Computers & Education 139 (2019) 1–15

Contents lists available at ScienceDirect

Computers & Education

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

Learning to be better at the game: Performance vs. completion contingent reward for game-based learning

Juneyoung Park, Seunghyun Kim, Auk Kim, Mun Y. Yi*

Korea Advanced Institute of Science and Technology (KAIST), South Korea

ARTICLE INFO

Keywords:
Game-based learning
Performance-contingent reward
Completion-contingent reward
Learning
Trying

ABSTRACT

The difficulty of designing intrinsically integrated game-based learning systems has led to alternative design strategies based on extrinsic integration. This study extends prior work on extrinsic integration design by examining the effectiveness of alternative reward structures in integrating learning and game. Specifically, a performance-contingent reward is proposed as a new integration mechanism and its effects on learning, motivation, engagement, and system perception are assessed, vis-à-vis a completion-contingent reward. A group of university students (N = 64) were involved in an empirical experiment designed to determine the effectiveness of the new reward structure in the context of English vocabulary learning and arrow-shooting gaming. The results from the experiment show that the proposed reward structure produces a statistically significant increase in the level of learning, motivation, and engagement. The results are highly encouraging for game-based learning research as the proposed approach is easily extendable, with design implications that are widely applicable.

1. Introduction

Game-based learning broadly refers to an educational system that implements game or game-elements as a motivational driver for students (Prensky, 2003; Van Eck, 2006). Games have established their status as one of the everyday activities of the current youth and have proven their extreme motivational value. Harnessing the motivational value of the game for education could provide groundbreaking leaps toward higher achievements among students (McGonigal, 2011; Tobias, Fletcher, & Wind, 2014, pp. 485–503). For the past few decades, researchers have studied and designed various types of game-based learning systems to understand the underlying effects and to identify successful designs (Gunter, Kenny, & Vick, 2006; Ke, 2016; Kim & Shute, 2015; Linehan, Kirman, Lawson, & Chan, 2011).

However, making education fun is difficult. The design of a successful game-based learning system requires the user to fully engage with the game, while also learning the educational material embedded in it. While the concept seems simple enough, the actual design can raise alarms for both the student and the teacher (Liu, Santhanam, & Webster, 2017). A student who is accustomed to commercial games is likely to find, for example, shooting down flying equations to be uninteresting. While a teacher may observe such gameplay to realize that not enough learning takes place during gameplay. Consequently, a poorly designed game-based learning can diminish both the entertainment and the educational benefit, eventually failing to appeal to both students and teachers. Therefore, it becomes highly necessary to develop game-based learning designs that work and to suggest a general design strategy that is robust and practical (Derboven, Zaman, Vissers, Geerts, & De Grooff, 2015; Ke, 2016; Meyer & Sørensen, 2009, pp. 69–82).

* Corresponding author.
E-mail addresses: j.park89@kaist.ac.kr (J. Park), ptptomr@gmail.com (S. Kim), kimauk@kaist.ac.kr (A. Kim), munyi@kaist.ac.kr (M.Y. Yi).

https://doi.org/10.1016/j.compedu.2019.04.016
Received 14 January 2019; Received in revised form 23 April 2019; Accepted 26 April 2019
Available online 04 May 2019
0360-1315/

Recent design strategies for game-based learning have been focused on the integration of learning and gaming with regards to user motivation (Brezovszky et al., 2019; Habgood & Ainsworth, 2011). Particularly, integration based on intrinsic motivation, known as intrinsic integration, has been the most popular. Intrinsic integration design combines the instructional elements with core game-mechanics in order to promote learning through highly motivating activities in the game. For example, shooting a battleship with a canon that uses fractions as the distance to teach basic mathematics would be considered an intrinsic integration (Lomas, Patel, Forlizzi, & Koedinger, 2013). When done well, intrinsic integration can provide an environment where users can learn simultaneously simply by participating in an enjoyable and engaging game activity. However, despite its popularity and potential, issues such as lack of generalizability, weak transfer of knowledge and difficulty of design have limited its progress (Santhanam, Liu, & Shen, 2016). Only a few studies have aimed to address these issues by adopting a design strategy known as extrinsic integration, where the instructional elements are attached to an external aspect of the game (Preist & Jones, 2015). Popular examples would include badges or trophies for learning often shown in gamification. Extrinsic integration design in its nature allows for a highly extendable design that can effectively deliver both learning and engagement. Developing a meaningful connection between learning and gaming, so that the two exclusive environments can encourage one another, is a vital research question for the success of extrinsic integration. However, the designs used in existing extrinsic integration designs lack the motivational nudge that can induce effort or will to perform (Huang & Hew, 2015).

Previously suggested extrinsic integration designs adapted a *'completion-contingent reward'*, in which the rewards are provided for completing a task regardless of performance or effort (Preist & Jones, 2015; Zhonggen, 2018). In this study, we aim to investigate the effects of performance-contingent reward on learning and motivational outcomes by comparing with those of completion-contingent reward. Specifically, we design and empirically evaluate an alternative reward structure based on a *'performance-contingent reward'*, which provides reward according to the performance from the activity (Deci, Koestner, & Ryan, 2001; Ryan & Deci, 2000). We believe that the consideration of performance will allow the reward to incentivize the user to put more effort into the activity, which can lead to enhancement in both learning and motivational outcomes. Also, the reward is designed specifically to be directly involved in the gameplay. As a result, the rewards would provide a diverse game-experience and higher performance for the user (Cheng, Lin, & She, 2015). We believe that our approach can contribute to the stream of research in game-based learning by providing both theoretical insights and practical implications. The highly extendable and replicable design of our approach can be deployed and updated in not only game-based learning systems but also commercial games.

For the investigation and comparison between completion-contingent reward vs. performance-contingent reward, we had the following research question:

Research Question: How performance-contingent reward in an extrinsically integrated game-based learning system affects learning and motivational outcomes.

In an effort to empirically validate our design, we developed an extrinsically integrated game-based learning system to teach English. We integrated a standard English learning system with a popular arrow-shooting game in a mobile setting using Unity. Two versions, a completion-contingent reward version and a performance-contingent reward version, were developed to directly assess their relative effects on learning and motivational outcomes. The systems were evaluated in a 3-day study involving 64 participants. The study results based on the quantitative data showed a statistically significant increase in the level of learning, motivation and some aspects of engagement for those who were assigned to the performance-contingent reward structure. We further examine qualitative data to understand the reasoning behind the difference and provide practical design implications for future studies in game-based learning.

In the following sections, we discuss the related work, system design, evaluation methods and study results. Then finally we discuss the findings from the empirical evaluation in our study and their implications with future research opportunities.

2. Related work

2.1. Extrinsic integration in game-based learning

In their work, Habgood and Ainsworth (Habgood & Ainsworth, 2011) formally identified two forms of integration design known as intrinsic integration and extrinsic integration. In their study, they directly compared the effectiveness of intrinsic integration and extrinsic integration with a control group and concluded that intrinsic integration was superior to extrinsic integration in delivering learning outcomes. While many studies accepted the strengths of intrinsic integration and extended the research to different settings (Lomas et al., 2013; Perry et al., 2014; Plass et al., 2012), the study by Preist and Jones explored the benefits that extrinsic integration could bring (Preist & Jones, 2015). While they accepted the premise that a well-designed intrinsic integration can outperform extrinsic integration, they hypothesized that extrinsic integration can be also effective when its design is improved.

The extrinsic integration design in Habgood and Ainsworth's study employs the same game from the intrinsically integrated system of the study (Habgood & Ainsworth, 2011). An action game with zombies that can be defeated using matching weapons. However, the design differs in the delivery of the learning material where in the intrinsic version they present the learning materials within the game through the game mechanics. Specifically, they displayed dividends on the enemy's chest and divisors on the weapon, by matching the two the user could slay the zombie and gain points. On the other hand, the extrinsic integration version used a normal game mechanic with matching zombies and weapons but included a mathematical quiz that needed to be completed to proceed at the end of each level. The mathematical quiz was designed to be identical to the matchings available in the intrinsic integration version except it was presented in the form of a multiple-choice quiz rather than a game mechanic. The key concept of the extrinsic version was to present the game as a reward for completing the multiple-choice quiz.

Preist and Jones (Preist & Jones, 2015) point out that the design provided by Habgood and Ainsworth (Habgood & Ainsworth, 2011) can be perceived as highly controlling to the users, as it interrupts the gameplay and must be completed to proceed. Preist and Jones considered this type of interruption from the reward perspective to be controlling and limiting the effectiveness of the system. They hypothesized that a reward design that is perceived to be less controlling would improve the game experience and therefore also the learning. Specifically, they adopted the micro-payment structure of the popular mobile game Clash of Clans and implemented a multiple quiz section that was freely accessible whenever needed and provided in-game credits which can facilitate gameplay. The design used in Preist's study differs with that of Habgood's in two aspects. First, the users were able to select to participate in the learning activity on their own will with no systematic constraints. The freedom to utilize the learning activity freely prevents it from being perceived too controlling. Second, the completion of the learning task provided a reward that contributes to the game experience. The in-game reward that facilitates gameplay provides extra incentive for the user as gaining the reward can elevate their game experience.

Another study that implemented a completion-contingent reward in an extrinsic integration setting is the work by Hwang et al. (Hwang, Hsu, Lai, & Hsueh, 2017). The study showed that implementation of game-based learning improved learning achievement and motivation with no significant change to the anxiety of the subjects. Although their work was not specified to be an extrinsic integration design or focused on the design approach of game-based learning, the implementation of the system reflects the fundamentals of an extrinsic integration design. The research aimed to evaluate the connection and the effect game-based learning had on language learning with consideration of student anxiety in the learning material. The system was divided into three phases based on the difficulty of the learning task and each phase consisted of a learning task, which provided abilities that were required to complete the gaming task. The design of integration between gaming and learning task remained to be a completion-contingent reward as the completion of the learning tasks provided the necessities to complete the game tasks. As such, there are opportunities for improvement as the participation in the game was limited when the learning tasks were not completed - there is still a possibility of a strong controlling effect (Deci et al., 2001). The procedure is more similar to that of Habgood's study rather than that of Priest and Jones (Habgood & Ainsworth, 2011; Preist & Jones, 2015).

Other than the studies mentioned above, implementing an extrinsic reward that contributes directly to gameplay has not been a common practice in game-based learning [(Infante et al., 2010; Ronimus, Kujala, Tolvanen, & Lyytinen, 2014; Zapata-Rivera, VanWinkle, Doyle, Buteux, & Bauer, 2009)?,]. Other studies that investigate the extrinsic association between game and learning were more focused on providing tokens or progress trophies. A prime example of such an approach can be found in the research area of gamification (Cheong, Cheong, & Filippou, 2013; Mekler, Brühlmann, Opwis, & Tuch, 2013; Sandberg, Maris, & Hoogendoorn, 2014; Santhanam et al., 2016; Tomaselli, Sanchez, & Brown, 2015). Gamification is a form of extrinsic integration with much less focus on the game aspect, where only game elements are used to decorate a formerly non-game activity (Deterding, Dixon, Khaled, & Nacke, 2011). Rewards are a common instrument in gamification to induce motivation or engagement, yet due to the nature of not actually having a game module included, many of the rewards in gamification are often consciously valued or accepted simply as an acknowledgment of progress.

The extant literature on extrinsic integration for game-based learning relies on a reward structure that has very limited influence on actual gameplay (Habgood & Ainsworth, 2011; Hwang et al., 2017; Preist & Jones, 2015). While the reward structure designed by Preist and Jones (Preist & Jones, 2015) provided a slightly more inclusive reward that reduced waiting time in-game, the reward structure in Habgood (Habgood & Ainsworth, 2011) and Hwang (Hwang et al., 2017) has virtually no impact on actual gameplay. Similar to how badges and trophies in gamification studies, reward structures with no impact on the game experience can limit its value to the user. In order to induce a more meaningful value to the rewards provided by the learning task, a reward should have a positive influence on the user's game experience. Through rewards that make the game experience more positive, the user's desire to obtain such rewards can motivate the users to participate more in the learning task and therefore increase the level of learning.

2.2. Rewards

Rewards are a critical part of what constitutes an entertaining game both as a motivator and as a feedback. Similarly, in the field of education, rewards have been used to incentivize students to learn and to recognize their efforts. In both areas, rewards are used as a mean to induce motivation for the activity, yet the controlling nature has made the use of rewards to be cautious (Amabile, 1993; Deci et al., 2001; Hitt, Marriott, & Esser, 1992). In this study, we utilize the psychological theories of Cognitive Evaluation Theory (CET) and Self-determination Theory (SDT) as guidelines to develop an effective reward structure. Specifically, we analyze how the rewards are given using the contingency types of tangible rewards provided in the CET. Also, how the rewards are valued by the user according to the locus of causality and regulatory style provided in the SDT.

CET is a psychological theory that explains the effect of external events such as rewards on intrinsic motivation through one's competence and autonomy. According to the CET, the effects of tangible rewards depend on the specific conditions of the task required for the reward to be given. CET provides a typology of contingencies for tangible rewards, elaborating on two contingency types, task-contingent reward and performance-contingent reward (Deci, Koestner, & Ryan, 1999).

A task-contingent reward can be further broken down to two categories of *'completion-contingent reward'* and *'engagement-contingent reward'*. A completion-contingent reward refers to the rewards given on the completion of a given task. In-game credits awarded for the completion of a task would be a completion-contingent reward (Preist & Jones, 2015). An engagement-contingent reward refers to rewards that are dependent on engaging in the activity but not necessarily completing it. For example, awarding points for continuous participation in the system would represent an engagement-contingent reward (Cooper et al., 2010).

Finally, a performance-contingent reward refers to a reward that depends on the performance of the individual in the task usually in the form of exceeding a certain criterion or matching a level of excellence. It has been observed that some gamification elements such as badges with specific performance standard can be considered a performance-contingent reward (Glover, 2013).

CET suggests that each contingency provides a different level of affirmation of competence that can offset the control from the reward. Completion contingent reward does convey some level of competence, especially if the person has a normative sense of how well they completed the task. However, the effect of a reduced sense of control is not expected to be strong. On the other hand, the recognition of excellence through a performance-contingent reward can provide substantial competence information that can negate the sense of control. In game-based learning, a performance-contingent reward that recognizes the user's effort and performance in the learning activity is more likely to lessen the controlling effect the learning activity has compared to a completion contingent reward.

Self-determination Theory (SDT) is a psychological theory of human motivation that identifies types of motivation and its development as a predictor of performance and emotional well-being. According to SDT, positive and negative effects of rewards are determined by the perceived locus of causality of control and how it aligns with extrinsic and intrinsic motivation (Ryan & Deci, 2000). On the spectrum of extrinsic and intrinsic motivation, the theory suggests that the extent to which the reward is internalized and aligned to one's values determines the motivational response. Specifically, SDT postulates six categories of styles for the regulation created from the reward, in the increasing order of internalization: non-regulation, external regulation, introjected regulation, identified regulation, integrated regulation and intrinsic regulation.

Non-regulation is a result of a complete lack of intent to act whereas intrinsic regulation is a result of fully intrinsically motivated behavior. External regulations are rooted in extrinsic motivation and are attempts to control the person's behavior through means such as rewards or punishments. Introjected regulation is a form of conformation usually involving one's desire to protect ego or self-worth, representing the very first step toward internalizing the regulation. While a controlling regulation can be negative to the person's motivation, the internalization of these regulations can induce autonomy and competence, leading to increased motivation. Among the four regulations of extrinsic motivation (i.e., excluding non-regulation and intrinsic regulation), identified regulation and integrated regulation are when the regulations become aligned with the person's values and needs. Specifically, identified regulation is when a person consciously assigns value to the regulation and integrated regulation is when the regulation itself becomes aligned with one's values.

In the context of game-based learning, rewards that users can consciously value, such as tokens or badges, are a form of identified regulation whereas rewards that can be used to enhance the game experience are a form of integrated regulation as the reward is aligned with the user's goal in enjoying the game. While both types of regulation have their strengths and weaknesses, there are educational benefits from internalizing the regulation. The internalization of the extrinsic regulation, according to Ryan and Connell (Ryan & Connell, 1989), has a significant impact on the person's autonomy, which has shown to positively influence engagement, performance and learning. The benefits from heightened autonomy and engagement could significantly alter the experience in the entire system and have impact on one's competence and performance, thereby positively contributing to learning achievement in a game-based learning system (Eseryel, Law, Ifenthaler, Ge1, & Miller, 2014; Filsecker & Hickey, 2014; Mekler et al., 2013).

3. System design

In order to empirically evaluate the effects of extrinsic integration in game-based learning, we developed a game-based learning system built upon an arrow shooting game. The concept of the game has been motivated by a popular Korean smartphone game "Everybody's game." The system consisted of two consecutive parts, first the education part and second the gaming part while a single session of gameplay is designed to start with the education part and proceed to the gaming part. The flow of the system is illustrated in detail in Fig. 1. The user may exit the session during any stage but cannot proceed to play the game without participating in the education section. In each session, 20 words are selected from the vocabulary list and they do not change throughout the session but the selection is done randomly for every new session. There were two different versions of the system corresponding to the experimental design: control group (completion-contingent reward) and treatment group (performance-contingent reward). The differences between the two versions are elaborated in the following section and briefly described in Fig. 1.

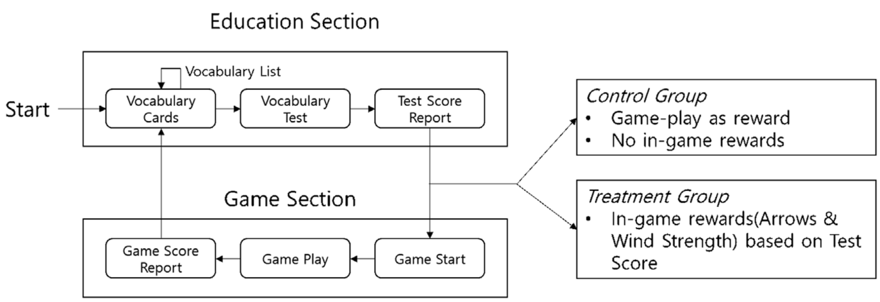

Fig. 1. System flow chart.

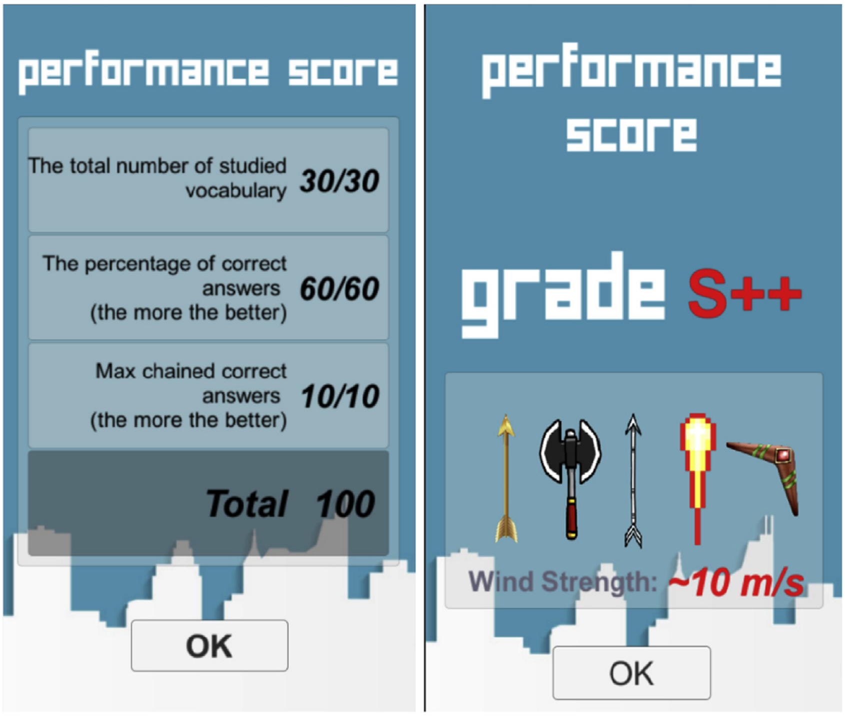

Fig. 2. Performance-contingent rewards shown for the treatment group (right). The wind strength shown indicates the maximum wind speed set for the upcoming game session. The rewards available are based on the performance score from the vocabulary test (left).

3.1. Integration design

In order to promote learning and maintain the entertainment of the game, our design of performance-contingent learning reward utilizes a unique mixture of intrinsic motivation and extrinsic motivation through the integration design. The user's intrinsic motivation to be better at the game induces a strong need for the rewards from the learning section, as they can be used to compliment the user's gameplay, inducing an extrinsic motivation through the reward and driving the user to perform better in the learning section. The link built by the two motivations creates an environment where the intrinsic motivation from gaming can continuously encourage the user to study harder.

Specifically, the two groups are presented with two different reward structures with different contingencies. Similar to that in Habgood and Ainsworth (Habgood & Ainsworth, 2011), the control group was rewarded with a completion-contingent reward, where the users proceeded to play the game. In the game, users were provided with 3 normal arrows as a reward to their completion of the learning activity. The wind strength for the control group was randomly assigned from 0 to 10(m/s).

In the treatment group, a performance-contingent reward was given according to their score in the quiz. As shown in Fig. 2 (left), the user's quiz score was evaluated in three categories to add up to 100. The total performance score was used to determine a letter grade (Fig. 2, right). According to their letter grade, the treatment group was presented with different arrows and controlled wind strength as a reward to their performance (Fig. 2, right). These arrows all differ in their types and their abilities but they were all designed and tested in a preliminary study so that they allow the user to gain a higher score than the normal arrow.

In total, there were 6 different arrows available. The specific criteria for the letter grade are shown in Table 1. These criteria were developed in a pilot study with 5 subjects who were not involved in the development of this study. In the pilot study, subjects were to determine if different letter grades provided noticeable differences in the gameplay and the game score. Specific details of available arrows and the effect of wind strength will be described in the following game description section. The golden arrow was specifically

Table 1
Letter grade criteria.

Test Score	Letter Grade	Wind Strength (m/s)	Number of Arrows	Golden Arrow
90~100	S	0	5	Available
80~90	A+	~3	4	Not Available
70~80	A	~5	3	Not Available
60~70	B+	~8	2	Not Available
~60	B	~10	1	Not Available

Fig. 3. Vocabulary card for the educational module, description (left) and example (right).

designed to represent an exceptional reward as it was determined to be the most accurate arrow available.

3.2. Education section

The education section that precedes the gaming section is largely set in two parts. The first part allows the user to freely navigate through a set of cards to study the given vocabulary, as shown in Fig. 3. The cards include details about the word including its meaning, word class, pronunciation and a sample phrase with the word. The user also has the option to view a list of these words in a long scrollable vertical table with the word and the word's meaning. The user can then move on to the quiz section where they need to solve 10 questions regarding the vocabulary they have studied in the study section, as shown in Fig. 4.

3.3. Learning content

The learning content of the education section was developed using the Graduate Record Examination (GRE) level vocabulary. The original set of words were extracted from the popular GRE vocabulary book ”WORD SMART for the GRE”. The original set was further developed using an online dictionary to add translations and example sentences. The final set of words were constructed of 2000 GRE-level vocabulary with information including translation, pronunciation, two example sentences and the translation of the example sentences.

The questions in the quiz section are all multiple choices designed to utilize all the available information in the vocabulary set. In total, there are three different types of questions including English to Korean translation, Korean to English translation and selecting an appropriate word for a blank space in a given sentence. The words presented as an answer choice in the questions are randomized from the entire vocabulary set and always ordered randomly in order to avoid any visual recognition of the answer from repetitive answer-sets.

3.4. Game section

As shown in Fig. 5 (left), the game's objective was to accumulate a higher score by hitting the targets that float horizontally on the top of the screen using the given arrows. The users can draw and shoot the arrow by touching or clicking on the screen of their smartphone device. The movement of the arrow depended on the type of arrow that is being shot and the wind moves the target at a random angle at a random strength (but within the maximum wind strength set for the game session). The score of the game is based on how accurately the user can shoot the arrow where each hit translates to points, starting from 100 at the bulls-eye and lower as it gets further from the center of the target. Also shown in Fig. 5 (right), a fever mode can be activated after a certain number of consecutive hits, depending on the stage level, where the user can accumulate a much larger sum of points as the fever target's score is

Fig. 4. Multiple choice questions in the educational module, description (left) and example (right).

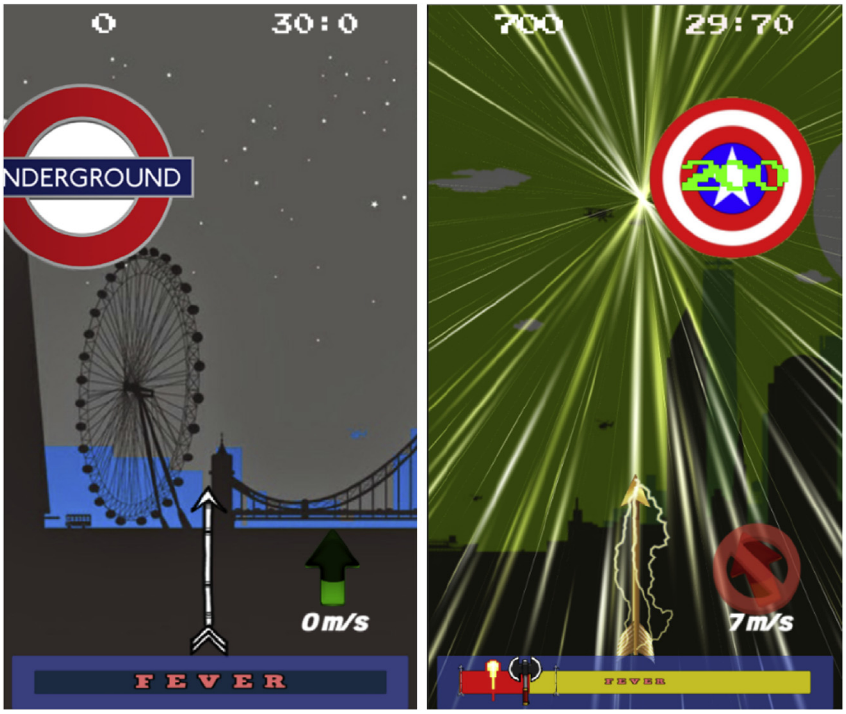

Fig. 5. The game module with a normal shooting mode (left) and a fever mode (right) that activates after several consecutive hits.

much higher than a normal target. The game ends when a user misses the target and has no remaining arrow to shoot.

In the game, there were 6 different arrow types including normal arrow, fireball, axe, boomerang, rocket and golden arrow. Normal arrow is the basic form that flew at normal speed. Fireball is a type of arrow that flew at half the normal speed but provided twice as many points when it hit the target. Axe is a type of arrow that rotated 360° when flying and flew at 0.75 times the normal speed. Boomerang is an arrow that had a circular trajectory towards the target, similar to a flying boomerang. Rocket had an exponentially growing speed to the movement, where it would start slow and accelerate. The golden arrow had twice the normal speed with lightning effects when it hit the target. All arrows were designed to be affected by the wind speed except for the golden arrow, which ignored the wind.

4. Method

4.1. Participants

Subjects were recruited on the official student forum of a large-size university in Korea, where the sign-up requested that people who were interested in both games and English vocabulary to volunteer for participation. A total of 64 students (42 male, 22 female; 42 undergraduate students, 22 graduate students; mean age 22.8) participated in the experiment and were given approximately $15 for 3 days of participation with pre and post survey including a vocabulary test within each survey. The subjects were randomly distributed to a control and a treatment group and showed no statistically significant difference in their motivation to play the game ($t = 0.13$, *n.s.*) to learn vocabulary ($t = 0.99$, *n.s.*), nor in their previous game experience ($t = 1.39$, *n.s.*) and English vocabulary studying experience ($t = 0.77$, *n.s.*). Only 4 subjects had previously studied for the GRE and their pre-test score showed no difference from the other participants.

4.2. Procedure

The experiment was conducted as a 3-day experiment. During the experiment, subjects were free to interact with the system on their own will and use the system as much as they wanted and as little as they wanted. During the pre-experiment session, subjects were briefly introduced to the experiment and were given a brief explanation about the purpose of the experiment. Next, they were presented with a vocabulary test and a pre-experiment survey, which asked for the subject's general information, experience and motivation in both game and English vocabulary learning.

Afterward we deployed the Android-based application onto each subject's smartphone device and basic instructions were given vocally. The application on launch provided a detailed tutorial on how to use the system. Specific details about the arrows and how they can be obtained were introduced in the tutorial. Subjects then used the application on their own will for 3 days and were gathered again for a vocabulary test and post-experiment survey, which included self-reported behavioral measures. Also, the subjects answered several questions regarding their system use in a follow-up interview. The study participants in each experimental condition were met separately and not told that there was an alternative condition.

4.3. Measures

All measures in the current study, except learning and perception of the system, were presented in a 7-point Likert scale, modified from the original version to fit the context of game-based learning. The specific questionnaire items used are found in Appendix 1.

4.3.1. Learning outcome

The learning outcome from the experiment has been measured through a pretest-posttest control group design to trace the impact of the intervention. Both the pre-test and post-test consisted of 25 multiple-choice questions. The tests were comprised of 25 words selected randomly from the vocabulary pool of 2000 words and all the questions were the same for all subjects. The two tests were identical but the subjects were not informed of it during the experiment and the order of the questions was shuffled. The learning outcome was measured by calculating the difference between the pre-test and the post-test for each user.

4.3.2. Autonomy and competence

The perceived autonomy and competence were measured as a part to determine the effect a reward structure has on the user as competence information. Perceived autonomy indicates the user's perceived level of control over the system's outcome. Perceived competence is a measure of the user's perceived level of his or her own capability to succeed within the system. Perceived autonomy and perceived competence are important predictors of intrinsic motivation according to the CET (Deci et al., 1999). These measures provide indirect evidence of how the rewards influence the user's motivation towards activity and insights to how the rewards are perceived to influence the user's experience. Specific questions for perceived autonomy are originated from Fu et al. (Fu, Su, & Yu, 2009) and perceived competence originated from Ryan and Deci (Ryan & Deci, 2000).

4.3.3. Engagement

The level of engagement was measured to demonstrate the user's state of involvement and absorption to the interaction with the system. Specifically in this study, the four different categories of cognitive absorption by Agarwal et al. (Agarwal & Karahanna, 2000) was used to measure engagement: enjoyment, immersion, temporal disassociation and curiosity. Enjoyment aims to measure the level of pleasure experienced during the interaction with the system. Immersion is a measure of the degree the user was able to ignore other attentional demands. Temporal disassociation measures the user's inability to register the passing of time during one's interaction with the system. Curiosity indicates the extent the interaction with the system arouses the user's sensory and cognitive curiosity. These engagement measures are used to assess the hedonic experience and how it alters based on the different reward structures in the current study.

4.3.4. Motivation

The subject's motivation to do well in each of the two contexts, gaming and learning, was measured using the trying to game and trying to learn measurement items, respectively, adopted from Ahuja and Thatcher (2005) (Ahuja & Thatcher, 2005). Theory of trying specifically postulates trying to act as a key mechanism in predicting behavioral outcomes and performance differences (Bagozzi & Warshaw, 1990). The current study employed the motivational measures based on the theory of trying because of the nature of game-based learning where the user's effort is paramount to its success. The measures of motivation for both gaming and learning are used to assess the different levels of motivation a user experiences based on the provided reward.

4.3.5. Perception of system

The perception of the system was measured to determine whether the subjects perceived the system to be a game or an educational application on a relative scale. The subjects were asked to rate the degree to which they felt the system was a game or an educational application in a scale of 1–10, 1 being strictly a game and 10 being strictly an educational application (Howard & McInnes, 2012; Vandercruysse, Vandewaetere, Cornillie, & Clarebout, 2013). The user's perception of the system is used to understand how the users viewed the system and how the view changed with the alternative reward conditions.

4.4. Interview

After the post-experiment survey and test, the subjects were asked several questions regarding their general response towards the system and their effort levels when using the system. They were encouraged to share their opinions on how positive or negative they felt about the system in terms of both learning and gaming and also opinions on how they tried to perform well in both learning and game.

5. Quantitative results

We conducted a series of statistical analysis to compare the differences between the control and treatment groups in 1) learning, 2) perceived autonomy and perceived competence, 3) engagement, and 4) perception of system. An independent T-test and the effect-size (Cohen's *d*) was used for the analysis.

5.1. Reliability and validity of measures

Before the analysis, the psychometric properties of the measurements were tested to assess their reliability and validity. First, the internal consistency for all constructs was tested using the Cronbach's α and composite reliability. Generally, when the measure exceeds 0.7, the construct is believed to be acceptable, and it is excellent when exceeding 0.9. As shown in Table 2, all constructs were at least over 0.7 with excellent consistency observed (over 0.9) for several measures. Furthermore, the convergent and discriminant validity were assessed using Average Variance Extracted (AVE) by comparing the square root of AVE (in bold) of each measure with the measure's correlation with other measures. As shown in Table 2, the convergent and discriminant validity is satisfied as the square root of AVE of every measure is higher than the correlation with other measures, indicating that the measurement items measure their intended construct well.

5.2. Learning

As shown in Table 3, the learning outcomes from the experiment showed that the treatment group (M = 10.969, SD = 3.848) had higher achievement than the control group (M = 8.563, SD = 4.399). An independent T-test showed the statistical significance of the difference with a $t = 2.329$, $p < 0.05$. Given that the difference of the pre-experiment test-score between the two groups showed no statistical significance with $t = -0.246$, $p = 0.807$, it is fair to assume that the two groups started with a similar level of knowledge.

Table 2
Reliability and validity of measures.

	Construct	Cronbach's Alpha	Composite reliability	AVE	Construct Correlations							
					1	2	3	4	5	6	7	8
1	Autonomy	0.777	0.867	0.685	**0.828**							
2	Competence	0.956	0.968	0.883	0.378	**0.940**						
3	Enjoyment	0.961	0.974	0.808	0.423	0.402	**0.900**					
4	Immersion	0.856	0.903	0.7	0.428	0.407	0.593	**0.837**				
5	Temporal Disassociation	0.874	0.941	0.888	0.334	0.36	0.658	0.728	**0.942**			
6	Curiosity	0.879	0.926	0.927	0.262	0.262	0.645	0.53	0.646	**0.963**		
7	Trying(game)	0.91	0.957	0.917	0.43	0.464	0.341	0.409	0.336	0.319	**0.958**	
8	Trying(learn)	0.901	0.953	0.91	0.239	0.317	0.327	0.466	0.352	0.362	0.694	**0.954**

Table 3
Sample T-Test Results. ***p < 0.001, **p < 0.01, *p < 0.05, + p < 0.10.

Construct		Group	Mean	SD	t-value	P-value	Effect size (Cohen's d)
Learning Outcome		control	8.563	4.399	2.329	0.023*	0.58
		treatment	10.969	3.848			
Perceived Autonomy		control	4.156	1.273	3.518	0.001***	0.88
		treatment	5.083	0.776			
Perceived Competence		control	2.734	1.079	2.254	0.028*	0.56
		treatment	3.484	1.542			
Engagement	Enjoyment	control	4.302	1.048	0.846	0.401	0.21
		treatment	4.594	1.646			
	Immersion	control	4.703	1.033	1.759	0.084+	0.44
		treatment	5.172	1.099			
	Temporal disassociation	control	3.484	1.247	2.542	0.014*	0.64
		treatment	4.375	1.54			
	Curiosity	control	3.979	1.206	1.277	0.206	0.32
		treatment	4.385	1.336			
Motivation	Trying (Game)	control	4.531	1.518	2.207	0.031*	0.55
		treatment	5.313	1.306			
	Trying (Learn)	control	4.281	1.524	2.037	0.046*	0.51
		treatment	5.031	1.42			
Perception of System		control	4.906	1.51	−3.115	0.003**	−0.78
		treatment	3.75	1.459			

Thus, the high learning gains of the treatment group show that performance contingent-rewards produced statistically significant learning enhancement.

To further understand the reason behind the learning gain, the system log data were analyzed in three categories for each subject including the number of sessions, the number of words exposed during the experiment and the time spent in the study section. The number of exposed words was calculated by counting the number of words from the pre-test to post-test each user studied in the system during the experiment. The results show that there were no statistically significant differences in the number of sessions (treatment: 23.44 times vs. control: 19.41 times, $t = 1.19$, $p = 0.24$) and the number of words exposed (treatment: 13.31 times vs. control: 13.59 times). However, the time spent in the study section showed a statistically significant difference (treatment: 10,762 s vs. control: 3,322 s, $t = 2.21$, $p < 0.05$). In sum, the results show that, despite not being more exposed or using the system more often, the subjects in the treatment group spent significantly more time studying the materials. As a result, subjects in the treatment group have achieved higher learning, as reflected in the learning outcome.

5.3. Perceived autonomy and perceived competence

The perceived autonomy of the treatment group (M = 5.083, SD = 0.776) was found to be significantly higher than the control group (M = 4.156, SD = 1.273) with $t = 3.518$, $p = 0.001$. The perceived competence of the treatment group (M = 2.734, SD = 1.079) was also found significantly higher than the control group (M = 3.484, SD = 1.542) with $t = 2.254$, $p < 0.05$. The results indicate that the treatment group perceived to have more control over the result of the task and to develop more competence than the control group, consistent with SDT (Ryan & Deci, 2000). For the users in the control and treatment groups alike, the perceived competence scores are below the middle point (3.5), showing that the participants overall felt that the game was not easy.

5.4. Engagement

Among the four categories of the engagement (enjoyment, curiosity, immersion and temporal disassociation) measures, immersion and temporal disassociation showed to have a statistically significant increase in the treatment group. For immersion, the treatment group (M = 5.172, SD = 1.099) and the control group (M = 4.703, SD = 1.033) showed marginally significant difference with $t = 1.759$, $p < 0.1$. For temporal disassociation, the treatment group (M = 4.375, SD = 1.540) and the control group (M = 3.484, SD = 1.247) showed a statistically significant difference with $t = 2.542$, $p < 0.05$. However, enjoyment ($t = 0.85$, $p = 0.4$) and curiosity ($t = 1.28$, $p = 0.21$) did not result in a statistically significant difference. While the system did not provide a higher level of enjoyment and curiosity to the subjects, it can be explained by the fact that the two groups were given the same game and the same learning contents. The differences in the experience for the two groups may have been subtle. The higher level of immersion and temporal disassociation experienced by the treatment group shows that there was a certain enhanced level of hedonic experience induced by the performance-contingent rewards.

5.5. Motivation

The treatment group (M = 5.031, SD = 1.420) showed a significant increase in trying to learn compared to the control group (M = 4.281, SD = 1.524). The difference between the two groups was statistically significant with $t = 2.037$, $p < 0.05$. For trying to game, the treatment group (M = 5.313, SD = 1.306) was higher than the control group (M = 4.531, SD = 1.518), with a statistically significant difference when assessed under *t*-test with $t = 2.207$, $p < 0.05$. The results indicate that the treatment group had a stronger effort level and motivation to do well for both learning and gaming compared to the control group. The performance-contingent reward intervention seems to have successfully motivated the subjects to try harder in the learning task for the reward.

5.6. Perception of system

For the perception of the system, treatment group (M = 3.750, SD = 1.459) reported a lower score than the control group (M = 4.906, SD = 1.510). The difference was statistically significant for $p < 0.01$ with $t = -3.115$. The results indicate that the treatment group perceived the system to be more of a game than an educational system whereas the control group perceived the system to be more of an educational system than a game.

6. Qualitative results

A series of interviews were conducted after the 3-day experiment to develop further insights on how users experienced each reward type. In general, the treatment group with the performance-contingent reward showed a more positive response and elaborated more on the reasons why they wanted to perform well in both learning and gaming.

6.1. Response to rewards

The subjects from the control group with the completion-contingent reward showed some positive responses toward the rewards. One subject stated *"In order to play the game again, I had to redo the studying and it got me going (P4)."* The subject showed that the motivation to play the game positively drove the subject to repeat the activity. However, as one subject explicitly commented, *"As I noticed that the studying had nothing to do with the game, I became lazy"* (P6), the minimum connection between the game section and the learning section was often not enough to make the users in this condition persevere.

The subjects from the treatment group with the performance-contingent reward showed that the rewards were directly responsible for the positive experience. A subject pointed out *"If I performed better at the test, I got more items so that was really exciting"* (P1) and *"The items in the game made me feel like I should study more to get better ones."* (P2). The subjects from the treatment group showed a clear understanding of the connection between the game section and the learning section while expressing much more positive responses towards the learning activity.

The different responses towards their rewards shown by the two groups highlight how a meaningful reward can induce a more positive attitude towards both learning and gaming in an extrinsic integration setting.

6.2. Effort

The subjects in each group clearly reflected differently in their level of effort for performance. Namely, when compared with the control group, a relatively smaller number of subjects in the treatment group indicated that they put little effort into performance. For the control group, 13 subjects (40.6%) expressed similar to *"I didn't put any efforts"* and 3 subjects (9.3%) indicated similarly for the treatment group. The results highlight how a completion-contingent reward can lack the ability to induce effort as users were only required to complete an activity. In contrast, performance-contingent reward induced the users to put conscious effort to perform well to gain the provided rewards.

6.3. Being more strategic

The treatment group responded with long and detailed responses in terms of their strategy with the learning section. Getting good grades was one of the strategies took by one of the subjects. A subject in the treatment group stated, *"It was crucial to get good grades in the learning section because the arrows with abilities were critical in getting high scores"* (P2). Some of the other subjects were more concerned about the speed of learning and how it affected their learning strategies. One subject said *"I studied the list from the learning instead of the cards, that way I could go through the vocabulary much faster and play the game more often"* (P7) while another responded *"I studied quickly to find out the words I don't know, that way I could get through the learning quickly and get high grade. After that, I concentrated on the game."* (P8). There was a general agreement towards the fact that they had to study quickly and get as many questions right as possible in order to be able to concentrate and get good scores in the game. There were even strategies where a subject got a question wrong on purpose. In the control group, most of the responses on the effort level were towards only the mechanical aspect of the game, as subjects respond *"I tried to time the release as carefully as possible"* (P9) or *"I checked the wind" (P4).* The responses are an indication of how the link between learning and gaming made by performance-contingent rewards allow the users to encompass both learning and gaming into consideration for better performance. On the other hand, completion-contingent reward creates more distinction between learning and gaming, resulting in a partition of activities.

7. Discussion

The ineffective designs of extrinsic integration have left the premise that extrinsic integration is less effective than intrinsic integration in the field of game-based learning (Habgood & Ainsworth, 2011). Yet extrinsic integration can address specific problems of intrinsic integration and an appropriate deployment can be educationally beneficial (Preist & Jones, 2015). The current study addresses the lack of available studies in further improving extrinsic integration by examining alternative integration mechanisms - performance-contingent reward as opposed to completion-contingent reward. Prior studies on extrinsic integration designs adapted a reward structure known as a completion-contingent reward, which only requires the user to complete a given task for the reward without any regards to the level of performance. Such reward structure limited the user's willingness to put effort into the learning activity, diminishing the system's effectiveness to educate. The current study employed a new reward structure, known as a performance-contingent reward, to provide a strong incentive for the user to try harder and to perform better, on the theoretical bases of CET and SDT. The results are encouraging as there are signs of improvement for learning and motivational outcomes.

7.1. Learning outcome

The current study found statistically significant evidence that performance-contingent rewards are superior to completion-contingent rewards for learning. The difference in the post-test scores between the two groups showed to be significantly different, while the pre-test scores showed no difference. The results demonstrate that extrinsic-integration designs can be further improved with a different reward structure. This prospective is encouraging especially because the rewards provided in the current study differ in type from those of successful previous studies (Habgood & Ainsworth, 2011; Preist & Jones, 2015). A combination or multiple reward structures could be further studied to find synergy and extend the research on extrinsic integration.

A possible explanation for the improved learning could be the increased level of motivation and effort level, indicated by the measure trying to learn. The level of trying of the treatment group was significantly higher than that of the control group. The results reflect the level of effort and motivation the user had in the learning activity (Ahuja & Thatcher, 2005). The evidence suggests that providing an incentive to perform well in the learning task can successfully motivate the users to put more effort and perform better in a game-based learning context. This explanation is further supported by the system usage data, as subjects from the treatment group spent more time studying compared to the control group. Especially considering that the treatment group did not have more exposure to the words from the pre-test to the post-test or to the number of sessions using the system, the increased level of learning observed should be interpreted with a significantly higher level of effort put into the actual studying.

There were some users who were disinterested from playing the game even though they belonged to the treatment group. A low level of effort was shown and the reward consequently failed to motivate those users. These exceptions are consistent with the self-determination theory (Ryan & Deci, 2000), which postulates that the external rewards can work differently depending on the level of internalization. The explanation provides useful practical advice, where extrinsic integration designs should be accompanied by a verification of user's interest in the game.

7.2. Motivation

A key success factor of game-based learning is whether the users were engaged with the system and was motivated to partake in the activity. The results from the current study show that the users from the treatment group were more immersed and more temporarily disassociated than those from the control group. The engagement level is also shown higher in the measure trying to game, as the treatment group reported to have tried more to perform well in the game activity than the control group. One possible explanation for the enhanced effort level and engagement can be found from the cognitive evaluation theory (Deci et al., 2001). As the CET states, when the reward provides enough competence information to sufficiently negate the controlling effect, the motivation for the activity would be positively influenced. A possible explanation for the higher level of perceived competence and autonomy in the treatment group could be that the performance-contingent rewards were successful in delivering sufficient competence information, thereby allowing the reward to positively influence the motivation of the users.

A possible explanation of why subjects in the treatment group were more likely to perceive the reward as competence information may be due to the meaningful link between learning and gaming. As the qualitative results clearly indicate, the treatment group built a clearer connection between the two activities in terms of overall performance. Therefore, the rewards provided for their performance in the learning activity not only indicated their learning achievement but also developed a meaningful chance to perform well in the game. In contrast, the control group built a distinction between the two activities, which limited the contribution of the reward to the overall performance. Therefore, while the reward was an indication of competence for successful completion of the learning activity, the limited contribution it made to the game prevented it from motivating the subjects more positively.

7.3. Perception

The users' perceptions of the system, which can alternate between game and learning, presented intriguing findings in terms of both quantitative and qualitative results. The current study measured the user's perception of the system type, whether it to be a game or an educational application, as an operational measure to see the game's quality (Salomon, 1984). The results showed that the two groups demonstrated a statistically significant difference in their perception of the system. The treatment group perceived the system to be more game-like whereas the control group perceived the system to be more educational. These contrasting perspectives were also observed in the qualitative interview, where the users from the treatment group elaborated that they accepted the learning task as a part of the game, as a preparation stage, while the users from the control group identified the game as a reward and an enjoyable extra activity. However, there are no other empirical studies that address this particular condition in game-based learning other than the study by Vandercruysse (Vandercruysse et al., 2013), who briefly mentions the phenomenon in their work. Other fields such as child education have evaluated similar concepts and demonstrated its link with the emotional well-being of children (Howard & McInnes, 2012). While the specific reason for the difference in the perception of system is beyond the scope of the current study, the finding poses an important point to consider for game-based learning research. Prior work in game-based learning had an underlying assumption that users would view and respond to the game-based learning system as the researchers had intended (Braghirolli, Ribeiro, Weise, & Pizzolato, 2016; Ronimus et al., 2014). However, the results from the present study suggest that such assumptions maybe prematurely made and require further attention to understanding how it influences the users in response to a game-based learning system.

7.4. Design implications

7.4.1. Extendable design

One of the key advantages of extrinsic integration is its highly extendable design nature due to the independent context in which learning and gaming are conducted. The current study focused on a reward structure that was designed to integrate the two different contexts with the objective of a higher level of learning while maintaining the enjoyment of the game. The current design can be easily extended to other subject domains of education and other gaming materials. There need to be considerations of the format in which the new subject domain is studied and the reward structure that can be suited for the particular game design. However, the implementation would be vastly simpler compared to designing an entirely new game-based learning system based on intrinsic integration. Furthermore, specific types of educational domains will be much more appropriate for extrinsic integration. For example, subjects that require practice and repetition, such as mathematics and languages, are best suited for the current design. Also, activities such as exam preparation could be complemented by the current design strategy.

7.4.2. Maintaining a fun environment

The findings from both quantitative and qualitative data point to one fundamental requirement for the current design to be successful - making the game fun. A key component of the current design's success is given through the rewards that are meaningful to gameplay, thus providing a strong incentive for the users to put the effort in the learning stage. Maintaining a fun environment is a particularly important consideration for future designs, as it would significantly alter the development of future systems. For example, subject materials that require a longer period to study or require multiple stages of progression would be too long to maintain a level of fun for a simple game. Along with the extended materials, the game would require a deeper design for it to remain fun and keep balance with them. Utilizing a game design that can be engaging for a lengthy period, such as online competition or highly randomized tasks would be necessary to fully benefit from current design in different educational and game environments.

7.4.3. Empowering the users

While common extrinsic rewards such as trophies and badges have their own value among users, their lack of direct contribution to a meaningful experience limits their influence on users. The current study designed the reward structure based on the motivational theories by Ryan and Deci (Ryan & Deci, 2000), in a way that it can positively influence the user's autonomy and competence. A reward that directly alters the gameplay, where the users perceive their chances of performing better in-game becomes higher, showed positive influences on their level of trying in both game and education. It would be particularly important for future designs of similar systems to focus on the discovery of what is meaningful in the game, and what rewards can empower the users to achieve their meaningful objective.

7.5. Limitations and future opportunities

There are several limitations to this study that should be considered along with the interpretation of the results. First, the study was conducted on a short-term period of three days. Although statistically significant learning improvements have been observed in the experiment, the short time period of the experiment limits the interpretation of the result to a short-term knowledge gain. In future research, conducting an experiment for a longer term or conducting a test after a delayed period of time to examine the subject's retention of the learning material may provide more robust results.

Second, the subjects were recruited voluntarily and they may not necessarily represent the behavior of the population. Thus, the results from this study may require a further study to identify the effect of the treatment in a non-voluntary setting.

Finally, the level of difficulty in the vocabulary was not controlled and was not incremental. The set of 2000 GRE English vocabulary may be more difficult than what the subjects were accustomed to. Although the effects of a fitted difficulty were not within the scope of this study, different categories of vocabulary to provide a variety of difficulty may have been more approachable and encouraging to some subjects. A future study may benefit from analyzing how difficulty in learning material can influence the experience of game-based learning.

Notwithstanding these limitations and future opportunities, it should be noted that the current study successfully demonstrated the efficacy of performance-contingent rewards in creating synergistic effects of game and learning by meaningfully connecting the two distinct but indispensable components of game-based learning. Our findings show that the seemingly two opposing forces, or foes, of game-based learning (i.e., game and learning) can not only live in harmony but also help each other. The results are further encouraging as the integration strategy is rapidly applicable to many other learning domains, without requiring the intricate mash-up of educational elements within the gaming scenarios.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2019.04.016.

References

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*, 665–694.

Ahuja, M. K., & Thatcher, J. B. (2005). Moving beyond intentions and toward the theory of trying: Effects of work environment and gender on post-adoption information technology use. *MIS Quarterly, 29*, 427–459.

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review, 3*, 185–201. https://doi.org/10.1016/1053-4822(93)90012-S.

Bagozzi, R. P., & Warshaw, P. R. (1990). Trying to consume. *Journal of Consumer Research, 17*, 127–140. https://doi.org/10.1086/208543.

Braghirolli, L. F., Ribeiro, J. L. D., Weise, A. D., & Pizzolato, M. (2016). Benefits of educational games as an introductory activity in industrial engineering education. *Computers in Human Behavior, 58*, 315–324. https://doi.org/10.1016/j.chb.2015.12.063.

Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodrguez-Aflecht, G., Pongsakdi, N., et al. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education, 128*, 63–74. https://doi.org/10.1016/j.compedu.2018.09.011.

Cheng, M. T., Lin, Y. W., & She, H. C. (2015). Learning through playing virtual age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18–29. https://doi.org/10.1016/j.compedu.2015.03.007.

Cheong, C., Cheong, F., & Filippou, J. (2013). Using design science research to incorporate gamification into learning activities. *Pacific asia conference on information systems* (pp. 156). .

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., et al. (2010). Predicting protein structures with a multiplayer online game. *Nature, 466*, 756–760. https://doi.org/10.1038/nature09304.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627–668.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*, 1–27. https://doi.org/10.3102/00346543071001001.

Derboven, J., Zaman, B., Vissers, J., Geerts, D., & De Grooff, D. (2015). The fun and the serious in an educational game: The monkey tales case. *Proceedings of the 33rd annual ACM conference on human factors in computing systems, ACM, New York, NY, USA* (pp. 2829–2832). . https://doi.org/10.1145/2702123.2702519.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). New York, NY, USA: ACM. https://doi.org/10.1145/2181037.2181040.

Eseryel, D., Law, V., Ifenthaler, D., Ge1, X., & Miller, R. (2014). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Journal of Educational Technology & Society, 17*, 42–53.

Filsecker, M., & Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. *Computers & Education, 75*, 136–148. https://doi.org/10.1016/j.compedu.2014.02.008.

Fu, F. L., Su, R. C., & Yu, S. C. (2009). Egameflow: A scale to measure learners enjoyment of e-learning games. *Computers & Education, 52*, 101–112. https://doi.org/10.1016/j.compedu.2008.07.004.

Glover, I. (2013). Play as you learn: Gamification as a technique for motivating learners. *Proceedings of EdMedia 2013–world conference on educational media and technology, association for the advancement of computing in education (AACE), victoria, Canada* (pp. 1999–2008). .

Gunter, G., Kenny, R. F., & Vick, E. H. (2006). A case for a formal design paradigm for serious games. *The Journal of the International Digital Media and Arts Association, 3*, 93–105.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20*, 169–206. https://doi.org/10.1080/10508406.2010.508029.

Hitt, D. D., Marriott, R. G., & Esser, J. K. (1992). Effects of delayed rewards and task interest on intrinsic motivation. *Basic and Applied Social Psychology, 13*, 405–414. https://doi.org/10.1207/s15324834basp1304_2.

Howard, J., & McInnes, K. (2012). The impact of children's perception of an activity as play rather than not play on emotional well- being. *Child: Care, Health and Development, 39*, 737–742. https://doi.org/10.1111/j.1365-2214.2012.01405.x.

Huang, B., & Hew, K. F. (2015). Do points, badges and leaderboard increase learning and activity: A quasi-experiment on the effects of gamification. *Proceedings of the 23rd international conference on computers in education, asia-pacific society for computers in education* (pp. 275–280). .

Hwang, G. J., Hsu, T. C., Lai, C. L., & Hsueh, C. J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education, 106*, 26–42. https://doi.org/10.1016/j.compedu.2016.11.010.

Infante, C., Weitz, J., Reyes, T., Nussbaum, M., Gómez, F., & Radovic, D. (2010). Co-located collaborative learning video game with single display groupware. *Interactive Learning Environments, 18*, 177–195. https://doi.org/10.1080/10494820802489339.

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research & Development, 64*, 219–244. https://doi.org/10.1007/s11423-015-9418-1.

Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education, 87*, 340–356. https://doi.org/10.1016/j.compedu.2015.07.009.

Linehan, C., Kirman, B., Lawson, S., & Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1979–1988). New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1979229.

Liu, D., Santhanam, R., & Webster, J. (2017). Towards meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly, 41*, 1011–1034.

Lomas, D., Patel, K., Forlizzi, J. L., & Koedinger, K. R. (2013). Optimizing challenge in an educational game using large-scale design experiments. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 89–98). New York, NY, USA: ACM. https://doi.org/10.1145/2470654.2470668.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world.* New York, NY, USA: Penguin.

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation?: An empirical analysis of common gamification elements. *Proceedings of the first international conference on gameful design, research, and applications* (pp. 66–73). New York, NY, USA: ACM. https://doi.org/10.1145/2583008.2583017.

Meyer, B., & Sørensen, B. H. (2009). *Designing serious games for computer assisted language learning – a framework for development and analysis.* Dordrecht: Springer Netherlands.

Perry, D., Robinson, J., Cruz, S., Aragon, C., Chowning, J. T., & Peters, M. (2014). Game design for bioinformatics and cyberinfrastructure learning: A parallel computing case study. *Concurrency and Computation: Practice and Experience, 26*, 2303–2315. https://doi.org/10.1002/cpe.3261.

Plass, J. L., Homer, B. D., Hayward, E. O., Frye, J., Huang, T. T., Biles, M., et al. (2012). The effect of learning mechanics design on learning outcomes in a computer-based geometry game. In S. Göbel, W. Müller, B. Urban, & J. Wiemeyer (Eds.). *E-learning and games for training, education, health and sports* (pp. 65–71). Berlin, Heidelberg: Springer Berlin Heidelberg.

Preist, C., & Jones, R. (2015). The use of games as extrinisic motivation in education. *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3735–3738). New York, NY, USA: ACM. https://doi.org/10.1145/2702123.2702282.

Prensky, M. (2003). Digital game-based learning. *ACM Computers in Entertainment, 1*, 21–23. https://doi.org/10.1145/950566.950596.
Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education, 71*, 237–246. https://doi.org/10.1016/j.compedu.2013.10.008http://www.sciencedirect.com/science/article/pii/S0360131513002935.
Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761. https://doi.org/10.1037//0022-3514.57.5.749.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68.
Salomon, G. (1984). Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*, 647–658. https://doi.org/10.1037/0022-0663.76.4.647.
Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education, 76*, 119–130. https://doi.org/10.1016/j.compedu.2014.03.006.
Santhanam, R., Liu, D., & Shen, W. C. M. (2016). Research notegamification of technology-mediated training: Not all competitions are the same. *Information Systems Research, 27*, 453–465. https://doi.org/10.1287/isre.2016.0630.
Tobias, S., Fletcher, J. D., & Wind, A. P. (2014). *Game-based learning.* New York, NY: Springer New York.
Tomaselli, F., Sanchez, O., & Brown, S. (2015). How to engage users through gamification: The prevalent effects of playing and mastering over competing. *Proceedings of the 36th international conference on information systems*. Association for Information Systems.
Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *Educause Review, 41*, 16.
Vandercruysse, S., Vandewaetere, M., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research & Development, 61*, 927–950. https://doi.org/10.1007/s11423-013-9314-5.
Zapata-Rivera, D., VanWinkle, W., Doyle, B., Buteux, A., & Bauer, M. (2009). Combining learning and assessment in assessment-based gaming environments: A case study from a New York city school. *Interactive Technology and Smart Education, 6*, 173–188.
Zhonggen, Y. (2018). Differences in serious game-aided and traditional English vocabulary acquisition. *Computers & Education, 127*, 214–232. https://doi.org/10.1016/j.compedu.2018.07.014.