Computers & Education 86 (2015) 18–29

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Learning through playing *Virtual Age*: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters

Meng-Tzu Cheng [a], Yu-Wen Lin [b], Hsiao-Ching She [b, *]

[a] Department of Biology, National Changhua University of Education, Jin-De Campus, No.1, Jin-De Road, Changhua 500, Taiwan, ROC
[b] Institute of Education, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 3 September 2014
Received in revised form
3 March 2015
Accepted 9 March 2015
Available online 18 March 2015

Keywords:
Secondary education
Applications in subject areas
Interactive learning environments
Teaching/learning strategies
Virtual reality

A B S T R A C T

Video games possess many unique features that facilitate learning. Meanwhile, teaching about evolution is never an easy task due to the existence of some barriers to its learning. *Virtual Age*, therefore, has been developed in an attempt to harness the power of gaming to increase student understanding of biological evolution. The aim of this study was to examine whether *Virtual Age* is effective for learning about evolution and to further explore the interplay of student concept learning, gaming performance, and in-game behaviors. A total of 62 7th graders took part in the study, and significant findings were revealed. The students did learn by playing *Virtual Age*, and their long-term knowledge retention was promising. The in-game behaviors, such as times and duration of viewing the relevant information embedded in *Virtual Age*, were significantly related to gaming performance (game score), which subsequently influenced learning outcomes. Moreover, the results of cluster analysis indicated that three clusters of low learning outcomes/low gaming performance, high learning outcomes, and high gaming performance emerged. Overall, *Virtual Age* is an effective game for learning about evolution based on its sound and sophisticated design. Implications derived from the study and suggestions for future work are proposed.

1. Introduction

In the past decade, a wide range of researchers and educators have investigated the potential of using video games in education due to the rapid growth of the gaming population among today's students (Annetta, 2008; Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Doubts about the influence of playing video games are now decreasing because more research supports the conclusion that playing video games can lead to positive consequences (Ferguson, 2007). The Serious Games Initiative, an association that promotes the use of video games for dealing with serious matters, has brought the term "serious games" (SGs) into wide use since 2002 (Djaouti, Alvarez, Jessel, & Rampnoux, 2011). Annetta (2008) further defined "serious educational games" (SEGs) as video games which are used for teaching and learning purposes in K-20 educational settings. Since then, a growing interest in investigating the effectiveness of SEGs has emerged.

Studies have now provided substantial evidence that the use of video games in education can be an effective way of increasing student knowledge acquisition, learning motivation, and classroom involvement (Annetta, Minogue, Holmes, & Cheng, 2009; Cheng & Annetta, 2012; Cheng, Su, Huang, & Chen, 2014; Papastergiou, 2009; Sung & Hwang, 2013; Tüzün, Yılmaz-Soylu, Karakuş, İnal, & Kızılkaya, 2009). However, in spite of growing support for the serious consideration of video games, researchers have suggested that more quality studies are needed to examine issues relating to the interactions between players and virtual environments because most of the currently available evidence focuses primarily on the consequences of learning by gaming (Young et al., 2012).

Evolutionary theory is considered to be an important cornerstone in modern biology (Mayr, 2001). The concepts of biological evolution include two different scales, micro and macro. Microevolution focuses on variations within a given population over generations, and

* Corresponding author. Institute of Education, National Chiao Tung University, 1001 University Rd, Hsinchu 300, Taiwan, ROC. Tel.: +886 3 5731665; fax: +886 3 5738083.
E-mail address: hcshe@mail.nctu.edu.tw (H.-C. She).

http://dx.doi.org/10.1016/j.compedu.2015.03.007
0360-1315/

macroevolution refers to major changes above the level of species over time. Natural selection is the most important mechanism by which species adapt and evolve (Reece et al., 2011). As it is an important idea within a biological model for explaining both the unity and the diversity of life on Earth, educators and scientists usually consider the theory of evolution to be a significant part of the biology curriculum. However, perhaps because the understanding of the theory is abstract and unintuitive (Bloom & Weisberg, 2007), evolution is never an easy topic for students to learn, and people often have many misconceptions (Kampourakis & Zogza, 2009).

With this in mind, our research team has developed an SEG called *Virtual Age* to assist students in their comprehension of biological evolution. The scientific concepts of evolution are realized, concretized, and gamified in *Virtual Age* in an attempt to harness the power of gaming to increase student understanding of biological evolution (Cheng, She, & Annetta, 2014). In addition, to investigate the effectiveness of using *Virtual Age* to teach students the concepts of evolution, this study sought to further explore student learning by gaming over time and the interplay of student concept learning outcomes, gaming performance, in-game behaviors, and the use of in-game characters in the virtual context of *Virtual Age* by using cluster analysis. Four research questions were addressed:

(1) Are there any differences in concept learning outcomes and gaming performance between high and low science achievers when such students play *Virtual Age* over an extended period of time?
(2) What are the correlations between student concept learning, gaming performance, and in-game behaviors?
(3) How do students use the in-game characters in *Virtual Age*?
(4) What is the interplay of student concept learning, gaming performance, and use of in-game characters?

2. Literature review

2.1. Teaching and learning of biological evolution

Scientists have been speculating about the origin and history of all living things on Earth for centuries. After more than a century of development, biological evolution, which is a unifying model that brings all the fields of biology together to explain both the unity and the diversity of life on Earth, is now regarded as a well-supported theory beyond dispute within the scientific community (Yates & Marek, 2013). Evolution includes two related phenomena, micro- and macroevolution, which together explain a broad range of processes, from the variations in a given population to the changes above the level of species over time.

Educators and scientists generally consider the theory of evolution to be a major concept which should be emphasized in the K-12 science curriculum because its explanatory and unifying powers allow for the investigation of a variety of biological questions in a scientifically meaningful way (National Science Teachers Association, 1997; Rutledge & Warden, 2000). A framework for K-12 science education proposed by the National Research Council (2012) claims that biological evolution is one of the four core ideas essential for a conceptual understanding of the life sciences enabling students to make sense of emerging research findings. The framework indicates that the science instruction for K-12 student learning of biological evolution should encompass the evidence of common ancestry and diversity (including both genetic information and fossil records), the mechanism of natural selection, the interactions between individuals and the environment (adaptation), and biodiversity and humans. Moreover, it has even been argued that not only students but also members of the public at large need to understand evolution because of its importance for understanding so many dimensions of the modern world (Smith, 2010).

The theory of evolution is central to our modern understanding of science; however, it is commonly agreed that teaching and learning about evolution is never an easy task due to the existence of various barriers to its learning for students and the general public alike. For example, biological evolution accounts for diversity over a long period of time, making it difficult to see the relevance of evolution in our daily lives. In order to understand evolution, one has to deal with concepts ranging from the micro scale (such as molecular biology) to the macro scale (such as population genetics). Moreover, a person's acceptance of evolution might be influenced by his or her worldview; for example, the idea of evolution conflicts with certain religious beliefs. The most challenging obstacle is that many misconceptions are held because an accurate understanding of evolution is anti-intuitive and clashes with individual naïve theories. These obstacles all make the learning of evolutionary concepts more difficult than learning about many other topics (Heddy & Sinatra, 2013; Rosengren, Evans, Brem, & Sinatra, 2012).

Given the aforementioned difficulties in learning about evolution, a number of instructional approaches have been proposed with solid theoretical foundations and various combinations of teaching techniques (for a review, see Smith, 2010). Researchers have indicated that it doesn't work to simply add evolutionary concepts to existing student knowledge, and that what is needed is to provide students with opportunities to approach the world and think about evolutionary processes in new and different ways than we generally do (Sinatra, Brem, & Evans, 2008). Despite the fact that many forms of instruction have been developed, however, research into using SEGs as tools for students to learn the theory of evolution remains lacking. Moreover, to our knowledge, there is still no study that has delved into the relationships between student in-game behaviors, the use of in-game characters, and the learning of evolutionary concepts through SEG play.

2.2. Serious educational games

Despite previous negative impressions, greater focus has now been placed on the positive impact of video games on learning (Gee, 2013). Video games provide players with various degrees of the subjective impression of immersion (Cheng, She, et al., 2014), and researchers argue that such immersion is the key leading to successful learning through multiple perspectives and situated experience in a digitally enhanced setting (Dede, 2009). Video games create a virtual world wherein players experience embedded activities in much the same way as when they think and act in the real world (Shaffer, Squire, Halverson, & Gee, 2005). They also continuously provide players with different challenges and tasks, and game features such as competitiveness and playfulness allow players to be engaged and immersed in the tasks so that they are willing to repeat the activities over and over (Garris, Ahlers, & Driskell, 2002). As a result, the cognitive skills of players are enhanced and learning subsequently occurs while video games are being played. The available empirical evidence has supported the fact that video game play is associated with the improvement of specific cognitive abilities and academic skills (Adachi & Willoughby, 2013; Dye & Bavelier, 2010; Spence & Feng, 2010).

The emergence of SGs and SEGs promoting the serious use of video games has significantly transformed our pedagogical perspectives. This is especially the case for science education. A context which allows students to actively explore the world, apply scientific principles, generate and test hypotheses, draw evidence-based conclusions, and predict consequences is considered necessary for best preparing scientifically literate citizens in the 21st century (Nelson, Ketelhut, Clarke, Bowman, & Dede, 2005). However, it is generally agreed that this context is difficult to demonstrate and offer and that scientific inquiry is hard to inspire through traditional lectures due to the constraints of classroom settings and the lack of experience and confidence among teachers. Fortunately, carefully designed SEGs which appropriately cover both inquiry skills and factual content successfully support scientific practices and facilitate understanding (Nelson & Ketelhut, 2007). As such, the integration of instructional content into a game format might provide students with more active and meaningful learning than traditional approaches offer (McClarty et al., 2012).

The effectiveness of SEGs in improving science learning has been examined by some empirical studies. For example, Cheng, Su, et al. (2014) developed an SEG, *Humunology*, for learning how human immunology is regulated. In terms of their understanding of procedural knowledge and higher level cognitive processes, study data indicated that students who learned by playing *Humunology* significantly outperformed their counterparts who learned by using web-based content. *River City*, a multi-user virtual environment (MUVE) that implements problem-based inquiry curricula, was created by Ketelhut, Nelson, Clarke, and Dede (2010). It was found that student understanding of biology was increased and motivation was improved after using *River City*. Moreover, student inquiry learning was enhanced and low science achievers did nearly as well as high science achievers with appropriate pedagogical strategy being employed. Wrzesien and Alcañiz Raya (2010), conducting an investigation on student learning of the basic notions of natural science and ecology through a serious virtual world game called *E-Junior*, also found higher engagement and participation, although no significant impact on knowledge itself was revealed. The findings of Wrzesien and Alcañiz Raya (2010) are in alignment with those of a study by Annetta et al. (2009) showing the same effect of using SEGs on student learning of genetics. Further, augmented reality educational games using handheld applications for supporting student-centered and intriguing opportunities have also been proposed to facilitate learning of environmental science (Klopfer & Squire, 2008).

Despite the growing support for serious consideration of video games, in a review of the literature, Young et al. (2012) suggested that more efforts are still needed to delve into the topic of SEG-based science learning. Because the majority of currently available research does not deal with the complex interactions among players, virtual environments, and even game mechanisms, the effects of using SEGs for science learning are not as clear as the effects of using SEGs for language and history learning. To assess how students act and what strategies they use while learning by gaming might give researchers more insights regarding student learning than can be derived from just examining their final scores on a game (Hsieh, Lin, & Hou, 2013). Moreover, despite the fact that some SEGs are now available for students to learn biology, e.g., *Immune Attack* (Kelly et al., 2007), *N-squad* (Klisch, Miller, Beier, & Wang, 2012), and so forth, relatively few studies have investigated how using SEGs affects student learning of biological evolution.

3. Game development and design

3.1. Learning objectives

Virtual Age is a web-based SEG created in Adobe Flash CS5 using ActionScript 3.0 with a MySQL database created using phpMyadmin for management. The game was designed based on the science curriculum standards in Taiwan. The standards suggest that K-9 science education should cover the topic of the history of life on Earth, the theory of evolution and evidence for biological evolution (Ministry of Education, 2001). Hence, in order to effectively embed the related concepts of evolution in the game format, the scientific concepts of evolution are realized, concretized, and gamified in *Virtual Age* (Cheng, She, et al., 2014). A team of two science education researchers with strong biology backgrounds and a middle school biology pre-service teacher were involved in developing the evolution game content and designing all the in-game characters, gaming rules, resource areas, and gaming environments. An animator was hired to render all the animated characters, resource areas, gaming environments, and the three animations for the opening, interlude, and ending according to the script for *Virtual Age* developed by the team of three researchers. Middle school biology teachers, college students majoring in biology, and science education researchers who are familiar with biology concepts were invited to repeatedly test the game to ensure its content validity and playability. Several learning objectives are set in advance. Namely, after playing *Virtual Age*, students are expected to be able to:

(1) Understand the birth of the Earth, the emergence and development of life on Earth, and the course of evolution.
(2) Realize the various environments of the Mesozoic and Cenozoic Eras and the morphology and characteristics of representative creatures of each era.
(3) Comprehend the mechanism of evolution.
(4) Know the relationships between creatures and the environment to which they have become adapted.

3.2. Game mechanism

The aforementioned learning objectives are embedded in the game format, with the Mesozoic Era and the Cenozoic Era as the setting for the game scenes and levels and with representative creatures from each era used as in-game characters. *Virtual Age* is a single-player game in which the player learns the mechanism of natural selection by manipulating a team of player characters (PCs) to compete with non-player characters (NPCs) controlled by the computer. Additionally, each level (era) has two more requirements for players to complete. Players have to generate an Ephippus and an Eomaia and keep them alive at the end of the Mesozoic Era, as both of them are representative species transitioning from the Mesozoic to the Cenozoic Era, and Australopithecus and Dryopithecus should be produced and kept alive at the end of Cenozoic Era because they are the ancestors of modern apes and humans.

Thirty-two in-game characters (avatars) are totally designed. These characters can be divided into six categories, namely, *Invertebrates*, *Fishes*, *Amphibians*, *Reptiles*, *Birds*, and *Mammals*. The critical and predominant species are *Reptiles* for the Mesozoic Era and *Birds* and *Mammals* for the Cenozoic Era. According to the actual ecological characteristics and niches, each avatar is created to have its own attributes. These attributes include the biomass required (i.e., the biomass required for reproducing or summoning it), health points, attack power, defense power, speed, ecological role (i.e., carnivore, herbivore, or omnivore), and habitat (i.e., the environments to which the given avatar is adapted). Each avatar also has a brief introduction describing its features and habits. The in-game characters are appropriately visualized in accordance with their real morphology in order to concretize the relevance of evolution that is generally difficult to see in our daily lives. All of the above information (that is, the character status) is provided in *Virtual Age* so that a learner can check it whenever she/he needs (Fig. 1, left).

In the game, many resource areas (which are made up of representative plants of each era) are also set up in which players have to use their herbivorous avatars to occupy and produce biomass (Fig. 2). Only when players produce enough biomass can they reproduce more offspring or summon new species to increase the character diversity of their team. Mutations resulting in different traits randomly occur during reproduction. *Virtual Age* is a turn-based game. Within a turn, the player has to manipulate each of the PCs in any order that he/she wants, and then each of the NPCs plays. The player may use a PC to move, attack, occupy, reproduce, take a rest, and so forth according to the attributes of the given character and his/her own strategies. A natural disaster system is randomly triggered at the beginning of each turn to determine which species and with what traits can survive. In order to introduce the interrelationship between creatures, the concept of symbiosis is integrated into the game mechanism. Once a symbiotic relationship has been established between the in-game character and its symbiont, the attributes of the character are enhanced. Moreover, the role that decomposers play in the environment is additionally introduced in *Virtual Age* by the design of using decomposers to decompose the corpses of characters that have died as a result of competition. Information about symbiosis and decomposers is also provided in *Virtual Age*, and similarly, a learner can check this information whenever she/he needs (Fig. 1, middle and right).

Three animations – opening, interlude, and ending – are inserted in the game. The opening animation describes the origin of the Earth and life. Moreover, it also provides an introduction to the Paleozoic Era. Between the two game levels, there is an interlude animation explaining the extinction of dinosaurs and the emergence of mammals and birds. Finally, an ending animation illustrating how modern apes and humans evolved and how modern culture began is presented after a player completes the game.

A scoring system and ranking board assessing the gaming performances of players are included in *Virtual Age* to increase the excitement and competitiveness of the game. A player's performance in the game is represented as a game score shown when the game is over and is calculated in a simple fashion according to the number and type of characters used, whether the characters are alive or dead, the degree of symbiosis achieved, and how many resource areas are occupied or not occupied by NPCs. Additionally, bonus points can be earned if continuous attacks on the same NPC are performed.

4. Materials and methods

4.1. Participants

This study was conducted in the summer of 2012. Two classes (in total, 75 7th graders) from a junior high school located in central Taiwan participated, although 13 were ultimately excluded from the analyses due to their failure to complete the whole research procedure or for having missing data. The school randomly assigned all the students into different classes using the method of heterogenous grouping, so a relatively even distribution of students with different abilities and instructional levels is assumed. The science achievement of these students was about average for the 7th grade level at their school. In order to investigate the effect of achievement level in this study, student achievement in science using a proficiency exam was collected and used as the basis for dividing students into three groups: high-achievers (the top one-third of students, score $\geqq 44$, N = 18), low-achievers (the lowest one-third of students, score $\leqq 34$, N = 23), and mid-achievers (the middle one-third of students, N = 21).

4.2. Data sources

4.2.1. Concept learning assessment

According to the learning objectives embedded in *Virtual Age*, a concept learning assessment including 24 multiple-choice questions was performed. Expert and face validity were ensured as one middle school biology teacher and two science education experts were invited to

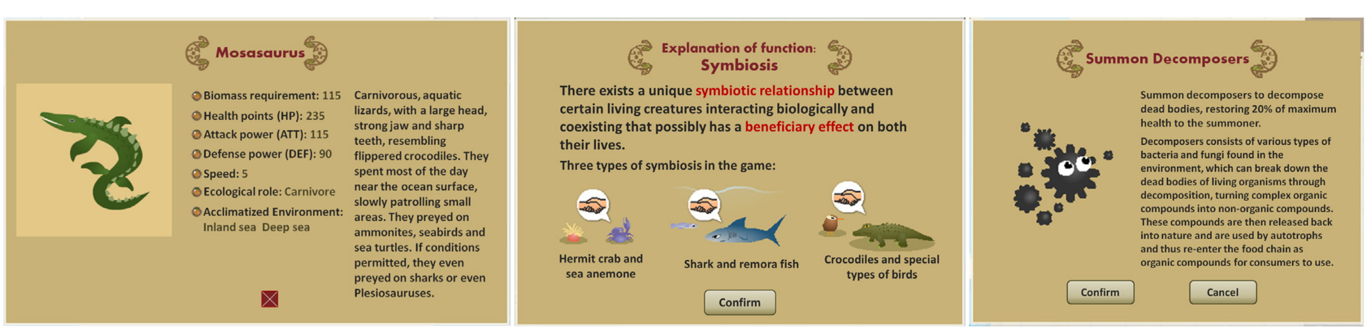

Fig. 1. Information embedded in *Virtual Age*. An example of the provided character status showing the visualization, attributes, and a brief introduction describing the features of an in-game character (left); an introduction of symbiosis (middle); and a description of decomposers (right).

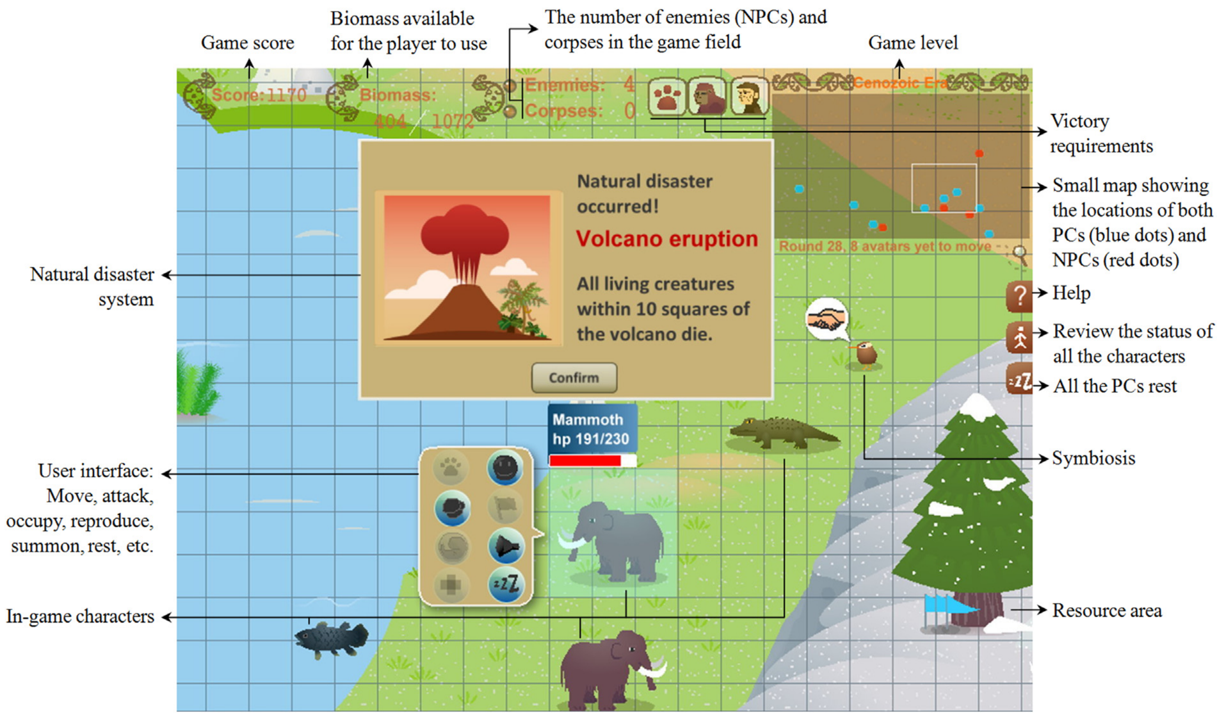

Fig. 2. A screenshot of Virtual Age. *Note*. PC: player character; NPC: non-player character.

review the assessment and to modify and reword any items which were considered inappropriate. Every question has four answer options. One is the correct answer and the other three are distracters. One point was given for each correct response to an item, which resulted in a maximum full score of 24 points. The KR_{20} of the assessment was 0.72.

4.2.2. Database recording

To further investigate how students interact with and use in-game characters in *Virtual Age*, a MySQL database was also created. Student in-game behaviors, including the in-game characters being used, the number of times of the in-game characters being reproduced, the duration and the number of times of viewing the status information, the duration of viewing the symbiosis and decomposer information, the total rounds (turns) being played, the total amount of characters being used, and student gaming performance (game score), are recorded and used for data analysis.

4.3. Experimental procedure

The participants were initially required to take the concept learning assessment as a pretest before the experiment. Then, the instructor and researcher took the participants to computer classrooms in their biology classes where the participants learned through playing *Virtual Age* on their own, with the instructor and researcher only serving as a facilitator to help solve any technical problems. All of the students in the study were asked to play *Virtual Age* for four sessions (45 min each). Two posttests were conducted to further examine the impact of playing *Virtual Age* on student concept learning over time. The first posttest (posttest 1) was conducted after the first session of play, and the second posttest (posttest 2) was taken after all four sessions were completed. Finally, a delayed test was administrated to the students one month later to evaluate the long-term effects of playing *Virtual Age* (Fig. 3).

4.4. Data analysis

4.4.1. Repeated-measures analysis of variance and regression analysis

A one-way repeated-measures analysis of variance (repeated-measures ANOVA) was run to see the difference in concept learning and gaming performance between high and low science achievers across time through playing *Virtual Age* and to examine how the in-game

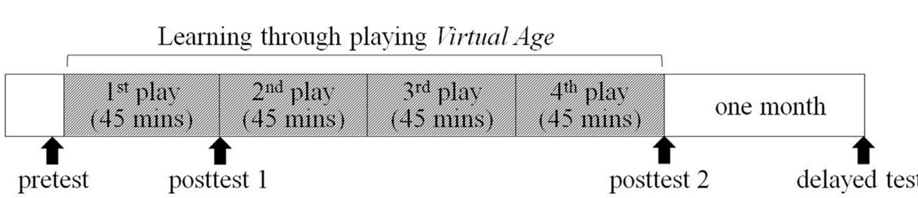

Fig. 3. An illustration of the experimental procedure of the study.

characters were used by students in *Virtual Age*. Student in-game behaviors were then retrieved from the database and their relations to gaming performance and learning outcomes were analyzed by using Pearson's correlation method.

4.4.2. Cluster analysis

Cluster analysis was employed to investigate potential patterns of student use of in-game characters, concept learning outcomes, and best game score (which refers to each student's highest game score out of all four of the games played). Parameter exploration was applied to determine the suitable parameters of the K-means algorithm in this study (Chen & Huang, 2013; De Jong, 1975), and three core clusters that produce more efficient outcomes were therefore obtained by using the K-means algorithm. One-way repeated-measures ANOVA and multivariate analysis of variance (MANOVA) were then calculated to further test the differences in concept learning outcomes and best game score across the three clusters.

5. Results

5.1. Concept learning outcomes

A one-way repeated-measures ANOVA with three achievement levels as between-subject variables and four assessment occasions of concept learning (pretest, posttest 1, posttest 2, and delayed test) as within-subject factors was calculated. Table 1 presents the descriptive statistics of student concept learning outcomes for overall performance and different achievement levels and the results of the one-way repeated-measures ANOVA. It is shown that though no effect of achievement level was found, significant differences in concept learning outcomes on different assessment occasions were revealed ($F = 4.64$, $p < .01$, $\eta^2 = 0.07$) ($\eta^2 = 0.01$ is a small effect size, $\eta^2 = 0.06$ is a medium effect size, and $\eta^2 = 0.13$ is a large effect size) (Stevens, 2009). The follow-up post-hoc tests indicate that student learning outcomes for posttest 1 were significantly better than those for the pretest ($p < .01$), and that the students also had a significantly better performance on the delayed test than on the pretest ($p < .001$). The obtained results imply that student concept learning was significantly improved over time by playing *Virtual Age*, and that the long-term effects on concept acquisition seem promising. Moreover, the learning of evolution-related concepts through playing *Virtual Age* was found to be fairly effective for all students, regardless of their achievement levels.

5.2. Gaming performance

A one-way repeated-measures ANOVA was run; three achievement levels were employed as between-subject variables, and student gaming performances (game scores) over four sessions of play (the 1st, 2nd, 3rd, and 4th plays) were used as within-subject factors, respectively. The results and descriptive statistics, including the means and standard deviations of the gaming performances for all the students and for the different levels of achievers, are depicted in Table 2. No effect of achievement level on student gaming performance was found, and a significant within-subject effect on student gaming performance ($F = 9.72$, $p < .001$, $\eta^2 = 0.18$) was approached. Through a series of post-hoc analyses, it was also found that the game scores students gained in the 2nd ($p < .01$), 3rd ($p < .001$) and 4th ($p < .001$) plays were significantly better than those that they achieved in the 1st play, and that the student gaming performances on the 4th play also outperformed those of the 2nd play ($p < .05$). The results suggest that, generally, student gaming performances became increasingly better as the number of plays increased, and that most of the students had their best performance on the 4th play.

5.3. Relationships between gaming performance, in-game behaviors, and concept learning

Pearson's correlation coefficients between student gaming performance and in-game behaviors for each play were calculated. Moreover, each student's best game score out of the four plays was particularly computed, and bivariate correlations between the best game score, in-game behaviors, and concept learning outcomes were also estimated to examine whether student performance in the game is related to their learning. The results in Table 3 indicate that, generally, the greater the number of times and the longer the duration that students spent viewing the relevant information in the game, the better the gaming performance. And especially for the 4th play, student gaming performance was significantly related to most of the in-game behaviors, such as the number of times a student viewed the status information ($r = 0.468$, $p < .01$), the duration of viewing the status information ($r = 0.282$, $p < .05$), and the duration of viewing the symbiosis information ($r = 0.415$, $p < .01$). Significant relationships were also approached between student gaming performance and the total rounds (turns) being played and the total amount of characters being used, regardless of how many times students had played *Virtual Age*. Finally, the results show that student best game score was positively and significantly related to concept learning outcomes for both the posttest 2 ($r = 0.268$,

Table 1
Results of one-way repeated-measures ANOVA showing the difference in assessment occasions and achievement levels in terms of concept learning outcomes.

	N	Pretest		Posttest 1		Posttest 2		Delayed test		*df*	MS	F	Post-hoc analyses
		M	SD	M	SD	M	SD	M	SD				
Overall	62	8.66	2.78	9.92	4.05	9.19	3.99	10.26	3.85				
High-achievers	18	9.17	2.88	10.56	5.01	9.94	4.23	11.56	4.13				
Mid-achievers	21	8.95	2.73	10.00	3.18	9.19	3.92	9.90	4.06				
Low-achievers	23	8.00	2.75	9.35	4.02	8.61	3.76	9.57	3.30				
Achievement level (low-, mid-, and high-achievers)										2	41.02	1.21	
Concept learning (pretest, posttest 1, posttest 2, delayed test)										3.00	32.80	4.64^{**}	posttest 1 > pretest (**) delayed test > pretest (***)

Note. $^{**}p < .01$, $^{***}p < .001$.

Table 2
Results of one-way repeated-measures ANOVA showing the difference in game-play occasions and achievement levels in terms of gaming performance.

	N	Gaming performance (Game scores)								df	MS	F	Post-hoc analyses
		1st Play		2nd Play		3rd Play		4th Play					
		M	SD	M	SD	M	SD	M	SD				
Overall	48	1634.79	862.09	2213.33	1274.63	2414.58	1190.67	2914.79	2012.30				
High-achievers	15	1765.33	692.22	2588.00	1467.73	2880.00	1400.81	3413.33	2338.42				
Mid-achievers	16	1579.37	761.89	2208.13	1267.10	2088.13	1007.21	3260.62	1848.30				
Low-achievers	17	1571.76	1093.80	1887.65	1066.77	2311.18	1084.18	2149.41	1712.48				
Achievement level (low-, mid-, and high-achievers)										2	7410773.96	2.21	
Gaming performance (1st, 2nd, 3rd, and 4th play)										2.37	17594548.33	9.72***	2nd > 1st (**) 3rd > 1st (***) 4th > 1st (***) 4th > 2nd (*)

Note 1. N = 48 (only participants who had game score recorded for all the four plays were analyzed).
Note 2. Gaming performance represents the original game score divided by 1000.
Note3. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

$p < .05$) and the delayed posttest (r = 0.343, $p < .01$). These findings indicate that, overall, the better the students performed in the game, the better they learned and the more they retained the scientific concepts they learned.

5.4. *Use of in-game characters in* Virtual Age

According to the log data exported from the database, we calculated the frequency of each category of characters being used (that is, the sum of times of each category of characters was used/the sum of times of the level was played) for the Mesozoic Era and the Cenozoic Era. Then, a one-way repeated-measures ANOVA with the six used categories of characters as within-subject variables was employed to analyze if any significant difference in the selection of the six categories was approached. The results demonstrate that significance in student use of different categories in both the Mesozoic Era (F = 104.34, $p < .001$, $\eta^2 = 0.63$) and Cenozoic Era (F = 68.71, $p < .001$, $\eta^2 = 0.54$) was revealed (Table 4), implying that students had a specific preference of selecting in-game characters in different game levels. Particularly, a series of post-hoc analyses show that students preferred to use reptiles and invertebrates as game characters in the Mesozoic Era and mammals and birds in the Cenozoic Era.

5.5. *Cluster analysis of in-game character use, gaming performance, and concept learning outcomes*

Three core clusters that produce more efficient outcomes were therefore obtained by using the K-means algorithm (Table 5). As a result of the analysis, the three clusters show different trends: low learning outcomes/low gaming performance (Cluster 1, N = 31), high learning outcomes (Cluster 2, N = 20), and high gaming performance (Cluster 3, N = 11). As shown in Table 5, students in Cluster 1 exhibited low concept learning outcomes across the four assessment occasions in general, their best game scores were the lowest, and the in-game characters they used as well as the frequencies of reproduction they employed in both Mesozoic and Cenozoic Era were relatively few. Learners in Cluster 2 had the best concept learning performance across the four assessment occasions compared to the other two clusters. In terms of their use of in-game characters, we can see that they knew what characters are the critical creatures in each geological age and therefore used them frequently. Hence, the characters they used the most often were reptiles in the Mesozoic Era and mammals in the Cenozoic Era. For Cluster 3, learners used the highest level of in-game characters of almost every category and employed the most frequent reproductions in both the Mesozoic Era and Cenozoic Era. The mean of their best game scores was the highest. Moreover, it is worthwhile to note that the mean performance of students in Cluster 3 on the delayed test was much more increased, although those students didn't perform really well on the pretest, posttest 1, or posttest 2.

Finally, we further examined if there were any significant differences in student concept learning and best game score across the three clusters. Table 6 shows the difference in assessment occasions for concept learning outcomes across the three clusters using repeated-measures analysis. It was found that no significant difference existed between the assessment occasions for the students in Cluster 1. However, significant differences in concept learning performances for different assessment occasions were revealed in Cluster 2 (F = 4.69,

Table 3
Pearson's correlation coefficient between in-game behaviors, gaming performance, and concept learning outcomes.

	Gaming performance (Game scores)				
	1st Play	2nd Play	3rd Play	4th Play	Best game score
Times of viewing the status info.	0.077	0.066	0.212	0.468**	0.033
Duration of viewing the status info.	−0.081	−0.071	0.154	0.282*	−0.004
Duration of viewing the symbiosis info.	0.305**	0.135	0.101	0.415**	0.270*
Duration of viewing the decomposer info.	0.325**	0.282*	0.394**	0.225	0.145
Total rounds (turns)	0.261*	0.320**	0.461**	0.677**	0.526**
Total amount of the characters used	0.263*	0.492**	0.350*	0.473**	0.385**
concept learning outcomes-posttest 2					0.268*
concept learning outcomes-delayed test					0.343**

Note 1. $^{*}p < .05$, $^{**}p < .01$.
Note 2. Best game score represents each student's highest game score of all the four plays.

Table 4
Results of one-way repeated-measures ANOVA showing the difference in student use of various categories of game characters in *Virtual Age*.

Game level	Category	M	SD	*df*	MS	F	Post-hoc analyses
Mesozoic Era (N = 62)	Invertebrates	1.35	1.11	1.89	184.62	104.34***	Invertebrates > Fishes (*)
	Fishes	1.06	0.51				Invertebrates > Amphibians (***)
	Amphibians	0.79	0.66				Invertebrates > Birds (***)
	Reptiles	3.07	1.38				Invertebrates > Mammals (***)
	Birds	0.22	0.31				Fishes > Amphibians (**)
	Mammals	0.20	0.35				Fishes > Birds (***)
							Fishes > Mammals (***)
							Amphibians > Birds (***)
							Amphibians > Mammals (***)
							Reptiles > Invertebrates (***)
							Reptiles > Fishes (***)
							Reptiles > Amphibians (***)
							Reptiles > Birds (***)
							Reptiles > Mammals (***)
Cenozoic Era (N = 60)	Invertebrates	1.33	1.53	3.18	118.14	68.71***	Invertebrates > Fishes (*)
	Fishes	0.84	0.69				Invertebrates > Amphibians (***)
	Amphibians	0.36	0.63				Invertebrates > Reptiles (***)
	Reptiles	0.63	0.67				Fishes > Amphibians (***)
	Birds	1.73	1.25				Reptiles > Amphibians (*)
	Mammals	3.44	1.70				Birds > Invertebrates (*)
							Birds > Fishes (***)
							Birds > Amphibians (***)
							Birds > Reptiles (***)
							Mammals > Invertebrates (***)
							Mammals > Fishes (***)
							Mammals > Amphibians (***)
							Mammals > Reptiles (***)
							Mammals > Birds (***)

Note 1. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.
Note 2. N = 60 (data of two students for the level of Cenozoic era was lacking because of their failure to complete the level of Mesozoic era).

$p < .01$, $\eta^2 = 0.20$) and Cluster 3 ($F = 2.90$, $p < .05$, $\eta^2 = 0.23$). Students in Cluster 2 had better mean performances on posttest 1 ($p < .001$) and the delayed test ($p < .01$) than on the pretest, and learners in Cluster 3 had the best performance on the delayed test compared to the pretest ($p < .01$) and posttest 2 ($p < .05$).

The MANOVA results showed significance [Wilks' criteria = 0.40 ($F = 6.32$, $p < .001$, $\eta^2 = 0.37$)], and the results of the ANOVA are displayed in Table 7. It is found that learners in Clusters 2 ($p < 0.05$) and 3 ($p < 0.05$) had a significantly higher level of best game score than learners in Cluster 1. And in terms of concept learning outcomes, learners in Cluster 2 generally outperformed their counterparts in Clusters 1 and 3 on the pretest, posttest 1, and posttest 2. However, the significant difference in the performance on the delayed test no longer existed between learners in Clusters 2 and 3, as learners in Cluster 2 only outperformed their counterparts in Cluster 1 on the delayed test.

Table 5
Student use of various categories of game characters, gaming performance, and concept learning outcomes of each cluster.

	Cluster 1 (N = 31) low learning outcomes/low gaming performance		Cluster 2 (N = 20) high learning outcomes		Cluster 3 (N = 11) high gaming performance	
	M	SD	M	SD	M	SD
M. Invertebrates	1.40	0.97	0.73	1.01	**2.32**	**0.98**
M. Fishes	1.12	0.40	0.87	0.57	**1.27**	**0.61**
M. Amphibians	0.87	0.62	0.31	0.40	**1.45**	**0.47**
M. Reptiles	2.95	1.38	**3.62**	**1.04**	2.41	1.66
M. Birds	**0.33**	**0.35**	0.14	0.23	0.05	0.15
M. Mammals	**0.31**	**0.43**	0.12	0.26	0.05	0.15
C. Invertebrates	0.79	0.79	1.00	1.33	**3.32**	**1.82**
C. Fishes	0.71	0.45	0.78	0.44	**1.32**	**1.23**
C. Amphibians	0.17	0.33	0.25	0.53	**1.05**	**0.93**
C. Reptiles	0.53	0.63	0.48	0.55	**1.18**	**0.72**
C. Birds	1.48	0.73	1.05	0.79	**3.64**	**1.19**
C. Mammals	2.87	1.20	**4.15**	**1.92**	3.64	2.03
M. Reproduction	0.41	0.56	0.47	0.59	**0.68**	**2.51**
C. Reproduction	1.52	0.66	2.25	1.67	**4.32**	**2.51**
Best gaming score	3135.81	1568.71	4656.50	2612.24	**4825.45**	**1961.16**
Pretest	7.84	2.41	**10.65**	**2.30**	7.36	2.80
Posttest 1	7.94	2.64	**13.80**	**3.49**	8.45	3.33
Posttest 2	7.65	2.64	**12.35**	**4.52**	7.82	2.48
Delayed test	8.52	3.01	**12.85**	**3.66**	10.45	3.05

Note 1. The highest means for each row are presented in bold.
Note 2. M. = Mesozoic Era; C. = Cenozoic Era.

Table 6
Results of one-way repeated-measures ANOVA showing the difference in assessment occasions for concept learning outcomes across the three clusters.

Variables	*df*	MS	F	Post-hoc analyses
Cluster 1				
Concept learning (pretest, posttest 1, posttest 2, delayed test)	3.00	4.36	0.72	
Cluster 2				
Concept learning (pretest, posttest 1, posttest 2, delayed test)	2.67	39.13	4.69^{**}	posttest 1 > pretest (***) delayed test > pretest (**)
Cluster 3				
Concept learning (pretest, posttest 1, posttest 2, delayed test)	2.85	21.51	2.90^{*}	delayed test > pretest (**) delayed test > posttest 2 (*)

Note. $^{*}p < .05$, $^{**}p < .01$.

6. Discussion

With advances in modern technologies, most of today's young people are introduced to virtual worlds at very young ages. They grow up playing video games, and because of they frequently interact with virtual environments via video games, their ways of learning are fairly different from those of their predecessors who were taught using traditional approaches (Prensky, 2006). Given the aforementioned phenomena, it is necessary to find ways to refocus our educational curricula around their needs. Therefore, *Virtual Age* was developed and the present study was conducted according to that precept. The findings obtained from our study are informative and thought-provoking with regard to the topic of SEG-based science learning.

In our study, we can see that students did learn from playing *Virtual Age* and they retained the knowledge they learned even a month later. Namely, our research provides evidence of a long-term effect regarding the retention of scientific knowledge acquired through using SEGs. These results are meaningful in two ways. First, the results are meaningful because studies examining the long-term effectiveness of SEGs are necessary yet still lacking (Hew & Cheung, 2010; Mikropoulos & Natsis, 2011). Moreover, most of previous studies generally investigated the short-term effect of learning through SEG play by simply using pretest-posttest designs, and none of them have explored the processes of learning by gaming over time. Empirical research into using SEGs to facilitate student learning has become an increasing trend in recent years. Despite the emerging interest in SEGs and the positive support being revealed by individual studies, however, research using meta-analyses have shown mixed results indicating that reliable conclusions about the effectiveness of SEG-based learning are not likely to be reached (Girard, Ecalle, & Magnan, 2013). This is partly due to the fact that most of the currently available studies have focused mainly on assessing the short-term effects on learning outcomes of exposure to SEGs. Therefore, in such studies, the impact on learning and retention might be underrated. SEGs support interactions of multiple sensory channels (Oblinger, 2004). Research in neuroscience and cognitive psychology has shown that the multisensory processing of information activates a larger network of brain areas than unisensory encoding invokes and benefits learning by facilitating short-term and even rather long-term memories (Shams & Seitz, 2008). It is shown that learning through the use of SEGs that activate multisensory encoding holds potential for improving knowledge retention. However, whether SEG-based learning practice is still effective and of long-term benefit when compared to learning by traditional methods that merely focus on unisensory learning is an important issue that requires further research. Second, as student understandings about the scientific concepts of biological evolution were significantly improved after playing the game and a medium effect size was produced, *Virtual Age* is an effective game for students to learn evolutionary concepts. The use of SEGs facilitating learning of some topics in biology has been supported by several studies (Annetta et al., 2009; Cheng, Annetta, Folta, & Holmes, 2011; Cheng, Su, et al., 2014; Klisch et al., 2012; Spires, Rowe, Mott, & Lester, 2011), yet SEGs developed for learning about evolution are few. Educators and scientists have been struggling over the years to find appropriate approaches for teaching evolution due to the existence of many difficulties that impede its learning. Hence, the heartening results from our study provide us with new directions and perspectives considering how SEGs which possess their own special and unique features can be put into use in future evolution education.

Improvements in student gaming performance (game score) over time were also revealed. Studies have argued that games contain a variety of features necessary to immerse players and even cause them to reach a state of flow (Csikszentmihalyi, 1990). The immersive experience encourages players to repeatedly solve the tasks in the game and mastery of skills is thus subsequently acquired (Cheng, She, et al., 2014). Hence, the students performed better and better in *Virtual Age* over time. What is interesting is that no effect of achievement level was found in this research. Low science achievers had similar learning outcomes and gaming performances as their counterparts

Table 7
Results of ANOVAs showing the differences in concept learning outcomes and best gaming score between the three clusters.

	df	MSQ	F value	Post-hoc analyses
Best gaming score	2	19464194.20	4.75^{*}	cluster 2 > cluster 1 (*) cluster 3 > cluster 1 (*)
Pretest	2	59.30	9.90^{***}	cluster 2 > cluster 1 (**) cluster 2 > cluster 3 (**)
Posttest 1	2	223.40	23.89^{***}	cluster 2 > cluster 1 (**) cluster 2 > cluster 3 (***)
Posttest 2	2	147.20	12.86^{***}	cluster 2 > cluster 1 (**) cluster 2 > cluster 3 (**)
Delayed test	2	114.45	10.00^{***}	cluster 2 > cluster 1 (**)

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

in the high achievement level. Namely, they can perform as well as high achievers through SEG-based science learning. As we know, low achievers may be easily ignored and become invisible in the classrooms because of their disengagement from lectures. This is particularly the case for science learning, which is typically considered to be more difficult and boring; students are more likely to lack confidence in learning about science than in learning about other subjects. However, SEGs built on sound learning principles not only concretize and gamify the scientifically abstract concepts, but provide learners with more engagement and personalized learning opportunities (McClarty et al., 2012). Low achievers are allowed trial and error and experience success continuously in SEGs without worrying about real-life consequences. As a result, the level of self-confidence might be increased and better performance is approached.

The most informative aspect of our study was the result of the log data being retrieved from the database to examine the interrelations between student in-game behaviors, the use of in-game characters, and concept learning resulting from playing *Virtual Age*. Researchers have appealed to studies which utilize log files to delve into the connections between players and the virtual environment and to further explore the use of avatars (Hew & Cheung, 2010); however, this kind of research remains scarce because for the most part, log data is only available for commercial games and is not accessible to researchers (Young et al., 2012). Therefore, we invested time and effort to create a database for log data recordings in *Virtual Age*. Valuable and interesting implications were obtained. The information embedded is helpful for students as they attempt to master the game, which subsequently results in better concept learning outcomes. Moreover, the design of various in-game characters visualizing the representative creatures of the Mesozoic Era and the Cenozoic Era and given different attributes successfully influenced student preferences in selecting which avatars to use. The categories that students prefer to use are exactly the critical and predominant species of each era. In other words, in addition to simply requiring students to render postplay assessment, the results further uncover the reason of how and why student learning is improved after exposure to SEGs. The overall design providing students with appropriate information and scientific concepts and different categories of avatars is sound and sophisticated, which makes *Virtual Age* effective for learning about biological evolution.

The results of cluster analysis and the follow-up tests have allowed us to understand more about student behavioral patterns and the impact of *Virtual Age* on student science learning. Learners in Cluster 1 used the least amount of in-game characters and employed the lowest frequency of reproduction. This implies that they might not have known how to play and further engage in *Virtual Age* and therefore did not learn through playing it. Learners in Cluster 2 gained the highest mean score on the pretest, showing that they had more domain knowledge than learners in the other two clusters prior to playing *Virtual Age*. They deftly used the critical characters in each geological age in the game, appropriately reproduced their in-game characters, and therefore had a pretty good average gaming performance. As a result, their concept learning gradually improved through playing *Virtual Age*, and their learning performance was the best compared to that of students in the other two clusters. Learners in Cluster 3 might be the most informative case in this study. Before using *Virtual Age*, they exhibited domain knowledge almost as low as that of learners in Cluster 1. However, they were very engaged in the game. They frequently used different categories of in-game characters and reproduced their characters as often as they could. Because of their strong participation, they had the best gaming performance and performed as well as learners in Cluster 2 on the delayed test.

These findings further tell us a more informative story behind the scenes. According to the different behavioral patterns of the three clusters, we obtained evidence of the interplays among student prior domain knowledge (that is, evolution-related knowledge), their interactions with *Virtual Age*, and their concept learning outcomes. In other words, if individuals already have more prior domain knowledge (like learners in Cluster 2), then they can master the game and have better learning outcomes. But if not, they can also learn by repeatedly playing *Virtual Age*, using different categories of avatars, and frequently reproducing their avatars (like Cluster 3). They can learn characteristics of representative creatures of each era and the relationships between creatures and the environment they have become adapted to by using different categories of avatars, and they can comprehend the mechanism of natural selection by frequently reproducing their in-game characters. The meaning is significant as the learning effect is long-term. Although some individuals have less prior domain knowledge and do not know how to play *Virtual Age* (like Cluster 1), the data from Clusters 2 and 3 shows us inspiring insights into instructional designs and game guides that can be developed to scaffold the learning of those in Cluster 1. This is what we have to do in the next step.

7. Conclusions and implications

In summary, this empirical study preliminarily uncovers important perspectives regarding student situated gaming-learning experiences which might be missed or devalued by traditional postplay judgments. The results support that the design of *Virtual Age* is effective for learning about biological evolution and further provide evidence on why the design is effective with respect to the situated interactions between learners and the game mechanism of *Virtual Age*. We believe this study brings valuable insights regarding the topic of SEG-based science learning. Some suggestions for future work are provided.

First, although long-term retention of scientific knowledge is supported in the study, four sessions of play and one-month intervals might still be too short as game play often continues for years outside of school. Thus, as also suggested in the review by Young et al. (2012), longer implementation of SEG-based instruction conducted to further examine student long-term changes in all aspects of learning is still encouraged. Second, although doing so would be a time- and effort-consuming task, it is suggested that practitioners and game designers consider the option of creating a database for recording log data when developing SEGs in the future. Such databases would allow the interrelations among players, virtual environments, and even game mechanisms to be obtained and thus could tell us more about not only the consequences but also the processes of student learning by serious gaming. Third, the analysis of data mining is an emerging trend in education in recent years because data mining can discover hidden information to better understand students and their learning environments (Baker & Yacef, 2009). Use of such analyses for SEG-based science learning is particularly warranted, as the interactions between students and the virtual worlds might be more complicated and multi-faceted than interactions in traditional classroom settings. Our study is a preliminary step, and future work with larger samples and data sets is strongly suggested. Finally, the ultimate goal of education is that no child be left behind. To help every student learn through *Virtual Age* play, more efforts still need to be invested in examining student gaming-learning interactions. Therefore, a more individualized design that provides adaptive feedback around personal needs according to the information derived from these empirical data can be created in the very near future.

Acknowledgements

This research was funded by the Ministry of Science and Technology (MOST), Taiwan, under grant contract no. 98-2511-S-009-006-MY3. All the supports are highly appreciated.

References

Adachi, P. C., & Willoughby, T. (2013). More than just fun and games: the longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades. *Journal of Youth and Adolescence, 42*(7), 1041–1052.
Annetta, L. A. (2008). *Serious educational games*. The Netherlands: Sense Publishers.
Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about Genetics. *Computers and Education, 53*(1), 74–85.
Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.
Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development, 53*(1), 86–107.
Bloom, P., & Weisberg, D. S. (2007). Childhood origins of adult resistance to science. *Science, 316*(5827), 996–997.
Cheng, M.-T., & Annetta, L. A. (2012). Students' learning outcomes and learning experiences through playing a serious educational game. *Journal of Biological Education, 46*(4), 203–213.
Cheng, M.-T., Annetta, L. A., Folta, E., & Holmes, S. Y. (2011). Drugs and the Brain: learning the impact of methamphetamine abuse on the brain through virtual brain exhibit in the museum. *International Journal of Science Education, 33*(2), 299–319.
Cheng, M.-T., She, H.-C., & Annetta, L. A. (2014). Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*. http://dx.doi.org/10.1111/jcal.12066.
Cheng, M.-T., Su, T., Huang, W.-Y., & Chen, J.-H. (2014). An educational game for learning human immunology: what do students learn and how do they perceive? *British Journal of Educational Technology, 45*(5), 820–833.
Chen, S. Y., & Huang, P.-R. (2013). The comparisons of the influences of prior knowledge on two game-based learning systems. *Computers & Education, 68*, 177–186.
Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661–686.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper-Perennial.
De Jong, K. A. (1975). *Analysis of the behavior of a class of genetic adaptive systems* (Ph.D. Dissertation). Ann Arbor, MI: University of Michigan.
Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66–69.
Djaouti, D., Alvarez, J., Jessel, J.-P., & Rampnoux, O. (2011). Origins of serious games. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 25–43). London: Springer.
Dye, M. W. G., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Research, 50*(4), 452–459.
Ferguson, C. (2007). The good, the bad and the ugly: a meta-analytic review of positive and negative effects of violent video games. *Psychiatric Quarterly, 78*(4), 309–316.
Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation and Gaming, 33*(4), 441–467.
Gee, J. P. (2013). *The anti-education era: Creating smarter students through digital learning*. New York: Palgrave Macmillan.
Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*(3), 207–219.
Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education, 97*(5), 723–744.
Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: a review of the research. *British Journal of Educational Technology, 41*(1), 33–55.
Hsieh, Y.-H., Lin, Y.-C., & Hou, H.-T. (2013). Exploring the role of flow experience, learning performance and potential behavior clusters in elementary students' game-based learning. *Interactive Learning Environments*, 1–16.
Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*(10), 1313–1340.
Kelly, H., Howell, K., Eitan, G., Holding, L., Swain, C., Burrowbridge, A., et al. (2007). How to build serious games. *Communications of the ACM, 50*(7), 44–49.
Ketelhut, D. J., Nelson, B. C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology, 41*(1), 56–68.
Klisch, Y., Miller, L. M., Beier, M. E., & Wang, S. (2012). Teaching the biological consequences of alcohol abuse through an online game: Impacts among secondary students. *CBE-Life Sciences Education, 11*(1), 94–102.
Klopfer, E., & Squire, K. (2008). Environmental detectives—the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development, 56*(2), 203–228.
Mayr, E. (2001). *What evolution is*. New York, NY: Basic Books.
McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). *A literature review of gaming in education*. New Jersey: Pearson Publishing. Retrieved from http://www.pearsonassessments.com/hai/Images/tmrs/Lit_Review_of_Gaming_in_Education.pdf.
Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769–780.
Ministry of Education. (2001). *Standards for nine-year continuous curriculum at elementary and junior high level in Taiwan*. Taipei, Taiwan: Ministry of Education, ROC.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academics Press.
National Science Teachers Association. (1997). An NSTA position statement on the teaching of evolution. *Science Scope, 21*(2), 26–27.
Nelson, B. C., & Ketelhut, D. J. (2007). Scientific inquiry in educational multi-user virtual environments. *Educational Psychology Review, 19*(3), 265–283.
Nelson, B. C., Ketelhut, D. J., Clarke, J., Bowman, C., & Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology, 45*(1), 21–27.
Oblinger, D. G. (2004). The next generatin of educational engagement. *Journal of Interactive Media in Education, 2004*(8). http://jime.open.ac.uk/jime/article/viewArticle/2004-8-oblinger/198.
Papastergiou, M. (2009). Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1–12.
Prensky, M. (2006). *Don't bother me, Mom, I'm learning! How computer and video games are preparing your kids for 21st century success and how you can help.* St. Paul: Paragon House, Cop.
Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2011). *Campbell biology* (9th ed.). San Francisco, CA: Pearson Education.
Rosengren, K. S., Evans, E. M., Brem, S., & Sinatra, G. M. (2012). *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. New York: Oxford University Press.
Rutledge, M. L., & Warden, M. A. (2000). Evolutionary theory, the nature of science & high school biology teachers: critical relationships. *The American Biology Teacher, 62*(1), 23–31.
Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104–111.
Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417.
Sinatra, G., Brem, S., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach, 1*(2), 189–195.
Smith, M. (2010). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education, 19*(6–8), 539–571.
Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology, 14*(2), 92–104.
Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning : effect of middle grade students' hypothesis testing strategies on learning outcome. *Journal of Educational Computing Research, 44*(4), 453–472.
Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences*. New York, NY: Routledge.
Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*(0), 43–51.

Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education, 52*(1), 68–77.
Wrzesien, M., & Alcañiz Raya, M. (2010). Learning in serious virtual worlds: evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55*(1), 178–187.
Yates, T., & Marek, E. (2013). Is Oklahoma really OK? A regional study of the prevalence of biological evolution-related misconceptions held by introductory biology teachers. *Evolution: Education and Outreach, 6*(1), 1–20.
Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., et al. (2012). Our princess is in another castle: a review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89.