically-driven approach, serving as a rough guideline/hint of interesting areas for further mining with other models/methods.

3.1 Research design

This study is conducted by making use of simple text extraction and Project Management cause - effect tools. Data for this research is collected through the combination of research, conference and journal articles to create a literature review data set. The sources of data are from journal articles extracted from Science Direct, IEEE Explore, Lancaster University One Search. These references are presented in Table 1.

Table 1. References used in this paper.

Area	*References*
Analytics	Daniel (2014) Luckin, Holmes, Griffiths, & Forcier (2016); Murphy, Redding, & Twyman (2016); Wong, Vuong, & Liu (2017); Salloum, Al-Emran, Monem & Shaalan (2018); Viberg, Hatakka, Bälter, & Mavroudi (2018); Howell, Roberts, & Mancini (2018)
Augmented reality	Antonioli, Blake & Sparks (2016); Toledo-Morales, & Sanchez-Garcia (2018)
Virtual reality	Getso, & Bakon (2017); Hu & Lee (2017)
Mobile learning	Göksu, & Atici (2013); Domingo, & Garganté (2016); Heflin, Shewmaker, & Nguyen (2017)
Cloud computing	Pardeshi (2014)
Ubiquitous computing	Friedewald, & Raabe (2011)
Gamification	Dicheva, Dichev, Agre, & Angelova (2015); Simões, Redondo, & Vilas (2013)
Policy	Roberts-Mahoney, Means, & Garrison (2016)
Open edu resource	Wiley, Bliss, & McEwen (2014); Scanlon, McAndrew, & O'Shea (2015)

Table 2 describes the techniques used and the objectives these techniques hope to achieve.

Table 2. Techniques used and their objectives

Technique	Purpose	Objective of technique
Text Mining	Data Extraction	• Process unstructured (textual) information collected from the combination of conference and journal articles. • Clean and Extract meaningful indices from the text to use for analysis. • This process will be done using R programming.
Word Cloud	Data Visualization	• Depict key terms. • Identify high frequency key terms.
Pareto Analysis	Data Analysis	• Identify the cause-effect factors. • Narrow down to most significant causes.
Ishikawa Diagram	Data Visualization and Analysis	• Analyze broad causes by looking at specific factors. • Categorize the potential causes and effects, hence identifying factors and sub factors contributing towards disruption. • Analyze data frequency of cause and effect. • Analyze the root cause.
Word Cloud + Ishikawa Diagram + Pareto	Data Analysis	• Identify common and if there are, unique factors between the top two from the Pareto Chart and Fishbone. • Compare results with existing literature to identify whether there are insights which are comparable and insights which literature did not detect.

Text mining is carried out using R as a mining tool to process text in order to detect key terms of major relevance. Processes in Text Mining using R are:

1. Create separate text files for the year ranging from 2013-2015, 2016-2018 for “Quality Education” and load into R.
2. The text is then filtered, parsed and pre-processed using Tokenization.

3. Convert the text to lowercase, remove numbers, comma of the text document and stop words such as “a”, “of” “the”, “is”, etc.
4. Text is then transformed into a Vector space, thereby allowing us to detect most frequent words and to further generate the word cloud.

4. Findings

4.1 Generated data (2013-2015)

The generated Word Cloud is presented in Figure 1. The Word Cloud depicts the most frequent terms which are considered to be the most relevant terms. Figure 1 shows “learning” “cloud” as the most frequent term and therefore the most impactful with regard to disruption. Other high-frequency words in the Word Cloud are “information,” “students” and “data.” This implies that from 2013-2015, the driver behind technology disruptions is due to the need for information and more easily accessible management of data e.g. the cloud.

Pareto Analysis in Figure 2 based on the same dataset illustrates the frequency and the line from left to right, the cumulative percentages. The almost comparable frequency of terms for “cloud”, “mobile” indicates the trend and direction for innovation as early as 2015 in managing data/information, catering to bigger audiences with computing devices.

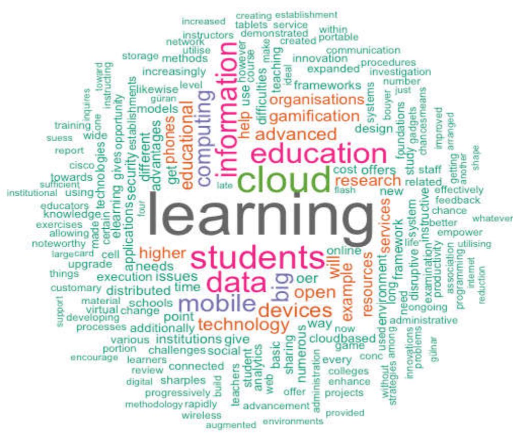

Figure 1. Word Cloud for 2013-2015 data

Answering the question “why does this happen?” we dig deeper, to better understand what is driving disruption in the Industry technology. The Ishikawa diagram in Figure 3 shows the sub-factors of “learning” as affected by education, computing, cloud, research, services, online, resources, higher, students. These reflect the concerns faced by institutions of higher learning, which need to carry out online activities such as research as well as academic services with suitable resource management. Similarly, “Cloud” affects research, education, learning, services, and information but highlight the cost factor, supporting the key concern with resource management.

As a result of combining these tools, the extension of *learning* accessibility made possible by online services, and the cloud’s higher manageability and cost-effectiveness highlight the two key main technology drivers.

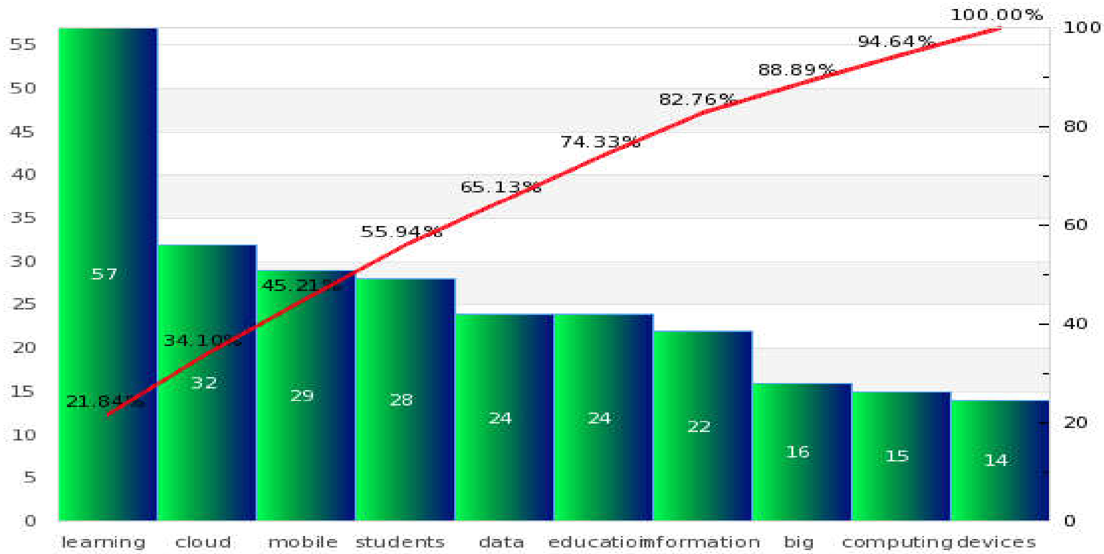

Figure 2. Pareto Analysis for 2013-2015 data

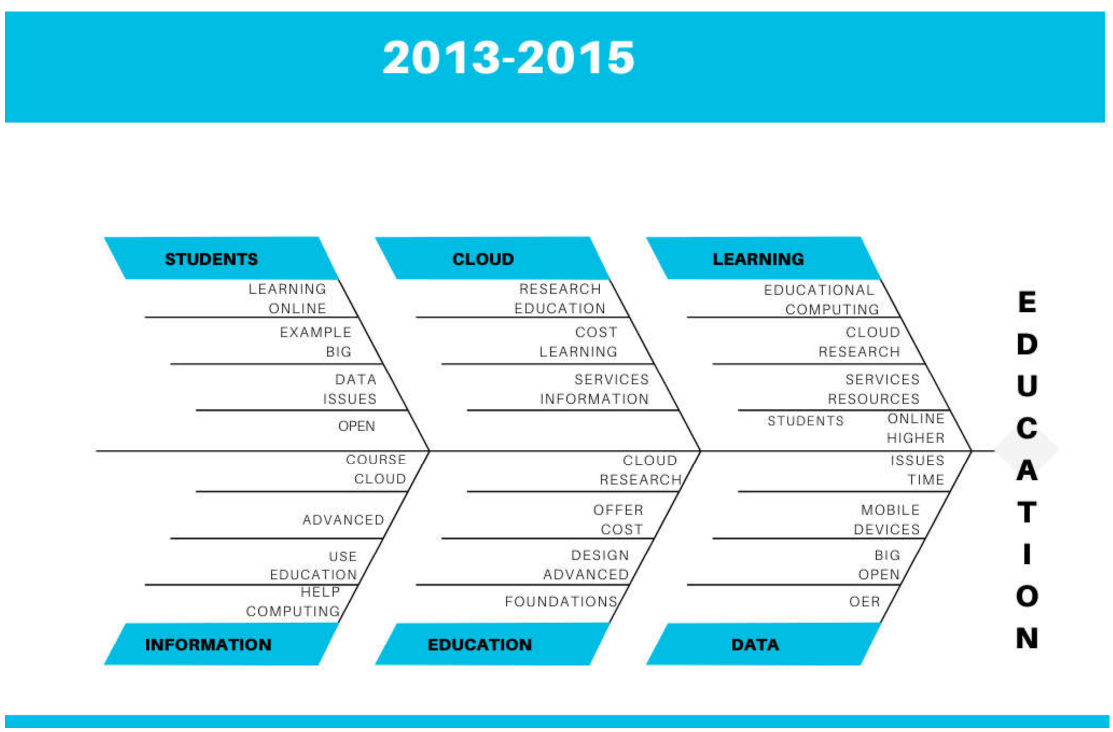

Figure 3. Ishikawa Diagram for 2013-2015 data

4.2 Generated data (2016-2018)

The categories, frequencies and cumulative relative frequencies from the Word Cloud generated from the 2016-2018 dataset is illustrated in Figure 4. Figure 5 presents the generated Word Cloud. In Figure 5, "learners," "teaching" and "learning" as the most frequent terms and therefore the most impactful with regard to disruption. The corresponding Pareto Chart is illustrated in Figure 6. There appears to be a shift to student-centered learning and "technology" as assistive.

Categories	Frequencies	Cum. Relative Frequencies (%)
learners	65	20.5
teaching	52	36.91
learning	47	51.74
technology	33	62.15
students	26	70.35
education	20	76.66
mobile	20	82.97
data	19	88.96
use	18	94.64
student	17	100
Total =	317	

Figure 4. Categories, frequencies and cumulative relative frequencies for the Word Cloud generation for the 2016-2018 data

Figure. 5. Word Cloud generated from the 2016-2018 data set

The Ishikawa diagram in Figure 7 shows “learners” is affected by learning, commitment, *model*, feedback, information, students, education, mobile, time where they impact application of disruptive technology towards education as contributing factors.

“Teaching” is affected by *enhanced*, education, *feedback*, *impacts*, school, *personalized*, *commitment*. These imply that *personalized learning* with avenues for *feedback* is gaining more interest in creating more commitment and impact to learning beyond greater accessibility and data management in the cloud (findings from the 2013-2015 data). There are also interests in using or developing “models” of teaching-learning to *enhance education*. A key factor underlying these needs is lack of *time*.

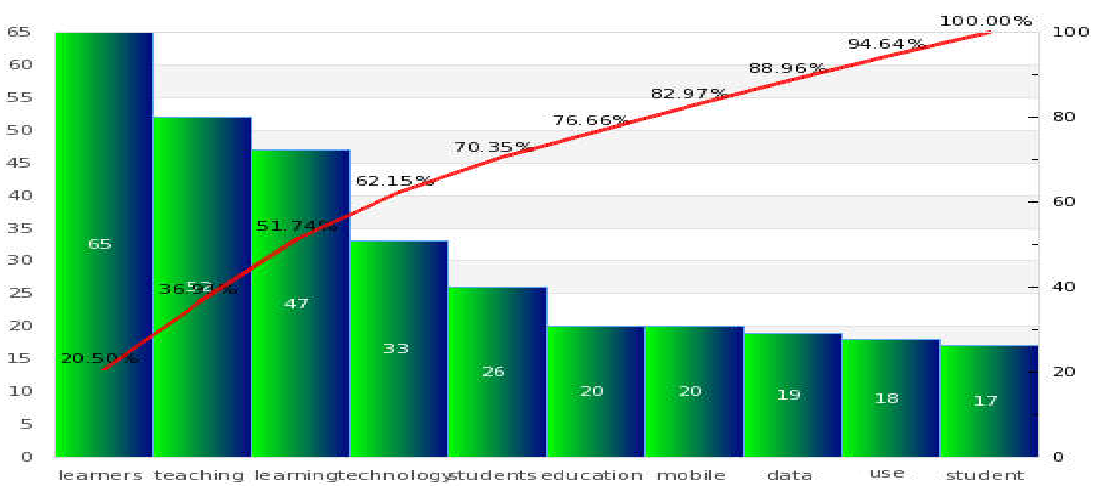

Figure 6. Pareto Analysis for the 2016-2018 data

This methodology is repeated for the healthcare domain and findings are consistent with existing literature, but faster due to its simplicity.

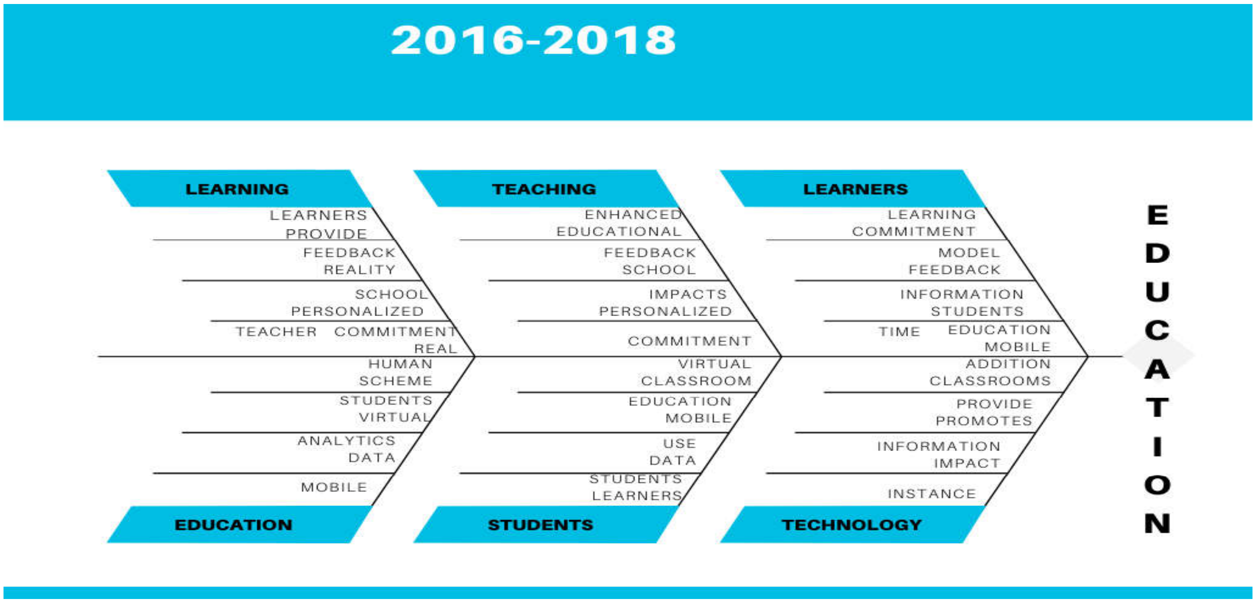

Figure 7. Ishikawa Analysis for the 2016-2018 data

4.3 Limitations of the study

This method has a certain degree of subjectivity as the data entry is partly dependent on human assessment/heuristics as its aim is only as a preliminary guide, not a full-fledged mining tool.

5. Conclusion

Problems and challenges can cause major technology product paradigm shifts or create entirely new ones. We use text extraction to feed into a Word Cloud generator and subsequently the frequency of terms from the Word Cloud become the data fed into the Pareto Chart and Ishikawa fishbone diagram. Consequently, we are able to identify major factors driving innovation for SDG4 Quality Education from 2013-2015 and, from 2016-2018. Our findings are consistent with literature and the general trends at that particular period of time but highlight associative factors in a simpler manner. Hopefully, with this less computation-intensive yet data-driven approach, with human input, it can serve as a rough guideline/hint of interesting areas for further mining with other models/methods and scoping, time, cost and risk management will become easier and faster.

Acknowledgement

We would like to thank the anonymous reviewers for their constructive feedback. This paper is extended from the second author’s capstone project completed Dec. 2018 at Sunway University, with the first author.

References

Antonioli, M., Blake, C. & Sparks, K. (2016). Augmented Reality Applications in Education. *The Journal of Technology Studies, 40*(2), 96-107.

Coccia, M. (2017). The Fishbone diagram to identify, systematize and analyze the sources of general-purpose technologies. *Journal of Social and Administrative Sciences*, *4*, 291- 303. [Accessed 15 August 2019].

Daniel, B. (2014). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904-920.

Dicheva, D., Dichev, C., Agre, G. & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study, *18*(3), 75-88.

Domingo, M. & Garganté, A. (2016). Exploring the use of educational technology in primary education: Teachers' perception of mobile technology learning impacts and applications' use in the classroom. *Computers in Human Behavior, 56*, 21-28.

Dunford, R. Su, Q., Tamang, E. & Wintour, A. (2014). The Pareto Principle. *The Plymouth Student Scientist, 7*(1), 140-148.

Friedewald, M., & Raabe, O. (2011). Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics, 28*(2), 55-65.

Getso, M. & Bakon, K. (2017). Virtual reality in education: the future of learning. *International Journal of Information Systems and Engineering, 5*(2), 30-39.

Göksu, İ. & Atici, B. (2013). Need for Mobile Learning: Technologies and Opportunities. *Procedia - Social and Behavioral Sciences, 103,* 685-694.

Heflin, H., Shewmaker, J. & Nguyen, J. (2017). Impact of mobile technology on student attitudes, engagement, and learning. *Computers & Education, 107*, 91-99.

Howell, J., Roberts, L. & Mancini, V. (2018). Learning analytics messages: Impact of grade, sender, comparative information and message style on student affect and academic resilience. *Computers in Human Behavior, 89*, 8-15.

Hu, A. E. & Lee, J. (2017). Virtual reality in education: A tool for learning in the experience age. *International Journal of Innovation in Education, 4*(4), 215.

Kiu C. C. & Lee, C. S. (2007). Learning Objects Reusability and Retrieval through Ontological Sharing: A Hybrid Unsupervised Data Mining Approach. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, Niigata, 2007, 548-550.

Lee C. S. & Singh, Y. P. (2004). Student modeling using principal component analysis of SOM clusters. *IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings.*, Joensuu, 480-484.

Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). Intelligence Unleashed. An argument for AI in Education. London: Pearson.

Murphy, M., Redding, S. & Twyman, J. (2016). *Handbook on personalized learning for states, districts, and schools.*

Pardeshi, V. (2014). Cloud Computing for Higher Education Institutes: Architecture, Strategy and Recommendations for Effective Adaptation. *Procedia Economics and Finance, 11*, 589-599.

Pinto. J. K. (2018). Project Management: Achieving Competitive Advantage (4th Edition). New York: Pearson.

Roberts-Mahoney, H., Means, A. & Garrison, M. (2016). Netflixing human capital development: personalized learning technology and the corporatization of K-12 education. *Journal of Education Policy, 31*(4), 405-420.

Salloum, S. Al-Emran, M. Monem A. & Shaalan, K. (2018). Using Text Mining Techniques for Extracting Information from Research Articles. *Studies in Computational Intelligence*, 373-397.Scanlon, E., McAndrew, P. & O'Shea, T. (2015). Designing for educational technology to enhance the experience of learners in distance education: How open educational resources, learning design and MOOCS are influencing learning. *Journal of Interactive Media in Education*, *1*(6), 1-9.

Simões, J., Redondo, R. & Vilas, A. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior, 29*(2), 345-353.

Toledo-Morales, P. & Sanchez-Garcia, J. (2018). Use of Augmented Reality in Social Sciences as Educational Resource. *Turkish Online Journal of Distance Education, 19*(3), 38-52.

United Nations Sustainable Development. (2019). About the Sustainable Development Goals. [Online]. Available: https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

Viberg, O., Hatakka, M., Bälter, O. & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110.

Wiley, D., Bliss, T. & McEwen, M. (2014). Open Educational Resources: A Review of the Literature. Handbook of Research on Educational Communications and Technology, In: Spector J., Merrill M., Elen J., Bishop M. (eds), *Handbook of Research on Educational Communications and Technology*. Springer, New York, 781-789.

Wong, P., Vuong, L. & Liu, K. (2017). Personalized learning: From neurogenetics of behaviors to designing optimal language training. *Neuropsychologia, 98*, 192-200.