Shih, J. L. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

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"

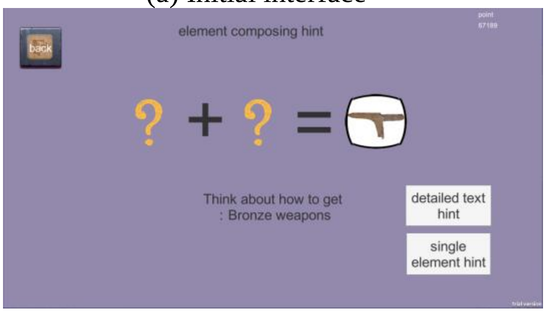
To attain our research purpose, we developed a serious game called “Hist Maker”, which is intended to support learning of historical knowledge. The game has several stages, and each stage delivers historical knowledge about one era in one country. The game can be run on a computer with Windows Operating 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.

(a) Initial interface

(b) Gameplay interface in a stage



(c) Hint system interface

(d) Post-test interface

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.

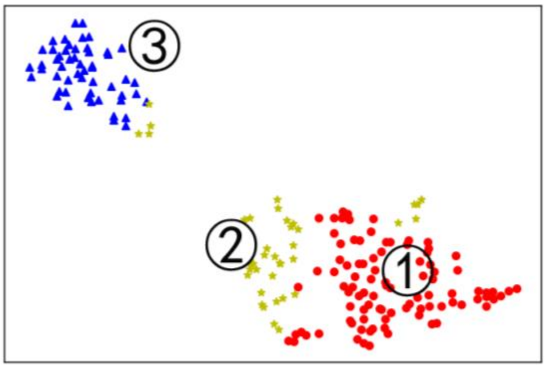


Figure 2. The result of clustering

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 *M(SD)*	Cluster 2 *M(SD)*	Cluster3 *M(SD)*	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: Conative 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.