Computers & Education 73 (2014) 149–159

Contents lists available at ScienceDirect

Computers & Education

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

Integrating self-explanation functionality into a complex game environment: Keeping gaming in motion

Deanne M. Adams*, Douglas B. Clark

Department of Teaching and Learning, Vanderbilt University, Nashville, TN 37203, USA

A R T I C L E I N F O

Article history:
Received 18 October 2013
Received in revised form
13 December 2013
Accepted 8 January 2014

Keywords:
Self-explanation
Educational games
Feedback
Physics learning

A B S T R A C T

Previous research has shown that either asking students to explain their answers or providing explanatory feedback can be effective ways to increase learning from an educational game. This study focused on an educational physics game about Newton's 3 Laws of Motion called *SURGE: The Fuzzy Chronicles*. Eighty-six middle school students played one of three versions of the game: (1) the base version with no tips or questions, (2) the self-explanation version with self-explanation questions prompts, and (3) the explanatory feedback version with gameplay tips. There were no significant overall learning differences between the three groups, but students in the base version successfully answered more questions about Newton's second law than students in the self-explanation group. This may have been due to students in the base condition progressing significantly further through the game than students in the self-explanation group. The results suggest that the cognitive load for gameplay as well as game flow must be managed in order for students to take advantage of explanation functionality in educational tools designed to increase deeper, germane processing.

1. Introduction

Recent meta-analyses looking at research with serious games (Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013) and simulation games (Sitzmann, 2011) suggest that games can be effective learning devices. Games can provide powerful affordances for motivation and learning. For example, individual studies have shown that games can promote conceptual understanding and process skills (e.g., Annetta, Minogue, Holmes, & Cheng, 2009; Clark et al., 2011; Hickey, Ingram-Goble, & Jameson, 2009; Klopfer, Scheintaub, Huang, Wendal, & Roque, 2009; Moreno & Mayer, 2000, 2004), can foster an epistemological understanding of the nature and development of science knowledge (e.g., Barab, Sadler, Heiselt, Hickey, & Zuiker, 2007; Neulight, Kafai, Kao, Foley, & Galas, 2007), and can produce gains in players' willingness and ability to engage in scientific practices and discourse (e.g., Barab et al., 2009; Galas, 2006; McQuiggan, Rowe, & Lester, 2008). Unfortunately, the field of educational games still lacks clear guiding principles for designing effective games based on empirical research. In the preface to their book, O'Neil and Perez (2008) point out that while there are may be hypothesized reasons for why educational games could be effective (e.g., interactivity and motivation) there is less actual research on how games should be designed to facilitate learning. Studies which aim to examine how changing game features can either affect learning outcomes or player motivation have been labeled value-added studies (Mayer, 2011). This study takes a value-added approach to examine the efficacy of adding self-explanation questions or explanatory feedback to an educational physics game to improve learning.

2. The case for self-explanation

According to the theory of multimedia learning, students must be cognitively active in order for productive learning to occur (Mayer, 2009). While a student may be behaviorally active while playing an educational game (e.g., pressing buttons), overt activity such as clicking on game elements does not guarantee that the student is cognitively active. Even during highly interactive educational games, students can "zone-out" or learn to "game the system" (Baker, 2008) instead of engaging conceptually with the educational material

* Corresponding author.
E-mail addresses: deanne.m.adams@vanderbilt.edu, deanne.adams@gmail.com, adams@psych.ucsb.edu (D.M. Adams), doug.clark@vanderbilt.edu (D.B. Clark).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2014.01.002

underlying the game mechanics. This issue is especially challenging for what Clark and Martinez-Garza (2012) call conceptually-integrated games. Conceptually-integrated games integrate learning goals directly into the game-play mechanics of navigating or manipulating the game-space (rather than simply situating learning goals and activities within the game space/world) such that players must master the learning material in order to succeed within the game. This framing of educational game design is similar to Habgood and Ainsworth's (2011) intrinsic integration approach in which the learning material is incorporated into the core mechanics of the game without interrupting the experience of flow (Csikszentmihalyi, 1990). One issue with conceptually-integrated/intrinsic integration games is that while players may spend a significant amount of time interacting with the game/conceptual mechanics, students may never explicitly articulate the relationships inherent in those mechanics. A learner can therefore potentially master the mechanics at an intuitive level without explicitly articulating those relationships in terms of formal learning goals. Therefore it is important that conceptually-integrated games encourage learners to explicitly articulate the connections between their game experiences and the formal learning goals (Clark & Martinez-Garza, 2012).

One proposed approach for connecting gameplay with formal learning goals is through using self-explanation. During self-explanation, students are asked to explain their problem solving processes to themselves as they engage in a learning activity. Key cognitive mechanisms involved in the process include: (a) filling in missing information by generating inferences, (b) integrating information within the lesson, (c) integrating new information with prior knowledge, and (d) confronting and correcting incorrect information (Roy & Chi, 2005). Although all of these processes can be cognitively demanding, they encourage students to engage in deeper, more meaningful learning by encouraging students monitor what they do and do not understand about the material. In a review of self-explanation and multimedia learning, Roy and Chi (2005) found that self-explanation effectively increases deep learning gains and that the benefits are larger for multimedia environments than for text-only lessons.

One issue Roy and Chi (2005) present is that self-explanation can place a large cognitive load on the learner. It is critical that the design of the instructional activity reduces as much of the cognitive load as possible so that students have adequate resources available to engage in such a cognitively demanding task. Similar to text-based and multimedia learning, if an education game overloads cognitive processing, then adding self-explanation prompts may not benefit students since they will be unable to engage in deeper processing.

Despite the amount of research exploring self-explanation in multimedia learning, significantly less research has explored self-explanation in educational or serious games. Moreno and Mayer's (2005) work with the simulation game *Design-a-Plant* (Lester, Stone, & Stelling, 1999) examined three possible design features for a simulation game: (1) interactivity (being given versus selecting the correct answer); (2) reflection (whether students are asked to orally explain their solutions); (3) and guidance (given corrective or explanatory feedback). Their first study found that students who were asked to give oral explanations for their answers showed no significant benefit over students who were not asked to reflect on their answers. This was true for both retention questions and transfer measures. The experiment did find that students performed significantly better when provided with explanatory feedback. Unlike corrective feedback, which only tells students whether they are correct or incorrect, explanatory feedback provides a rationale for why the student's answer is correct or incorrect. The authors proposed that the lack of an effect for the reflective self-explanation task may have been due to the fact that the interactivity of the lesson already required students to engage in active cognitive processing. Therefore there was no added benefit from adding another mechanism aimed at encouraging deeper processing.

Moreno and Mayer's (2005) second experiment examined whether manipulating interactivity affected the efficacy of including reflective activities in the *Design-A-Plant* game. The results showed an interaction between the interactivity level of the game and whether students were asked to reflect on their answers. Students in the interactivity condition performed best on retention measures when they were not asked to reflect on their answers while students in the non-interactive version of the game performed best on retention and far transfer items when asked to engage in reflection. The authors suggested that this occurred because students in the non-interactive version were required to reflect only on correct answers while students in the interactive condition were also prompted to reflect on their incorrect choices. The authors' third experiment confirmed this proposition. Students in the interactive version of the game were asked to explain correct solutions provided by the game after the players had made their selections (correct or incorrect). The overall results showed that students were more likely to give correct explanations when provided with the correct solution either in the non-interactive group or the interactive group when the correct solution was provided. This set of three studies suggests that explanation is only effective in a game environment that does not already encourage active cognitive processing through interactivity and if students are only reflecting on correct information.

Similar results were also found by Mayer and Johnson (2010) with a game-like environment dealing with electrical circuits. During the game, students needed to determine how the electrical current in a circuit changes when batteries and resistors are added in serial and parallel arrangements. The study compared the benefits of using self-explanation and explanatory feedback on performance on a transfer task which took place during the final/10th level of the game. Instead of providing their own oral self-explanation when prompted by the game, students in the self-explanation group were asked to select from a list of possible reasons for the correct solution. The study showed that students who received self-explanation prompts, explanatory feedback, or a combination of the two, demonstrated significantly higher performance on the transfer task compared to students in the base version of the game. In addition, compared to the control condition, participants in the three experimental groups also performed significantly better while playing the game in terms of their accuracy on the first nine levels. However, there were no significant differences between the three experimental conditions on game performance suggesting that there were no advantages of receiving both self-explanation and explanatory feedback, or receiving only self-explanation over explanatory feedback. The study shows that both explanatory feedback and self-explanation can increase student learning outcomes on transfer tasks as well as increase students' effectiveness while playing the game. The self-explanation results were replicated in another study with the same game (Johnson & Mayer, 2010).

Although Mayer and Johnson's (2010) results were similar to Moreno and Mayer's (2005) findings, the types of self-explanation used in the two studies were very different. Hausmann and Chi (2002) had found that having students type their self-explanations into a computer while learning from a multimedia lesson caused students to engage more in paraphrasing the information instead of constructing explanations. Students in the self-explanation group showed no benefit over the read only condition in terms of understanding the material possibly due to the cognitive demands of having to produce their answers or the typing format suppressing more spontaneous thinking typically seen with oral explanations. While Moreno and Mayer (2005) used oral self-explanation with *Design-A-Plant* to address this problem, this may not be feasible for game-like environments in classroom settings. Teachers cannot always engage in one-on-one interactions with students. Integrating self-explanation effectively into the game is an important design consideration. Mayer and Johnson's (2010) results suggest that providing students with possible explanation choices is an effective way to increase learning. An additional study by Johnson and Mayer (2010) revealed that having students select an explanation from 8 possible options was more effective that having them generate an explanation by typing their reasoning into the game. The results showed that the self-explanation generation group did not significantly differ from the base version of the game on transfer performance. The authors theorized that the inclusion of the self-generated explanations disrupted the flow of the game and increased extraneous processing. The study points to the importance of making sure that the self-explanation portions of educational games should facilitate generative and intrinsic processing while not adding needless extraneous processing.

The work of Mayer and colleagues suggests that self-explanation will not be effective when it adds extraneous processing or when the game already encourages deeper processing. Therefore it is important to consider how to balance cognitive demands when implementing self-explanation functionality into a game. Although having students select an explanation from a set of pre-generated options can alleviate cognitive load, this approach does not guarantee that students will engage in deeper processing. For example, students could "game" the system by selecting every multiple choice answer until the game indicates that the correct answer has been chosen (Baker et al., 2008). In their educational game study with shadows, Hsu, Tsai, and Wang (2012) found that the performance of the participants in the self-explanation group varied significantly depending on the engagement level of those participants. Students in the self-explanation group were divided into high and low engagement levels according to the ratio of correct versus incorrect/"I don't know" responses to self-explanation prompts that students had to answer after making a mistake during the game. While there were no significant overall differences between their control and self-explanation conditions, high self-explanation engagement students scored significantly higher than both the control participants and the low engagement self-explanation participants on the retention test. While self-explanation prompts may facilitate learning for students who actively engage in the process, there is no guarantee that students will engage cognitively with those prompts.

3. Present study

The present study examines the effects of adding either self-explanation questions or explanatory feedback into an educational game dealing with Newton's laws of motion. Previous research by White (1993) has shown that simulation environments can help teach physics by creating simple, controlled environments that can be manipulated by the learner. While Mayer and colleagues' research on the electric circuits and *Design-A-Plants* games demonstrated benefits for prompting students to reflect on correct solutions as well as receiving explanatory rather than corrective feedback, both of these games dealt with much simpler game environments. In *Design-A-Plant*, students only make three decisions regarding which components to use when constructing their plants (8 possible root types, 8 stem types and 8 types of leaves) (Lester et al., 1999). Students playing the electric circuits game only had to problem solve using combinations of up to 4 elements at a time (one or two batteries and resistors in single, parallel or serial configurations). The game used for the present study, *The Fuzzy Chronicles*, requires students to deal with more elements while engaging in problem solving such placing forces and adjusting their quantity and direction. The self-explanation questions used in *The Fuzzy Chronicles* are similar to the prompts used by Johnson and Mayer (2010), although students are only selecting from four possible answers choices (rather than eight) to explain why certain actions are important to their solutions and why they may have failed to solve the level. Similarly, students in the explanatory feedback condition receive tips and feedback relating to the core physics concepts. We predict that students who receive either explanatory feedback or self-explanation question prompts should demonstrate greater pretest–posttest gains than the students who receive the base version of the game.

In addition to looking at differences between the three conditions, this study will also explore possible interactions between self-explanation and explanatory feedback with gender. Previous research has found differences in video game experience and usage between males and females with males typically spending more time playing games (Homer, Hayward, Frye, & Plass, 2012; Paraskeva, Mysirlaki, & Papagianni, 2010). Physics learning has also shown a significant gender gap favoring males although females can benefit from instruction that uses more interactive strategies (Lorenzo, Crouch, & Mazur, 2006). Due to these issues, gender will be taken into consideration as a factor that could possibly affect learning outcomes or gameplay progression.

4. Methods

4.1. Participants

Participants included 100 eighth grade students from a diverse public middle school with a 38.56% rate of free/reduced lunch eligibility. Ten students were removed from the sample due to not having their parents' consent in order to use their data. Students were also removed if they did not complete the entire posttest or if they accessed and completed any part of the posttest prior to the official posttest administration. This resulted in an additional 4 participants being removed from the analysis leaving a total of 86 students (Male = 49, Female = 37). The students were randomly assigned to one of the three game conditions: 28 in the no self-explanation or explanatory feedback group (base game), 27 in the explanatory feedback tips only condition (explanatory feedback), and 31 in the self-explanation questions group (self-explanation). A chi-squared analysis revealed equal numbers of males and females across the three groups, $X^2(2, N = 86) = 1.97$, $p = .37$.

4.2. Materials

Students in this study played *The Fuzzy Chronicles*, which is a conceptually-integrated educational physics game focusing on force and motion. The narrative of the game is that the player is a space explorer navigating through space trying to save a friendly group of aliens known as "Fuzzies" from an evil group known as "Pricklies". The stages for the game are laid out on a star map with each star representing a mission with three possible levels of difficulty (bronze, silver, and gold). After completing the bronze level of difficulty, students unlock the silver difficulty level for that mission as well as the bronze difficulty level for the next mission. Completing the silver difficulty level of a mission unlocks the gold difficulty level of the mission. Each level takes place on a grid that contains a starting point as well as a goal (a portal/door) that must be reached in order to successfully complete the level. Each square in the grid represents one square meter of space. Players navigate amongst various obstacles and sub-goals between the starting and ending points of a level.

The game begins with a tutorial given by a Fuzzy general who explains how to plan out and manipulate the ship's trajectory as well as how to place actions on the timeline action bar. When starting a level, students plan out the trajectory of their ship by setting waypoints. Each waypoint represents were an action takes place such as changing the speed or direction of the ship or performing an action on a Fuzzy. Even if the player reaches the goal, the game will not let them proceed unless their solution matches the planned path. This was designed so that students paid attention to how the paths and the actions on the timeline action bar corresponded. To move the spaceship, students place rocket boost actions on the timeline action bar at the bottom of the screen. A sample screen shot of the game including the action bar can be seen in Fig. 1. Each square on the action bar represented a second's worth of time. For each force they had to indicate the amount of force (10 N, 20 N, 30 N, 40 N, 60 N, 80 N, 100 N, and 120 N) and the direction in which the force was going to be applied (left, right, up, and down). Rocket boosts have a duration of 0.1 seconds, so a rocket boost with 20 Newtons of force changes the spaceship's velocity by 2 meters per second in the direction of the rocket boost if it is not carrying any Fuzzies. The spaceship has a mass of 1 kg and each Fuzzy pod ship also has a mass of 1 kg. Only two forces could be applied at a time although certain special actions (shields, electric Fuzzy fields, and disguises) could be used in conjunction with two forces. For any of the actions involving Fuzzies (picking up, dropping, and launching), only one action could be used during that second's worth of time.

For each level there is a starting screen in which the Fuzzy general gives the players tips about the level and tells them what new or key game mechanics are involved. After setting their path and placing forces on the action bar, students can press the lever above the action bar to launch their spaceship and see if their solution can guide the spaceship to the end portal safely. During a level, students receive feedback from a robot character called Fuzzbot. The major manipulation for the current study deals with type of feedback that students received after the following actions during game: (a) before testing their solutions by pulling the lever, (b) after an incorrect solution, and (c) upon successfully completing a level. Students in the base condition only received correct/incorrect feedback for their solution attempts, such as Fuzzbot saying, "You blew it!" after an unsuccessful attempt and "Great Job!" after a successful attempt. Students in the explanatory feedback group received a tip after the same game actions. Tips included items such as saying that according to Newton's 1st law the spaceship will not move unless an unbalanced force acts upon it. For the self-explanation group, instead of receiving tips, students had to answer questions after the three different types of game actions. The questions were placed at these three positions to mirror the predict–observe–explain (POE) teaching method often used by science teachers to address preexisting misconceptions in science (Chinn & Malhotra, 2002; White & Gunstone, 1992). The questions at the beginning of the stage require student to predict how certain actions would affect their ship. For example, students would have to explain why it was important to apply a force at the beginning of the stage because Newton's 1st law states that an object at rest will stay at rest unless acted upon by an unbalanced force. Questions after each failure were based off of observing the behavior of their ship during the attempt and trying to explain why the solution was incorrect. Finally, the last question was use to explain what happened during the game level in terms of physics principles. Each question had four answers to choose from and students could not return to the game until they answered the question correctly. Self-explanation students did not see scores for the individual question items, but they did receive an overall score after successfully completing a level based on the number of attempts they made averaged over all of the questions.

Missions in the game were arranged so that game mechanics dealing with different laws of motion were slowly introduced. The game included a total of 35 missions with three levels of difficulty each. The first 11 missions involved applying forces to alter the speed and direction (velocity) of the spaceship dealing predominately with the 1st law relationships. The next five missions involved picking up and dropping Fuzzies (which includes increasing and decreasing the mass of the spaceship). These levels demonstrated the relationship between velocity and mass as related to the 2nd law of motion, F-MA. The next 7 missions involve launching Fuzzies. When a Fuzzy is launched, the spaceship is pushed in the opposite direction of the launch demonstrating the 3rd law of motion (for every action there is an equal and opposite reaction). The final 12 missions used different combinations of elements from the earlier levels while adding new game elements to act as obstacles and challenges during the game.

The three versions of the game were hosted on a local server using the WISE (web-based inquiry science environment) interface software as a project host. Students played the game individually on Macbook Air computers with 11 inch screens.

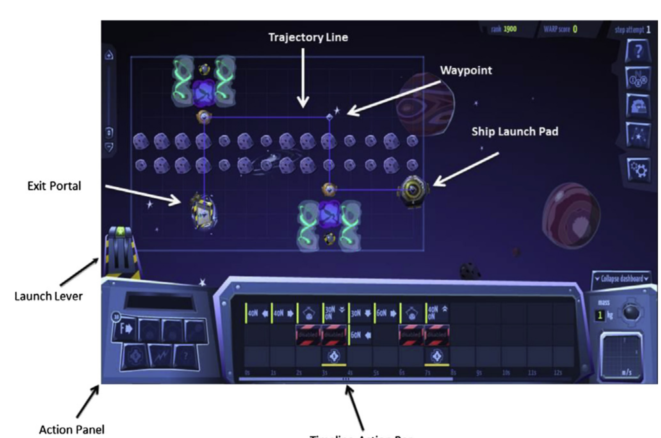

Fig. 1. Screen shot of a level from *The Fuzzy Chronicles*, including action plan that required students to launch Fuzzies.

4.3. Measurement

To assess learning gains between the pretest and the posttest, a 20-item multiple-choice test was used. A sample item from the test can be seen in Fig. 2. Questions asked students to determine what would happen if a certain force was applied either to a moving or stationary object, what combinations of forces could create a certain trajectory, what would happen if the mass or force acting on an object changed, and what would happen if an object was thrown using a particular force in a specific direction. More specifically, six multiple-choice questions dealt with Newton's 1st law of motion in one dimension, five dealt with Newton's 1st law of motion in two dimensions, five dealt with Newton's 2nd law of motion, and four questions dealt with Newton's 3rd law of motion. In addition to the multiple-choice test, students also completed a short-answer posttest, which involved a sheet of paper containing two of the multiple choice questions (one dealing with 1st law and one dealing with 2nd law) with the correct solutions circled on them. Students were asked to explain why that answer was correct according to Newton's laws of motion. The two short answer questions can be seen in Fig. 3. Only 1st and 2nd law questions were included on the short answer pretest due to previous pilot data showing that the majority of students only reached game levels dealing with these relationships.

4.4. Procedure

Before starting gameplay students received a consent form for their parents to sign as well their own assent forms. Students were told that their data would not be used in the analyses if either they or their parents were uncomfortable with participating. The classes were awarded with a free pizza lunch as compensation for their participation in the project. The experiment took place during five 1-hour class periods over the course of one week. Four separate classes of students (all with the same teacher) participated in the experiment. Students were randomly assigned to one of the three conditions within each class period. For each class, students had already been assigned to lab groups at separate tables with up to four students at a table. A project code was placed on the table for students to use when logging into the WISE interface to create their individual student accounts. The code for each table determined the condition to which the students were assigned.

During the first day students were guided through setting up their accounts on WISE. They were then asked to review the tutorial before completing the pretest. Students were then allowed to play the game at their own pace. For the game, students were instructed to attempt to complete as many of the bronze levels as they could in order to experience as much of the educational content as possible before attempting the silver or gold levels. Students were given minimal help from the experimenters, but students were encouraged to seek help from the person sitting next to them if they were having trouble completing a level. This was done for two reasons: to minimize reliance on the experimenters for help and to encourage students to progress as far as possible in the game. However, the students were discouraged from playing the game jointly. Students continued to play the game for the next three days for approximately 45–50 minutes each day depending on the length of the teacher's announcements and administrative tasks. On the fifth and final day, students played the game for 30 more minutes before they were asked to complete the multiple choice posttest with no time constraints. After students finished the electronic posttest, they were given the paper and pencil short answer test.

5. Results

Means and standard deviations for the pretest, posttest, and gains scores can be found in Table 1. There were no significant differences between the three groups on the pretest, $F(2,83) = .82$, $MSE = 13.64$, $p = .44$. A repeated measures ANOVA examining test (2) × condition (3) found that there was a significant overall effect for testing session, $F(1,83) = 21.97$, $MSE = 113.85$, $p < .001$, $d = .50$, with participants scoring significantly higher on the posttest after playing the game with an average gain of 1.62 points ($SD = 3.27$). Thus students demonstrated significant pre-post gains overall. The repeated measure ANOVA revealed no significant interaction between testing session and condition, $F(2,83) = 2.42$, $MSE = 12.52$, $p = .10$, therefore the condition students were assigned to did not affect overall learning gains.

To examine whether the game condition had an effect on learning for any of the subsets of questions (1st law in 1 dimension, 1st law in 2 dimensions, 2nd law, and 3rd law), a repeated measure ANOVA was conducted examining pre and posttest performance on the four

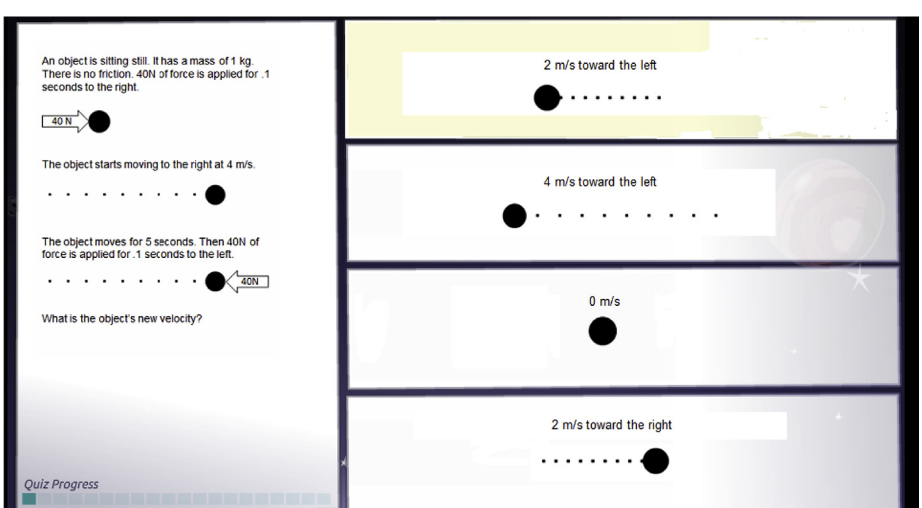

Fig. 2. Sample question from the electronic physics pretest/posttest.

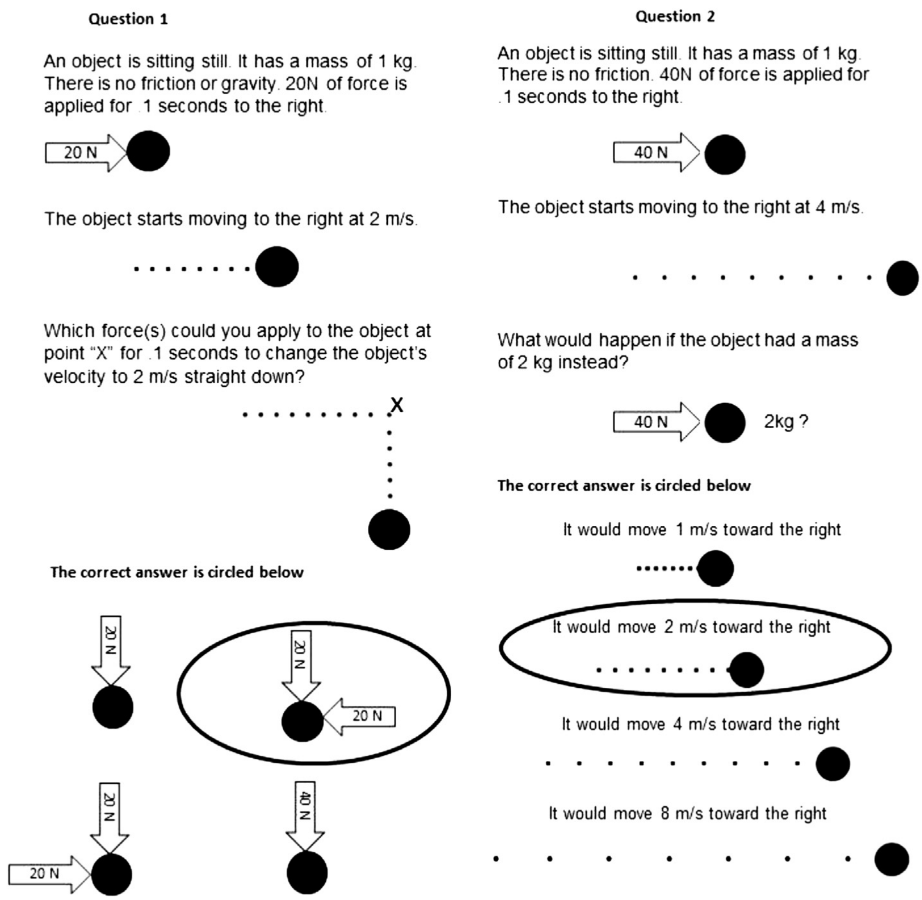

Fig. 3. Short answer paper and pencil test items.

question types for the three conditions. When looking at the main effect for testing session, participants performed significantly better on the posttest for questions dealing with 1st law in one dimension, $F(1,80) = 15.31$, $MSE = 21.38$, $p < .001$, 1st law in two dimensions, $F(1,80) = 13.79$, $MSE = 13.09$, $p < .001$, and 2nd Law, $F(1,80) = 4.88$, $MSE = 5.45$, $p = .03$. However, there was no significant effect of testing session on questions dealing with the 3rd law, $F(1,80) = .04$, $MSE = .03$, $p = .84$. When looking at the interaction between game condition and testing session, the only significant difference was for 2nd law questions, $F(2,80) = 3.19$, $MSE = 7.38$, $p = .05$, with the base-game group demonstrating significantly higher learning gains when compared to the self-explanation group ($p = .04$, $d = .70$). There were no other significant differences between the three groups.

The finding that the self-explanation group demonstrated significantly lower learning gains compared to the base-game group on the posttest on questions dealing with 2nd law is likely explained by examining game progression differences between the three groups. Analyses of the number of levels completed by students across the three conditions showed significant difference between the three groups, $F(2,83) = 4.18$, $MSE = 107.48$, $p = .02$. Bonferroni pair-wise comparisons revealed that the base-game group ($M = 19.50$, $SD = .96$) completed significantly more levels compared to the self-explanation group ($M = 16.03$, $SD = .91$) ($p = .03$). There was also a non-significant trend for the explanatory feedback group ($M = 19.11$, $SD = .98$) to complete more levels than the self-explanation group ($p = .07$). There was no significant difference between the level progression of tips and base-game groups. Therefore, students who were not required to answer physics question prompts completed more levels of the game.

Looking at game level progress broken down by law focus, a chi-squared analysis revealed that significantly fewer participants in the self-explanation group reached levels dealing with launching Fuzzies/3rd law relationships, $X^2(4, N = 86) = 16.57$, $p = .002$, compared to the other two conditions. The majority of participants in the self-explanation group (64.5%) reached the section of levels that focused primarily on the 2nd law relationships (e.g. picking up and dropping Fuzzies), while the majority of participants in the explanatory feedback group (59.3%) and the base-game group (53.6%) reached the section of levels that focused primarily on the 3rd law. The data suggests that participants in the base-game group, who had more experience with the implications of the changes in mass and velocity involved in picking up and launching Fuzzies, were more successful at answer questions dealing with key 2nd law relationships. The fact that fewer students

Table 1
Means and standard deviations for pretest and posttest for participants across the three conditions.

Condition	Pretest	Posttest	Gain score
Base game	8.07 (4.57)	10.79 (5.50)	2.71 (3.31)
Self-explanation	8.55 (3.92)	9.61 (4.65)	1.06 (3.02)
Explanatory feedback	7.19 (3.70)	8.30 (3.78)	1.11 (3.34)
Total	7.97 (4.07)	9.58 (4.76)	1.62 (3.27)

overall reached levels dealing with 3rd law relationships could explain the lack of a significant difference between the pretest and posttest for questions related to the 3rd law of motion.

5.1. Self-explanation engagement and learning differences

Hsu et al. (2012) found that the engagement level of the participants in their self-explanation condition had a significant effect on learning gains. It is therefore possible that within our self-explanation group participants who actively tried to answer the questions correctly may similarly have demonstrated higher performance than those who “gamed the system” by randomly clicking answers until the game indicated that the correct explanation had been selected. For every question participants were provided four possible explanation options. For each incorrect answer choice selected their score decreased by 25%. Scores were then averaged across the pre-flight, in-flight, and post-flight questions to give an average explanation score for each game level. These scores were then averaged over the completed levels for each self-explanation participant. There was a significant positive correlation between pretest and self-explanation score, $r(30) = .59$, $p = .001$, and posttest, $r(30) = .49$, $p = .006$, but there was no significant correlation between learning gain scores and self-explanation scores, $r(30) = -.02$, $p = .91$. The lack of a significant relationship with learning gains suggests that the student’s ability or willingness to answer self-explanation questions may have been more a function of prior knowledge levels rather than information that they acquired through the game. To examine this possibility further, a median split was used to divide students into high and low self-explanation engagement groups. A repeated measures ANOVA looking at performance over the two testing sessions between high and low self-explanation participants showed no significant interaction between testing session and self-explanation engagement level, $F(1,28) = .13$, $MSE = .60$, $p = .73$, suggesting that receiving higher engagement scores was not significantly related with higher learning gains from the game. One possible reason for this result is that because participants in the self-explanation group received feedback after answering the self-explanation prompts, all of the students were exposed to the same correct information. Due to how the self-explanation scores were calculated, participants with higher scores were individuals who were more likely to answer the question prompts correctly on the first or second try. These were mostly likely individuals with higher prior knowledge considering the significant positive correlation between pretest performance and explanation score. Therefore participants with high prior knowledge received higher scores on the self-explanation prompts but did not necessarily learn more from playing the game since all students received the same correct information.

5.2. Gender differences on game performance and learning gains

Examining the possible effects of gender, there was no significant difference in pre-test performance between males and females, $F(1,80) = .19$, $MSE = 3.25$, $p = .66$. For difference scores between the pretest and the posttest, there was no significant main effect for gender $F(2,80) = .77$, $MSE = 8.07$, $p = .38$, nor was there significant interaction between gender and condition, $F(2,80) = .58$, $MSE = 6.09$, $p = .56$. Looking at game progress, there was a significant main effect of gender on number of levels completed, $F(1,80) = 5.02$, $MSE = 115.61$, $p = .03$, $d = .55$, with males ($M = 19.33$, $SD = 5.47$) completing significantly more levels than females ($M = 16.54$, $SD = 4.56$). In addition there was a significant interaction between condition and gender in terms of number of levels completed, $F(2,80) = 3.73$, $MSE = 85.82$, $p = .03$. While participants of both genders completed the fewest number of levels in the self-explanation group, female participants in the explanatory feedback group completed the most levels while males in the base-game group completed the most levels. Means and standard deviations for game progression for the three groups broken down by gender can be found in Table 2.

5.3. Short answer analysis

For the short-answers portion of the posttest, students were scored according to the following criteria: (1) intuitive/game based understanding independent of formal terminology; (2) attempted formal connection (accurate or otherwise) to connect the explanation with one of Newton’s laws (either a number or the definition); (3) accurate formal connection.

The first of the two short-answer questions involved making a 90° turn. The challenging part of this question for students involves the leftward force, which is necessary to stop the rightward motion otherwise the object will travel in a diagonal down and to the right. Answers that were seen as only reflective of an intuitive understanding would include information regarding applying a leftward force while an response that including both and intuitive and a more formal understanding would also mention that an object in motion will remain in motion unless acted upon by an outside force or that the question related to the 1st law of motion.

In terms of whether or not the student’s response included intuitive/game based explanation, there was no significant difference between the three conditions, $X^2(2, N = 84) = 1.35$, $p = .51$. Therefore students across the three conditions were equally as likely to include the necessity of a leftward horizontal force to reduce horizontal velocity to zero before applying a downward force. In terms of whether or not the students made an attempted formal connection between the problem and one of Newton’s laws there was no significant difference in the likelihood that the students would include any information about Newton’s law(correct or incorrect) in their answers, $X^2(2, N = 84) = 4.41$, $p = .11$. However, when examining whether participants made accurate formal connections, there was a significant difference across the three conditions. Participants in the base-game group were more likely to mention the incorrect 3rd law (for every action there is

Table 2
Average level reached by male and female participants across the three conditions.

Condition	Females	Males
Base game	15.83 (4.84)	22.25 (6.17)
Self-explanation	15.44 (4.73)	16.67 (1.99)
Explanatory feedback	19.44 (2.60)	18.94 (5.77)
Total	16.54 (4.56)	19.33 (5.47)

an equal and opposite reaction) compared to the explanatory feedback condition, $X^2(4, N = 84) = 10.81$, $p = .03$, to explain why a leftward horizontal force must be used. There was no significant difference in the likelihood that participants would make an accurate formal connection to the 1st law of motion. This may indicate that the inclusion of explanatory feedback helped to address misconceptions regarding the laws but did not encourage students to necessarily think in terms of the three laws. Only one student in the base-game group included Newton's 2nd relationship information in their response. This response was recorded as correct due to including information in relation to how force affects the acceleration of the ship.

Looking within the self-explanation group for high versus low engagement participants, there was a significant difference in the likelihood that participants' answer for the first question would include needing to use a force to stop the horizontal motion, $X^2(1, N = 29) = 7.81$, $p = .005$, with 78.6% of high engagement students stating that a forced needed to be used to stop the spaceship before making the spaceship move straight down, but there were no significant differences in either attempts to make formal connections or the accuracy of formal connections.

For the second question, dealing with 2nd law relationship, students had to explain how a change in mass from 1 kg to 2 kg would affect the change in velocity when the amount of force was kept constant (See Fig. 3). For this question, there was no significant difference between the three groups for whether students gave an intuitive based explanation describing the relationship between force and mass, $X^2(2, N = 84) = 1.34$, $p = .51$. In addition there was no significant difference between the three condition concerning attempted formal connects, $X^2(2, N = 84) = 1.50$, $p = .47$, or whether students made the accurate formal connections to the 2nd law relationships, $X^2(6, N = 84) = 3.98$, $p = .68$. The majority of student did not mention anything pertaining to Newton's 2nd law (77.4%). Only a few mention anything about the correct law (11.9%), however, only a small number made the error of mentioning either the 1st law (8.3%) or the 3rd law (2.4%). There were no significant differences between the high and low engagement self-explanation student groups on the second short-answer question for any of the short-answer scoring criteria.

6. Discussion

6.1. Findings by condition

The results showed that students in all three conditions demonstrated significant higher performance on the post-test compared to their performance on the pre-test. Between the three groups there was only one significant difference; participants in the base-game condition had significantly larger gains between the pretest and the posttest on questions dealing with Newton's 2nd law. This is the opposite of the predicted outcome, that requiring students to answer self-explanation questions would help connect their game experiences to more formal physics representations. The simplest and most plausible explanation for this outcome is that students who progressed furthest through the game learned the most. Students in the base-game condition were able to complete significantly more levels than the self-explanation condition, and the explanatory feedback condition also showed a similar non-significant trend of completing more levels than students in the self-explanation-condition. The 2nd law relationships between force, mass, and changes in velocity were the focus of levels later in the sequence and reinforced by elements in the levels which introduced picking up and launching Fuzzies (a game mechanic designed to highlight 3rd law relationships). Thus the self-explanation group had less opportunity to explore 2nd law relationships in game. Unfortunately the extra time invested in explaining solutions during earlier levels did not result in increased learning outcomes for 1st law relationships for the self-explanation group in comparison to the other two groups. For the paper posttest, student in the base version of the game were more likely to make the misconception of attributing applying an unbalanced force to stopping an object to the 3rd law of motion although this misunderstanding may not have been detrimental to them while playing the game. There were no significant differences between conditions for whether participants made accurate formal connections with the 1st law of motion. Therefore, while having explanatory tips may have encouraged students to avoid this misconception, it did not necessarily lead to a connection between the intuitive understanding fostered in the game and formal representations of Newton's 1st law.

One possible limitation for the experimental manipulation was that students were allowed to help each other during gameplay. While this may have helped students to progress further in game and limited their interaction with the experimenters, interacting with other students may have decreased the effectiveness of the self-explanation and feedback manipulations. Although a purer experimental manipulation would restrict all outside interactions during gameplay, the experimenters believed that allowing students to seek help from partners at their lab tables (who were all within the same condition) allowed for a more authentic classroom experience. Students were discouraged from playing the game together so that they would attend to the question and feedback prompts. Future studies could limit this interaction to see if the absence of any outside assistance would increase reliance on in-game help features.

6.2. Findings by gender

Previous research has found that using an educational game to teach physics can be more effective for males versus females when using a conceptually-integrated/intrinsic integration game design approach (Echeverría, Barrios, Nussbaum, Améstica, & Leclerc, 2012). The results from the current study showed no significant differences in learning gains between males and females in general or between the three conditions. Males, however, progressed further through the game missions. The interaction between gender and condition on level progression suggests that female players may have benefitted more for feedback in terms of how to play the game. This may have been due to differences in prior game playing experience. Unfortunately no data on previous commercial game experiences was gathered so no conclusions can be made as to interaction between gender and game-playing experience with regard to level progression.

6.3. Findings by engagement

Unlike Hsu et al.'s (2012) findings, which showed learning differences within their self-explanation condition between high and low engagement students, our results showed no differences in learning gains between students with high versus low self-explanation scores.

Although high-scoring self-explanation participants were more likely to mention canceling out the horizontal force on the short answer problem, it is unclear whether or not these differences were due to the intervention or if it was due to the relationship between higher prior knowledge levels and self-explanation scores.

In terms of engagement for the tips conditions, it is unclear whether students took advantage of the information provided at the three different feedback points. Students in the tips condition may have shown higher learning gains if they spent more time reading the hints and using that information to develop more formal based understandings of the relationships in the games. Future research should examine whether time spent reading the explanatory feedback affects gameplay and learning outcomes.

6.4. *The interaction between cognitive load and self-explanation*

Despite the self-explanation group receiving additional information pertaining to physics, why were they unable to complete as many levels as the base game condition? Similar to result from Moreno and Mayer (2005) the interactivity within the game environment may have affected gameplay. Unlike the *Design-A-Plant* studies, where interacting may have encouraged deeper processing, the amount of interaction in *The Fuzzy Chronicles* may have overloaded cognitive processing making students unable to engage reflectively with the self-explanation questions. The basic game design for *The Fuzzy Chronicles* has a large amount of intrinsic processing load (i.e., the amount of cognitive load required to understand the basic material in a lesson). According to Sweller (2010), one of the critical factors in intrinsic load is element interactivity, which is how many elements the learner must process simultaneously in order to complete the learning goal. Moreno and Mayer's (2005) *Design-A-Plant* game only asked students to make a total of 3 selections to construct a plant while *The Fuzzy Chronicles* requires students to place anywhere between 2 and 40 actions. During a typical level *The Fuzzy Chronicles* students must: (a) travel through multiple line segments while adding forces in order to start, stop, and change the speed of the spaceship; (b) add actions such as picking up Fuzzies and applying shields; and (c) ensure their timeline action bar matches their planned route in the game area.

The basic game design also includes a spatial separation between the action bar and where the actual action takes place in the game area. This may also increase cognitive load due to violating the spatial contiguity principle (Mayer, 2009) in which students learn better when corresponding elements are presented near to each other rather than far apart. Students have to make eye movements between the play area and the action bar in order to integrate how the actions placed on the bar correspond to the movements of the spaceship.

Additional cognitive load may have been created by students mentally simulating the actions of their spaceship before pulling the launch lever to test out their solutions using the game. This may have been the key factor that may have caused the progression differences between the three game conditions. During the game students can test out their solution by "pulling" the launch lever on the left hand side of the screen. With the cognitively demanding design of the game, students may have broken down parts of the level into smaller segments instead of focusing on the entire solution. The segmenting principle in multimedia learning is a way that complex material can be broken down into smaller segments in order to manage intrinsic or essential processing load (Mayer, 2009). Segmenting can similarly be used in games in order to break down complicated game mechanics into smaller, more manageable sub-goals. For example, a student could first focus on moving three squares and then focus on creating the correct combination of forces to move along a diagonal path instead of trying to place all of the forces for the two line segments at once. Using the launch lever would allow students test to see whether they had successfully completed each segment separately.

Unfortunately, the design difference between the three conditions may not have encouraged students to equally adopt this strategy. Students in the self-explanation group had to answer questions before launching the simulation, after a failed attempt to complete the level, and upon successful completion of the level. This design mechanic forced students in the self-explanation condition to answer at least one to two questions for every attempt. Questions had to be answered correctly before the game would allow the students to return to the game. In contrast, students in the explanatory feedback version were presented with feedback tips at the same game points. The students could easily press a button to return to gameplay with or without reading the entire explanatory tip. In the base-game group, students only received correct/incorrect feedback from the game for their attempts. Considering the game requirements, the self-explanation group may have been discouraged from making more attempts to avoid having to answer so many questions. The questions interrupted the flow of the solution process therefore incentivizing students to focus more on getting the entire solution correct rather than focusing on natural consecutive segments. Focusing on the entire solution instead of segments could overwhelm student's working memory capacity making it harder for them to find problems within their solution pertaining to particular segments.

Allowing students to simulate their action plans can be considered an example of what Kirsh and Maglio (1994) call an "epistemic action." Epistemic actions are actions in which humans alter their physical environment in order to facilitate cognition. Kirsh (1995) classified epistemic actions into 3 categories according to how they can facilitate cognition: by simplifying choice, simplifying perception, or by simplifying internal computation (Kastens, Liben, & Agrawal, 2008). For example, Kirsh and Maglios' (1994) research has shown that players can offload the cognitively demanding process of mental rotation by using the simple rotation function in the commercial game *Tetris.* Instead of having to visualize what a *Tetris* shape will look like after being rotated by 90° increments, it is faster for the player to press a key and get immediate visual feedback from the game for how the piece appears at all four orientations. Other research has shown that epistemic actions can facilitate learning in scientific areas such as geology (Kastens et al., 2008). Specifically, Kastens et al. (2008) found that both students and geologist would use epistemic actions such as rotating geological models so they aligned with their sketches and juxtaposing two models to help determine which best represented the 3D geological structure. The efficacy of complicated educational games might be enhanced by allowing students to offload parts of the gameplay through epistemic actions (such as allowing students to simulate the action plan in *The Fuzzy Chronicles*).

In future research it will be important to examine the degree to which allowing students to offload processing onto the game, such as segmenting levels into smaller units, might support more effective learning outcomes. Future research could also examine how a student's strategy can affect learning from an educational game. Do students who use more trial-and-error learn less about physics compared to students that plan out his or her entire solution before trying to solve the level? It would also be of interest to see if students who have higher visualization skills do not need to use game mechanics in order to test out their planned trajectories.

One other major difference between *The Fuzzy Chronicles* and games that have benefitted from adding self-explanation and explanatory feedback was the interaction between the interactive demands of the games and the coupling of feedback with players' responses. In the *Design-A-Plant* and the electric circuit games, students made only one to four selections and either received feedback or were asked to explain their choices. This allowed the game interface to closely match explanations and feedback with the player's actions. As mentioned in the previous section, *The Fuzzy Chronicles* interface can require students to place up to 40 different actions in order to solve a level. An error in any of those actions can result in the student being asked a question or receiving explanatory feedback related to one of the laws.

The less constrained game design for *The Fuzzy Chronicles* makes it harder for the system to recognize the exact errors made by the students. If a player's spaceship goes off course, the game may ask the student how using an unbalanced force could change the spaceship's trajectory (1st law of motion), however, it cannot identify the exact reason for why the spaceship went off course (e.g. applying a force too late or not using a downward force to stop upward motion). Research on gaming the system has shown that there is a high correlation between students seeing an electronic tutor as being unhelpful for learning and gaming behavior (Baker et al., 2008). If students did not think that the system was able to identify their particular mistakes this may have led to the belief that the game was unhelpful at providing guidance for how to successfully complete levels. To examine this possibility, future studies could limit the number of decisions each student has to make for each level and couple the questions and feedback specifically to those decisions. This would decrease overall cognitive load and help students integrate the information they are receiving from playing the game with the physics information from the self-explanation questions.

One final difference between previous research which has shown benefits for self-explanation and the current study is when students were asked to engage in self-explanation. Moreno and Mayer's (2005) research points to the importance of having participants explain correct solutions. Johnson and Mayer (2010) also asked participants to explain correct solutions. In Hsu et al.'s (2012) study, self-explanation was only effective for students who correctly explained what mistake they had made during the game. It is unclear from Hsu et al.'s (2012) results whether students would have benefitted from receiving feedback to their self-explanation choices. One issue with the current study is that participants were asked to answered questions after three separate types of games events: (1) before making an attempt which could have been correct or incorrect, (2) after an unsuccessful attempt which was an incorrect solution, and (3) after a successful attempt/correct solution. Future studies should separate the timing of self-explanation questions into separate experimental conditions to determine the impact of reflecting only on successful attempts versus the impact of reflecting on unsuccessful as well as successful attempts.

All of the above issues point to a disruption in the experience of flow while gaming (Csikszentmihalyi, 1991). During a state of flow participants are completely absorbed in an activity. Flow experiences have been defined by characteristics such as concentration, time distortion, and sense of control (Kiili, 2005). Kiili and Laineme (2008) draw a distinction between the experience of flow while playing an educational game and consciously reflecting on the learning material. While flow should encourage behaviors that are spontaneous and automatic, knowledge construction is a conscious and active process. Therefore controlling the game should be automatic and spontaneous while the actual educational content, such as the self-explanation prompts, should require the players to reflect on the material. Both Johnson and Mayer (2010) and Kiili and Laineme (2008) raise the issue that when working memory is overloaded by either extraneous material or the intrinsic difficulty of the gameplay, students are less likely to experience flow. The design of the game must encourage fluid interactions so students can free up their cognitive resources to focus on processing the learning material and problem solving during the game. Having students make an overwhelming number of decisions during a game level, asking questions during the problem solving process, and discouraging offloading cognitive demands through trial and error are all possible reasons why game flow could have been disrupted during *The Fuzzy Chronicles*. Future designs of this particular game must focus on encouraging flow while playing the game so that students have the cognitive resources available to take advantage of the learning prompts.

6.5. Incorporating self-explanation into games: final thoughts and next steps

Evidence from previous research supports the following conclusions in terms of determining when self-explanation is productive to include in education and simulation games support the following conclusions:

1) Self-explanation adds no additional benefit when interactivity with the game is enough to encourage adequate processing to understand the to-be-learned material (Moreno & Mayer, 2005).
2) In order to decrease cognitive load participants should be given self-explanation options to pick from instead of having to generate explanations (Johnson & Mayer, 2010; Mayer & Johnson, 2010). Using options also allows the game to give feedback to correct any misconceptions that students have for their explanations.
3) Participants should likely be asked to reflect only on correct answers (Moreno & Mayer, 2005)

Results from this present study suggest that the following design aspects should be taken into consideration when implementing self-explanation:

4) Self-explanation paired with errors made by the player will not be effective if the prompts are not closely coupled with the error. Students may not see the prompts as informative if there is a mismatch between the questions and the errors the student made.
5) Learner behavior needs to be taken into account when deciding how to use self-explanation prompts. If the game places a lot of cognitive demands on the player using fewer self-explanation prompts may be more effective.
6) If students make a large number of errors while problem solving during the game, the questions could actually be viewed as punishment and discourage students from progressing through the game using techniques such as segmenting, experimenting by altering variables in the level, and trial and error.

All six of these points need further investigation to understand when and how educational game can integrate self-explanation. The present study suggests that self-explanation questions and explanatory feedback are not effective in a conceptually integrated educational game when the overall cognitive load level is high or when the self-explanation functionality disrupts gameplay and becomes "punitive." One beneficial affordance of games is immediate feedback, which encourages experimentation and supports learning. By forcing students to answer questions after unsuccessful trials, the design of self-explanation functionality in the current study seems to have discouraged more exploratory strategies such as segmenting the problem into more manageable chunks to decrease cognitive load. When cognitive load is too high, the learner is unable to engage in deeper cognitive processing in order to integrate what is happening in the game with the educational concepts, then they cannot take advantage of additional information. Game mechanics must manage cognitive load, possibly by allowing them the opportunity to offload processing that is not central to the learning goals.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110782 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

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., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.
Barab, S. A., Sadler, T., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: a framework for socio-scientific inquiry. *Journal of Science Education and Technology, 16*(1), 59–82.
Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S., et al. (2009). Transformational play as a curricular scaffold: using videogames to support science education. *Journal of Science Education and Technology, 18*, 305–320.
Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: how is conceptual change impeded? *Journal of Educational Psychology, 94*, 327–343.
Clark, D. B., & Martinez-Garza, M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuhler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age*. Cambridge: Cambridge University Press.
Clark, D. B., Nelson, B., Chang, H., D'Angelo, C. M., Slack, K., & Martinez-Garza, M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers and Education, 57*(3), 2178–2195.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York: Harper Perennial.
Echeverría, A., Barrios, E., Nussbaum, M., Améstica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: a structured methodology for the design of games for the conceptual understanding of physics. *Computers and Education, 59*, 806–816.
Galas, C. (2006). Why Whyville? *Learning and Leading with Technology, 34*(6), 30–33.
Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169–206.
Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *International Journal of Cognitive Technology, 7*, 4–14.
Hickey, D., Ingram-Goble, A., & Jameson, E. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*(2), 187–208.
Homer, B. D., Hayward, E. O., Frye, J., & Plass, J. L. (2012). Gender and player characteristics of video game play of preadolescents. *Computers in Human Behavior, 28*, 1782–1789.
Hsu, C.-Y., Tsai, C.-C., & Wang, H. Y. (2012). Facilitating third graders' acquisition of scientific concepts through digital game-based learning: the effects of self-explanation principles. *The Asia-Pacific Education Researcher, 21*(1), 71–82.
Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*, 1246–1252.
Kastens, K. A., Liben, L. S., & Agrawal, S. (2008). Epistemic actions in science education. In C. Freksa, N. S. Newcombe, P. Gardenfors, & S. Wolfl (Eds.), *Lecture notes in computer science: Vol. 5248. Spatial cognition VI. Learning, reasoning, and talking about space* (pp. 202–215). Spring-Verlag Berlin Heidelberg.
Kiili, K. (2005). *On educational game design: Building blocks of flow experience*. Tampere, Finland: Tampere University of Technology Press.
Kiili, K., & Lainema, T. (2008). Foundation for measuring engagement in education games. *Journal of Interactive Learning Research, 19*(3), 469–488.
Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73*(1–2), 31–68.
Kirsh, D., & Maglio, P. (1994). On the distinguishing epistemic from pragmatic action. *Cognitive Science, 18*, 513–549.
Klopfer, E., Scheintaub, H., Huang, W., Wendal, D., & Roque, R. (2009). The simulation cycle: combining games, simulations, engineering and science using StarLogo TNG. *E-Learning, 6*(1), 71–96.
Lester, J. C., Stone, B., & Stelling, G. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9*, 1–44.
Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics, 74*(2), 118–122.
Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.
Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). Charlotte, NC: Information Age Publishing, Inc.
Mayer, R. E., & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*(3), 241–265.
McQuiggan, S., Rowe, J., & Lester, J. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. In *Proceedings of the 2008 SIGCHI conference on human factors in computing systems, Florence, Italy* (pp. 1511–1520).
Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: the case for personalized multimedia messages. *Journal of Educational Psychology, 92*, 724–733.
Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*, 165–173.
Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-base multimedia game. *Journal of Educational Psychology, 97*(1), 117–128.
Neulight, N., Kafai, Y. B., Kao, L., Foley, B., & Galas, C. (2007). Children's participation in a virtual epidemic in the science classroom: making connections to natural infectious diseases. *Journal of Science Education and Technology, 16*(1), 47–58.
O'Neil, H. F., & Perez, R. S. (2008). *Computer games and team and individual learning*. Oxford, UK: Elsevier.
Paraskeva, F., Mysirlaki, S., & Papagianni, A. (2010). Multiplayer online games as educational tools: facing new challenges in learning. *Computers and Education, 54*, 498–505.
Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). New York, NY: Cambridge University Press.
Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528.
Sweller, J. (2010). Cognitive load theory: recent theoretical advance. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). New York, NY: Cambridge University Press.
White, B. Y. (1993). ThinkerTools: causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1–100.
White, R., & Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.
Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/a0031311. Advance online publication.