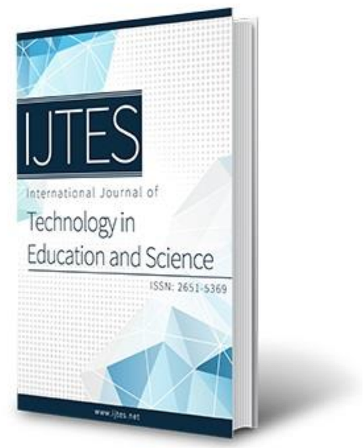

www.ijtes.net

The Effects of Game and Student Characteristics on Persistence in Educational Games: A Hierarchical Linear Modeling Approach

Seyedahmad Rahimi
Florida State University, USA

Valerie Shute
Florida State University, USA

Qian Zhang
Florida State University, USA

To cite this article:

Rahimi, S., Shute, V., & Zhang, Q. (2021). The effects of game and student characteristics on persistence in educational games: A hierarchical linear modeling approach. *International Journal of Technology in Education and Science (IJTES), 5*(2), 141-165. https://doi.org/10.46328/ijtes.118

istes
International Society for Technology, Education and Science

International Journal of Technology in Education and Science (IJTES) is affiliated with
International Society for Technology, Education, and Science (ISTES): www.istes.org

IJTES
International Journal of Technology in Education and Science
2021, Vol. 5, No. 2, 141-165 https://doi.org/10.46328/ijtes.118

The Effects of Game and Student Characteristics on Persistence in Educational Games: A Hierarchical Linear Modeling Approach

Seyedahmad Rahimi, Valerie Shute, Qian Zhang

Article Info

Article History

Received:
22 May 2020
Accepted:
30 August 2020

Keywords

Persistence
Game-based learning
Stealth assessment
Game design elements
21st-century skills

Abstract

Persistence is an important part of student success—both in and out of school. To enhance persistence, we first need to assess it accurately. Digital games can be used as vehicles for measuring and enhancing persistence. The purpose of this study is to test the effects of (a) game-level characteristics (i.e., game mechanics and conceptual difficulty), and (b) student-related characteristics (e.g., students' incoming knowledge and gender) on persistence in a game called *Physics Playground*. The participants in this study were 137, eighth and ninth-grade students from a K-12 school in Florida. We used a Hierarchical Linear Modeling Approach (HLM) to analyze the data. The major findings are (1) the degree of difficulty relating to both the physics concepts and game mechanics of each game problem are significant predictors of persistence, with the former being more effective than the latter in predicting students' persistence, and (2) the number of gold and silver trophies students attained in the game were the only significant student-level predictors of persistence. We conclude by discussing the findings, the implications, limitations, and future research related to this study.

Introduction

A little more persistence, a little more effort, and what seemed hopeless failure may turn to glorious success.

—Elbert Hubbard

Researchers and organizations have emphasized how important persistence is for students' success—in school or generally in life (Duckworth, Peterson, Matthews, & Kelly, 2007; Eisenberger, 1992; Eisenberger & Leonard, 1980; Moore, & Shute, 2017; Partnership for 21st Century Learning, 2015). Our educational systems must prepare students for success in life in the 21st century (Reigeluth & Karnopp, 2013; Rotherham, & Willingham, 2010; Shute, 2007)—e.g., via fostering students' persistence. Therefore, research focusing on assessing and fostering such skills is timely and valuable. One promising area of research involves using digital games to both assess and foster 21st-century skills. However, before we can foster students' 21st-century skills (e.g., persistence), we must first assess them accurately and also examine various game elements that can affect those skills, both positively and negatively. Also, we need to use appropriate tools to enhance these skills. Digital games show great potential for assessing and supporting 21st-century skills (Loup, Serna, Iksal, & George, 2016; Qian & Clark, 2016; Ventura, Shute, & Small, 2014). In this study, we used data collected from an educational game to measure students' persistence, and then, using a hierarchical linear modeling approach investigated various game-related factors (e.g., game difficulty and conceptual difficulty) and student-related factors (e.g., prior knowledge and gaming background) to see if/how they predicted student persistence. The findings of our study can be helpful to educational game designers and game-based learning researchers who want to create educational games that can foster students' 21st-century skills, such as persistence.

Persistence

The everyday definition of persistence involves continuing in a course of action despite difficulty or opposition. Theoretically, persistence may be viewed as a sub-facet of two larger constructs: *conscientiousness* (Shute, & Ventura, 2013), and *grit* (Duckworth et al., 2007). Conscientiousness is one factor in the Five-Factor Model of Personality (McCrae & Costa, 1987), and includes persistence, organization, carefulness, and dependability (Barrick & Mount, 1991). Grit is typically characterized as consisting of both perseverance (i.e., persistence) and passion for achieving long-term goals (Duckworth et al., 2007). Duckworth and colleagues found that grit did not correlate with IQ. Also, while strongly correlating with conscientiousness, grit showed an incremental predictive validity of success in the "accomplishment of widely valued goals" (p. 1087) beyond IQ and conscientiousness. These findings are important because success in life is often attributed to IQ more than other measures of individual differences.

In this study, we use Feather's definition of persistence—the ability to work on hard or unsolvable tasks without any restrictions on time or number of attempts (Feather, 1962). Persistence is being seen as a critical attribute predicting success in various aspects of life, both in and outside of school contexts. However, maintaining persistence when faced with challenges is difficult (Israel-Fishelson & Hershkovitz, 2019), and we as researchers and educators can help enhance students' persistence by developing or using tools such as digital games capable of assessing and fostering students' persistence. Before we can enhance persistence, we first need to assess it accurately (Shute & Wang, 2016). Games appear to be a viable approach toward that end.

Digital Games and Persistence

Digital games are a large part of children's lives, with about 97% of them playing a type of digital game at least an hour a day on average in the United States (Granic, Lobel, & Engels, 2013). Recent research has shown that well-designed digital games can enhance personality traits such as persistence (e.g., Ventura, Shute, & Zhao, 2013). One reason that digital games can be used to enhance persistence is by providing immediate feedback (Gee, 2005; Shute & Ke, 2012). According to Elliot and Dweck (2013), when children receive feedback on various activities based on their efforts (e.g., "you worked hard on this problem") and not based on their existing aptitudes (e.g., "you are smart"), they form a belief system that it is possible to enhance their performance by applying more effort. Games provide immediate feedback relative to players' efforts (Shute & Ke, 2012).

For example, Eisenberger and Leonard (1980) conducted a study to investigate the effect of feedback on 128 college students' persistence in solving an unsolvable task (i.e., a simple game in which students were instructed to find differences between two drawings where the drawings were, in fact, identical but the participants did not know that). The researchers found that continuous failure coupled with feedback indicating the impossibility of solving such a task reduced students' efforts on subsequent tasks, but continuous failure with feedback indicating students' insufficient effort in solving the tasks increased efforts on subsequent tasks. Believing that it is possible to solve a game problem by more effort, students can improve their persistence when facing failures during gameplay until they succeed (Eisenberger & Leonard, 1980; Ventura, & Shute, 2013). Therefore, digital games can be used as ideal vehicles for enhancing students' persistence.

Another reason that digital games can be used to enhance persistence can be the incremental or adaptive challenges existing in well-designed digital games (Shute & Ke, 2012). Such challenges can help players enter and sustain the state of *flow* (Csikszentmihalyi, 1990). When in a flow state, players lose their sense of time and experience deep engagement in their current task. Such a state can keep players engaged with the game for hours without giving up. Another factor enhancing players' engagement when playing video games is interactivity. Research shows that the level of interactivity and game-like interactions in learning environments can increase students' persistence (Croxton, 2014; Kai, Almeda, Baker, Heffernan, & Heffernan, 2018; Sümer, & Aydın, 2018). This amount of engaging time-on-task can be used to assess and support various attributes, including persistence. Other digital learning environments might not be as strong as digital games in inducing and facilitating the state of flow. Thus, digital games can be seen as vehicles to help players incrementally develop their persistence." Next, we discuss the literature on the effects of *game mechanics* and *conceptual understanding difficulty* relative to players' persistence in digital games.

Digital Games Difficulty and Persistence

In psychometrics, there are two kinds of difficulty definitions associated with items or tasks—a *lay* definition and a more rigorous *psychometric* definition (Almond, Kim, Velasquez, & Shute, 2014). In layperson terms, an item or a task is difficult if it takes much effort to solve, and few people can solve it. The psychometric definition of difficulty relates to the requirement of possessing a high level of competence to solve that item or task. Almond and his colleagues (2014) refer to these two types of item difficulty as *the game difficulty* (or task difficulty) and *psychometric difficulty*. Similar to test items, game problems (*note:* we use *game problem* or simple problem throughout the paper to refer to *game level*) can show these two types of difficulties. In this paper, we refer to the two difficulties as *game mechanics* (GM; i.e., how much effort is needed to solve a problem in a given game) and *physics understanding* (PU; i.e., the level of competence required to understand the underlying concept targeted by the problem).

Karumbaiah, Rahimi, Baker, Shute, and D'Mello (2018) investigated the effect of GM and PU (details about these two measures of difficulty are provided in the Method section) in the game *Physics Playground* (*PP*; Shute, Almond, & Rahimi, 2019) and found that PU was more predictive than GM regarding students' frustration—which typically led to quitting the problem— ($p < .05$, Cohen's $d = .96$). GM was not a significant predictor of frustration. Investigating the effect of these two difficulty indices on *persistence* (the aim of the current study) can shed light on how the difficulty of game problems can affect students' persistence. Persistent behavior eventually can affect performance in the game (i.e., solving more problems) and learning (i.e., better game performance can lead to higher learning gains).

Eisenberger and Leonard (1980) investigated the effects of task difficulty (i.e., *game* difficulty) on college students' persistence in an experiment. In that experiment, 192 college students completed a training stimuli session (i.e., exposure to different task difficulty levels and getting mentally prepared for the post-test) and then completed a post-test. During the training session, students were divided into four groups (48 students per group): *unsolvable* who received very difficult tasks, *complex* who received tasks with medium difficulty, *easy* who received easy tasks, and *control* who just were told to read the prompts of each task without trying to solve them. After the training session, all four groups were given six pairs of side-by-side drawings in the post-test. Each pair had six differences in the drawing (students did not know how many differences the pairs of pictures had), and they had to find all differences in each of the six pairs. Students had 120 seconds per pair (12 minutes in total). Students received one point for each correct difference they discovered. Results showed that the students in the unsolvable group showed greater persistence than the complex group (t (184) = 2.34, $p < .05$), who in turn showed greater persistence than the easy group (t (184) = 3.45, $p < .001$). There were no significant differences in persistence between the easy and control groups (t (184) = 1.62, $p > .10$). Eisenberger and Leonard (1980) concluded that initial failure (due to the conceptual task difficulty or psychometric difficulty) could increase the effort and consequently, the persistence of subjects on a subsequent task.

In a recent study (n = 2,040 first to sixth graders), Israel-Fishelson and Hershkovitz (2019) investigated the relationship between game difficulty and persistence in an online game called *CodeMonkeyTM* which is a puzzle game for developing computational thinking skills. This study's main findings indicated that persistence positively associates with game problem difficulty and that the most determined learners were highly persistent across topics in achieving the best solution. Israel-Fishelson and Hershkovitz, however, did not break down the difficulty of game problems into conceptual understanding and game mechanics. The results of this study and the other studies we included above suggest a more detailed investigation of the relationship between game-related characteristics (especially game difficulty) and students' persistence. In this study, we split game difficulty into these two components (i.e., GM and PU) for further investigations.

Students Characteristics and Persistence

Various student characteristics can also impact persistence, either positively or negatively. For example, confidence (i.e., the expectancy of success when learning a new concept or dealing with a new task) can affect students' persistence (Keller, 1987). Students with a more extended history of gaming are likely to have higher confidence in a game-based learning environment and show more persistence when faced with challenges in games than students with minimal gaming backgrounds and with lower confidence levels. Also, high incoming knowledge can be seen as a source of confidence affecting both game performance and persistence (Shute et al., 2015). Similarly, Jackson, Gardner, and Sullivan (1993) found that the best predictor of engineering students' persistence (i.e., continuing in engineering) was the students' average GPA. That is, students with more knowledge and skills received higher GPAs and consequently could persist in engineering more than students with lower GPAs. Therefore, it is essential to take these two variables (i.e., gaming backgrounds and prior knowledge) into account when investigating persistence predictors in game-based learning.

Another essential variable to consider when studying persistence in playing educational games is gender. In general, research shows that girls show more persistence than boys in school (Perez-Felkner, McDonald, & Schneider, 2014). However, girls have a higher tendency to attribute their failure to their level of confidence in their ability than the difficulty of the task at hand than boys (Kiefer & Shih, 2006). This tendency to attribute failure to self-ability for girls (which is rooted mostly in gender stereotypes) results in less desire to show persistence when faced with a difficult task (e.g., solving a math problem) (Kiefer & Shih, 2006). When it comes to digital games, boys are more likely to be or identify as gamers than girls—again, rooted mostly in gender stereotypes and what the male-dominant digital games market dictates (Andrews, 2008; Jenson & de Castell, 2011; Richard, 2013). This fact can exacerbate the effect of stereotypes for girls leading to less desire to show persistence in the digital gameplay context. However, Jenson and de Castell (2011) found that these gender differences in gameplay can be diminished once girls are "afforded genuine access, support, a "girlsgamer" model, and the right to choose what, when, and with whom they would play" (p. 175). The current study's findings can shed light on gender differences and persistence in the context of a STEM-related digital game.

How to Measure Persistence?

It is hard to measure constructs like persistence (Almond, Kim, Velasquez, & Shute, 2014). We can no longer rely only on self-report measures (see Ryans, 1939, for a review on traditional self-report measures of persistence). With new technologies and advances in the learning sciences (Shute, Leighton, Jang, & Chu, 2016), it is possible to design innovative assessments to assess hard-to-measure constructs like persistence accurately. For example, *stealth assessment* (Shute, 2011) uses technology-rich environments (e.g., digital games) as vehicles for assessment. The stealth assessment models are designed using the evidence-centered design assessment framework (ECD; Almond, Mislevy, Steinberg, Yan, & Williamson, 2015; Mislevy, Steinberg, & Almond, 2003), which are then embedded into a digital game. Ongoing performance data are collected in log files as the learner interacts with the game (e.g., the amount of time a player spends working on various game problems in the game as a measure of persistence). The stealth assessment automatically scores and accumulates the collected data using statistical methods (e.g., Bayes nets), making real-time inferences about the learner's current level of the targeted competencies (e.g., persistence).

For example, Ventura and Shute (2013), used log data in a digital 2-dimensional game about physics, then called *Newton's Playground* (*NP*) to measure persistence. They collected data on the time students spent on unsolved, as well as solved-but-difficult game problems as an in-game measure of persistence. Results showed that the in-game measure of persistence was validated as it correlated ($r = .22, p < .05$) with an external performance-based measure of persistence (see Ventura, Shute, & Zhao, 2013 for more details about the external performance-based measure of persistence). In another study, Ventura, Shute, and Small (2014) used a similar measure of persistence in the same game and investigated how persistence relates to learning. Results showed that the in-game assessment of persistence relates to the posttest scores of low performers controlling for gender, video game experience, pretest, and game enjoyment ($r = 0.26$, $p < .05$). We used the same game, which is now called *Physics Playground*, and used stealth assessment to measure persistence in this study.

In another study, DiCerbo (2014) used stealth assessment of persistence in a virtual world called *Poptropica* with different virtual islands where players can explore and fulfill various quests. In this study, time spent on difficult quests (unsolved) and the total number of tasks completed per quest from the log files were used to assess persistence. Results from a confirmatory factor analysis showed that both of these two indicators were significantly loaded on one factor—persistence. In general, the previous research on stealth assessment of persistence shows that persistence can be operationalized using various time indicators (e.g., time spent on unsolved and solved but difficult tasks/problems).

Physics Playground

Physics Playground (*PP*; Shute, Almond, & Rahimi, 2019) is a problem-based, 2-dimensional game where each problem's primary goal is hitting a red balloon with a green ball. For example, Figure 1 shows a game problem called "Little Mermaid." The player has drawn a springboard (in red) and attached a weight (in blue) to shoot the ball up and hit the balloon by deleting the weight. In the game, the player draws objects to create various simple machines or "agents of force and motion" (e.g., lever, ramp, pendulum, springboard) to solve a problem.

Figure 1. The Springboard Solution for the Problem *Little Mermaid*

In the version of the game used in the current study, there were seven playgrounds, each with 10 to 11 problems for 74 problems, arrayed from easy to difficult. The difficulty of a game problem was determined based on various factors including the relative location of the ball and balloon (i.e., if the ball was placed below the balloon, the problem is more difficult than when the ball is placed above the balloon), the number of obstacles between the ball and balloon, and the number of simple machines required to solve the problem. The version of *PP* used in this study was nonlinear. That is, students could freely choose the next problem they wanted to play and revisit any game problem they played. However, *PP*'s problems are presented to the students in a linear order; thus, students tend to follow the game problems in a linear order (Kim, & Shute, 2015). Across various studies (e.g., Kim & Shute, 2015; Shute et al., 2015; Shute & Ventura, 2013; Wang, 2017), the game has been shown to support students' acquisition of physics concepts via gameplay. Across various studies (e.g., Kim & Shute, 2015; Shute et al., 2015; Shute & Ventura, 2013; Wang, 2017), the game has been shown to support students' acquisition of physics concepts via gameplay.

Students could see their progress in the game by the silver and gold trophies displayed on the screen's upper-left part. In this version of *PP*, students could earn a gold trophy for a solution solved efficiently—i.e., under a certain number of objects used which varied based on the problem's difficulty (i.e., the easier the problem, the smaller the number of objects one could use to get a gold trophy). A silver trophy was given to any solution which did not qualify for a gold trophy. These trophies could motivate students to come up with elegant and efficient solutions. Students who got silver trophies for some simple machines per problem could replay that problem to turn their silver trophies to gold. Students could use the dashboard on the top left corner of the screen to see how many trophies they collected per simple machine. Also, when they navigated through the game, they could see the trophies they received per problem. In general, the feedback students received in this version of *PP* were game-related (i.e., the number of problems solved, and gold or silver trophy received for a solution).

Purpose of the Study and Research Questions

The purpose of the present study is to re-examine an existing dataset by Shute and colleagues (2015) to test the effects of (a) *game-level* characteristics (i.e., game mechanics difficulty, physics understanding difficulty, the simple machine used in the problem's solution, and the primary physics concept linked to the problem), and (b) *student-related* characteristics (e.g., incoming physics knowledge, performance in the game, gaming background) on student persistence in *PP*. This study is different from the previous studies that used this dataset in four ways: (1) the outcome variable in the previous study was *learning* and in this study is *persistence*; (2) we are asking different research questions focusing on game-level and student-level characteristics predicting students' persistence in *PP*; (3) the statistical method used in this study is different (i.e., multi-level modeling) than the previous study; and (4) we are not using the field observations of students affective states which were collected in the prior study. Specifically, in this study, we aimed to address the following research question:

> *What game-related variables (e.g., game-problem difficulty, simple machine used,* primary *physics concept of the problem), and student-related variables (e.g., gender, age, gaming frequency, or students' progress in the game) can influence student persistence in an educational game?*

Hypotheses

Related to game-related variables, we hypothesized that game difficulty indices (GM and PU) would have a positive relationship with students' persistence. We expected to see a stronger positive relationship between PU and persistence than GM and persistence—as Karumbaiah and colleagues (2018) detected a stronger relationship between PU than GM and frustration, we expect to see a similar pattern here. The other game-related variables also can be influential in how students persist in the game. That is, we hypothesize to see more persistence on game problems with hard-to-implement solutions and difficult primary concepts (note that these two features are related to GM and PU).

Regarding student-related variables, we expect to see a positive relationship between students' gaming frequency and incoming knowledge, and persistence—as more skilled and knowledgeable students are more confident and can persist more when faced with challenges than less skilled students. We expect to see a significant and negative relationship between progress in the game (number of gold and silver coins) and persistence. The reason can be that skilled students can successfully play through the problems, and the problems might seem easy for them; therefore, they would not show or get a chance to show persistence. Also, we hypothesize that boys will show greater persistence than girls as based on the literature in a STEM-related digital games context, girls might show less desire to show persistence.

Method

Participants

The participants included 137 students (8th and 9th graders; 80 female and 57 male) from a K-12 school in Florida. The target population for playing *PP* is students between 7th to 10th grade. Students in the 8th and 9th grades were selected for this study because of the alignment of the *PP*'s content to the Next Generation Sunshine State Standards relating to Newtonian physics, at those grade levels. Each student was given a $25 gift card upon the completion of the study.

Research Design and Procedure

This study used a single-group, pretest-posttest correlational research design. All students played the same version of the game, in a computer lab with thirty computers, across four days, for about 2.5-3 hours of testing. On Day 1, students completed a pretest (i.e., a qualitative physics test) for 15 minutes and a background questionnaire (with gaming frequency questions) and a demographic questionnaire—all administered online. Then, students completed a performance-based online test of persistence (described in the next section). Finally, after introducing the game and some instructions by the researchers, the students started playing. After finishing the tutorials in playground 1, students could play any problem in any playground. On Days 2 and 3, students continued playing the game—approximately for an hour per day. Finally, on Day 4, students played the game and completed the post-test and a questionnaire about the game.

Measures

Physics Understanding

Two matched isomorphic forms (Form A and Form B—counterbalanced between pretest and posttest; each with 16 items) were used to measure students' physics understanding. These forms included pictorial multiple choice-type questions (see Figure 2). The Cronbach's α for Form A was .72, and for Form B it was .73 (Kim &

Shute, 2015; Shute & Ventura, 2013; Wang, 2017).

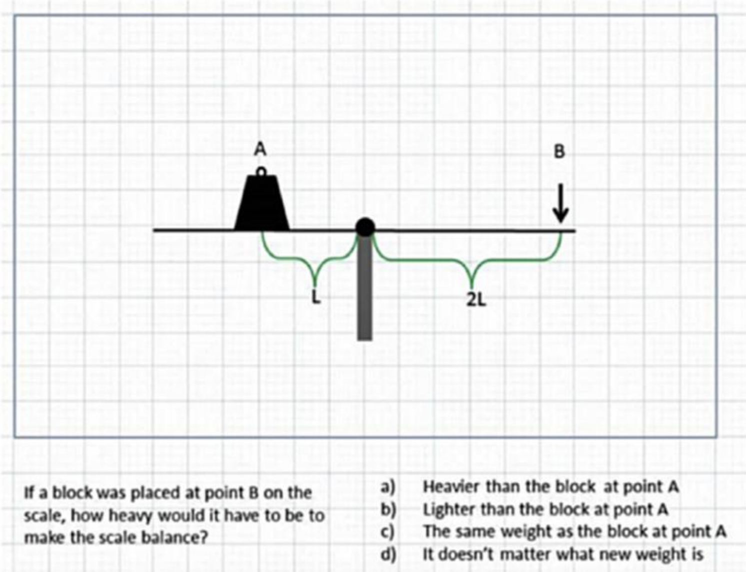

Figure 2. An Example Item of Physics Understanding (from Shute et al., 2015)

The In-game Measure of Students' Progress

PP keeps track of the events which happen during gameplay (e.g., game problems played, trophies collected, timestamps of various events, objects drawn) in the log files. We parsed the log files to generate tallies of gold trophies (i.e., when the solution was elegant—under a specific number of objects drawn) and silver trophies given for any solution not qualify to get a gold trophy. The trophies were recorded based on the simple machines used (e.g., lever trophies, springboard trophies). Some problems could be solved by more than one simple machine; hence, one could get more than one trophy per problem. We summed up all the gold trophies and silver trophies achieved for each student for this study.

The In-game Measure of Students' Persistence

We collected in-game data on time spent playing a problem in two ways. First, we examined time spent on problems that were ultimately unsolved. If a problem was not solved, the trophy of the problem in the log file was set as "none." Second, we examined time spent playing solved problems. In this case, we collected data on times above the average time it took the other students to solve that particular problem. We first calculated the average time spent per problem to collect this data, then selected problems whose solution time was greater than the average. Overall, we could include 3,803 data points (from 137 students playing different problems) in our dataset.

Data Analysis

The nature of the data collected from *PP* is nested. That is, each student plays multiple problems—i.e., times spent playing problems are nested within each student's gameplay (see Table 1). Therefore, using Hierarchical Linear Modeling (HLM; Raudenbush & Bryk 2002) is warranted, where level-1 units are the game problems and level-2 units are the students.

Table 1. Nested Structure of Students' Gameplay in *PP*

Student	Game Problem	Game Problem's Time
Student 1	Game Problem 1	Time 1
Student 1	…	…
Student 1	Game Problem *k*	Time *m*
…	…	…
Student *i*	Game Problem 1	Time 1
Student *i*	…	…
Student *i*	Game Problem *k*	Time *m*

The outcome variable in this study is the in-game measure of student persistence (or PERS for short). Using RStudio (Version 1.0.136; RStudio Team, 2016), SPSS (Version 25; IBM Corp, 2017), and HLM7 (Garson, 2013), we created the level-1 dataset with the following variables:

Game-Related Variable (Level 1)

- Student ID: Each student had a unique ID that was matched to the level-2 dataset later in the analyses using HLM7.
- Persistence in minutes (PERS): The time on the log data was recorded in milliseconds. We computed a new variable by dividing that time by 60,000 and created this variable in minutes.
- Game mechanics (GM) difficulty: Game mechanics (GM) difficulty: Two physics experts consulted with the team to develop a rubric to rate each game problem in terms of GM difficulty. This rubric included ball position (above or below the balloon; when the ball is above the balloon gravity can help make the game problem easier to solve, whereas when the ball is below the balloon, the player needs to defy gravity to hit the balloon making the game problem more difficult), whether the name of the problem gave a hint (e.g., "Timing is Everything" or "Need Fulcrum"), the number of obstacles between the ball and the balloon, whether the player needs to be precise to solve the problem, and the number of sub-goals required to solve the problem (e.g., when the player needs to move an obstacle out of the way and then hit the balloon using a pendulum). The sum of all of the mentioned indices became the GM score for each problem. The GM score could range from 0 (very easy) to 5 (very hard). Two raters scored each problem independently and resolved any disagreements.
- Physics understanding (PU) difficulty: Again, the physics experts worked with the *PP* research team to identify the competency model of physics understanding in *PP* (see Figure 3). This comprised the primary and secondary physics concepts for each problem. Finally, based on the pedagogical difficulty of the physics concepts, we developed the Physics Understanding difficulty (PU) rubric. PU consists of the conceptual order of the primary concept of the problem (force and motion = 0, momentum and energy =

1, torque = 2), the need to either balance forces (i.e., equilibrium or Newton's third law = 1) or conservation of energy (i.e., energy can transfer or conservation of momentum = 1), and the primary and secondary concepts are subtopics of same parent topic (e.g., Newton's first and second laws = 0) or from a different parent topic (e.g., Newton's first law of motion and energy can transfer = 1). The sum of all of the mentioned indices became the PU score for each problem. PU could range from 0 (very easy) to 5 (very hard). Again, two raters scored each problem independently and resolved disagreements.

Note: Only GM was used to determine the order of the problems. We included PU for the analyses in this study.

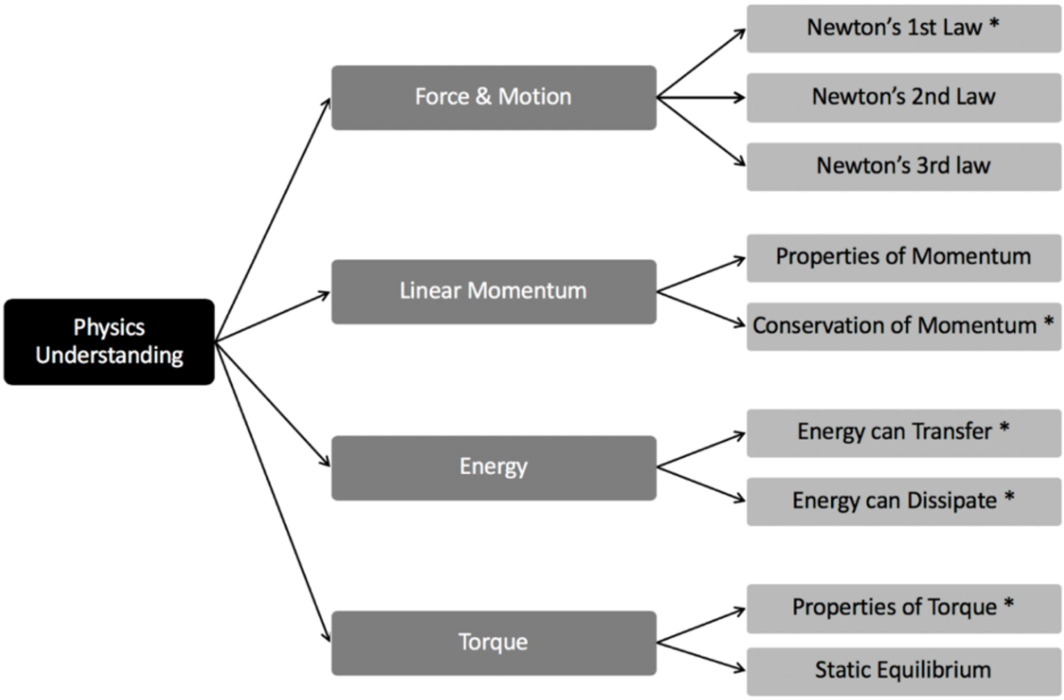

Figure 3. Physics Understanding Competency Model used in this Study

Note. * included concepts in this study.

- The physics simple machines used: *PP* can identify what physics simple machines (i.e., ramp, lever, springboard, or pendulum) the player used. When the problem is not solved, the physics simple machine used is recorded as "none." Therefore, we had five categories for this categorical variable: NONE, RAMP, springboard (SB), pendulum (PEN), and LEVER. We created four dummy variables with LEVER as the reference group because lever was a simple machine targeting two concepts of energy can transfer (ECT) and properties of torque (POT).
- The primary concept of each game problem: our physics experts identified the primary and secondary physics concepts for each problem. The game problems included in this study (n = 74) had the following primary physics concepts: Newton's first law of motion (NFL), conservation of momentum (COM), energy can transfer (ECT), energy can dissipate (ECD), and properties of torque (POT). We created four dummy variables (NFL, COM, ECD, and POT) with ECT as the reference group because ECT was the most frequent concept among all the problems in our study.

Students-Related Variables (Level 2):

We first aggregated the level-1 dataset. That is, the average time spent on the selected game problems in the level-1 dataset—i.e., persistence means—was calculated for each student. The initial level-2 dataset had only a student ID, and the persistence mean for each student. Next, we added other level-2 variables to this dataset. The level-2 variables included:

- Student ID: to match level-1 with level-2 datasets.
- Gender: Male = 0, Female = 1.
- Gaming frequency (Game BG): a scale to measure students' gaming frequency per day from 0 to 6 times per day.
- Pretest: students' physics understanding pretest score.
- Posttest: students' physics understanding posttest score.
- Golds: the total number of gold trophies collected by students during the game.
- Silvers: the total number of silver trophies collected by students during the game.

Results

Descriptive Statistics

We first examined the descriptive statistics of the variables. We aggregated the variables on level-1 to generate descriptive statistics for a typical person (Tables 2 and 3). We also included other student variables such as gaming frequency, pretest, posttest, golds, and silvers in Table 1. A total of 3,803 observations across all the students' gameplay data were included in the level-1 dataset, and 137 cases were included in the level-2 dataset.

Table 2. Descriptive Statistics of the Level-2 Variables (n = 137).

Variables	*Mean*	*SD*	*Min.*	*Max.*
Persistence	3.56	1.35	1.02	6.61
Game Mechanics	2.27	.27	1.00	2.88
Physics Understanding	2.64	.26	1.00	3.23
Gaming Frequency	2.20	1.68	.00	6.00
Pretest	7.11	2.04	2.03	12.00
Posttest	7.52	1.94	3.24	12.04
Golds	8.93	4.87	1.00	34.00
Silvers	23.17	7.94	4.00	51.00

In multi-level modeling, having a sufficient sample size at level-2 (in this case, students) is important for obtaining accurate parameter estimates (Maas & Hox, 2005). Mass and Hox suggested a sample size of at least 50 at level-2. Hence, our sample size of 137 students at level-2 was sufficient.

Table 3. Descriptive Statistics of the Categorical Variables (n = 137)

Simple Machines Used		
Variables	*% (on level-2)*	*Frequencies (on level-1)*
NONE	46	2010
RAMP	17	557
SB	12	403
PEN	10	331
LEVER	15	502
Physics Concepts of Problems		
ECT	43	1610
POT	22	887
NFL	31	1173
COM	3	102
ECD	1	31

Note. NONE = no simple machine was used and the problem was unsolved, RAMP = the problem was solved using a ramp solution, SB = springboard, PEN = pendulum, ECT = energy can transfer, POT = properties of torque, NFL = Newton's first law, COM = conservation of momentum, ECD = energy can dissipate.

Overall, students spent about three minutes on each problem in the game. The game problems included in this study had a medium average difficulty ($\overline{GM}$ = 2.27, SD_{GM} = .27, and $\overline{PU}$ = 2.64, SD_{PU} = .26). The distribution of problems for the primary concept is: ECT 43%, POT 22%, NFL 31%, COM 3%, and ECD 1%. That is, for example, 43% of the data comes from ECT problems. A similar interpretation can be made for the simple machine used. That is, 46% of the data come from unsolved problems, and collectively, 54% of the data come from solved problems. Finally, students reported 2.19 (*SD* = 1.68) times of playing digital games per day.

The data of our persistence measure was positively skewed (Skewness = 3.07, Kurtosis = 18.25). To fulfill the normality assumption, we computed a log transformation (e.g., Benoit, 2011) of the persistence data for applying multilevel models (see Figure 4).

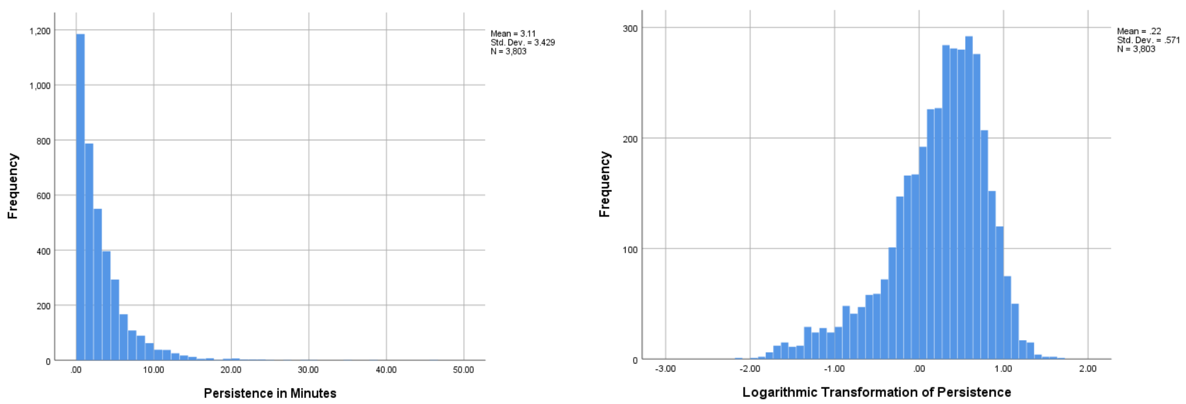

Figure 4. Distribution of Persistence across all the Game Problems and Individuals (left), and Logarithmic Transformation of Persistence (right) (n = 3,803)

HLM Results

Fully Unconditional Model

To examine the variability of the data within each student's gameplay and between the students, we first ran a fully unconditional model without level-1 and level-2 predictors:

$$Log(PERS)_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

where $PERS_{ij}$ is the outcome of interest (persistence) of student *j* in game problems *i*; β_{0j} is the mean of persistence score for student *j* across game problems; r_{ij} is the level-1 residual whose variance depicts within-students variability of persistence scores; γ_{00} is the average of students' persistence across all students; and u_{0j} is the level-2 residual whose variance depicts between-students variability of persistence scores.

Based on the fully unconditional model, we computed an intra-class correlation (ICC) to determine the proportion of between-student variation. The higher the ICC, the more variation of the persistence measure comes from level-2—from individual differences in students' gameplay. In this case, ignoring data dependencies and using regression analysis can yield misleading results (Raudenbush & Bryk, 2002). The ICC for this data is .13, and the between-students variance of individual means is statistically significant (χ^2 (136) = 749.11, $p < .001$). Thus, between-students variability cannot be ignored and a multilevel modeling approach for data analysis is warranted.

Leve-1 (Game Problems) Predictors Added

Next, we entered all level-1 predictors to the model and after removing those with nonsignificant effects, we specified the following model:

$$Log(PERS)_{ij} = \beta_{0j} + \beta_{1j}\, GM_{ij} + \beta_{2j}\, PU_{ij} + \beta_{3j}\, NONE_{ij} + \beta_{4j}\, RAMP_{ij} + \beta_{5j}\, SB_{ij} + \beta_{6j}\, PEN_{ij}$$
$$+ \beta_{7j}\, POT_{ij} + \beta_{8j}\, NFL_{ij} + \beta_{9j}\, COM_{ij} + \beta_{10j}\, ECD_{ij} + r_{ij}$$

where $PERS_{ij}$ is the outcome of interest (persistence) of problem *i* played by student *j*; GM_{ij} is the game mechanics difficulty of problem *i* played by student *j*; PU_{ij} is the physics understanding difficulty of problem *i* played by student *j*; $NONE_{ij}$, $RAMP_{ij}$, SB_{ij}, and PEN_{ij} are the dummy variables for the simple machine used in solving problem *i* by student *j* with *LEVER* as the reference group; and POT_{ij} , NFL_{ij} , CM_{ij} , and NFL_{ij} are the dummy variables for the primary concept of problem *i* played by student *j* with *ECT* as the reference group. β_{0j} is the intercept estimated for each student's persistence; β_{1j} the expected change of persistence score of student *j* with a one-unit increase in GM_{ij} controlling for all other predictors; β_{2j} the expected change of persistence score of student *j* with a one-unit increase in PU_{ij} controlling for all other predictors; β_{3j}, β_{4j}, β_{5j}, and β_{6j} are the mean differences of the persistence of student *j* playing problems that were not solved, solved by a ramp, a springboard, or a pendulum and the persistence mean of student *j* playing problems that were solved by a lever (the reference group) controlling for all other predictors; and β_{7j}, β_{8j}, β_{9j}, and β_{10j} are the mean differences of the persistence for student *j* comparing game problems whose primary physics concept was properties of torque (POT), Newton's first law (NFL), conservation of momentum (COM), or energy can dissipate (ECD) with those game problems whose primary physics concept was energy can transfer (ECT) (the reference group) controlling for all other predictors. Finally, r_{ij} is the level-1 residual.

The averages of β_{0j}-β_{10j} are called *fixed effects*, which depict the average effects across individuals. Note that we first included residuals for all random effects (or individualized effects, β_{0j}-β_{10j}) in the model, allowing them to be different across individuals. However, we found that none of the variances for these residuals were significant except for that of the intercept. That is, the effects of those level-1 predictors on persistence can be regarded as the same across students. Therefore, we excluded the residual terms for β_{1j}-β_{10j} and only kept the residual for β_{0j} (denoted by u_{0j}) in the model.

Exploratory Analysis of Level-2 Predictors

To find the level-2 predictors with potentially significant predictive power on the random intercept, we ran an exploratory analysis in HLM7. Specifically, we selected the level-2 predictors that were significantly correlated with the residuals of the random intercept at level-1. The selected level-2 predictors based on the exploratory analysis are gender (GENDER), gold trophies (GOLDS), and silver trophies (SILVERS) for β_{0j}. When we included GENDER in the model, there was not significant difference between boys and girls for β_{0j}. Hence, the final level-2 model with removing the nonsignificant predictors is:

$$
\begin{aligned}
\beta_{0j} &= \hat{\gamma}_{00} + \hat{\gamma}_{01}\, GOLDS_j + \hat{\gamma}_{02}\, SILVERS_j + u_{0j} \\
\beta_{1j} &= \hat{\gamma}_{10} \\
\beta_{2j} &= \hat{\gamma}_{20} \\
\beta_{3j} &= \hat{\gamma}_{30} \\
\beta_{4j} &= \hat{\gamma}_{40} \\
\beta_{5j} &= \hat{\gamma}_{50} \\
\beta_{6j} &= \hat{\gamma}_{60} \\
\beta_{7j} &= \hat{\gamma}_{70} \\
\beta_{8j} &= \hat{\gamma}_{80} \\
\beta_{9j} &= \hat{\gamma}_{90} \\
\beta_{10j} &= \hat{\gamma}_{10\,0}
\end{aligned}
$$

$GOLDS_j$ and $SILVERS_j$ are the student-level (level-2) predictors that were found to be useful for explaining the variability in the intercept β_{0j}. Results are shown in Table 4.

Table 4. The Final Estimate of Fixed Effects

Fixed Effect	Raw Coef. (SE)	% Change	Stnd. Coef.	t-ratio (*d.f.*)
For Intercept 1, β_0				
Intercept 2, $\hat{\gamma}_{00}$	.48 (.07)	61.61		6.73 (134) **
GOLDS, $\hat{\gamma}_{01}$	-.01 (.002)	-1.00	-.09	-4.13 (134) **
SILVERS, $\hat{\gamma}_{02}$	-.004 (.001)	-.40	-.06	-2.40 (134) *
Difficulty Indices				
For GM slope, β_1				
Intercept 2, $\hat{\gamma}_{10}$	.03 (.01)	3.05	.04	2.27 (3,628) *
For PU slope, β_2				
Intercept 2, $\hat{\gamma}_{20}$	.10 (.01)	10.52	.19	10.39 (3,628) **
Simple Machine Used for a Solution				
For NONE slope, β_3				
Intercept 2, $\hat{\gamma}_{30}$	-.56 (.03)	-42.88		-22.22 (3,628) **
For RAMP slope, β_4				
Intercept 2, $\hat{\gamma}_{40}$	-.18 (.03)	-16.47		-6.43 (3,628) **
For SB slope, β_5				
Intercept 2, $\hat{\gamma}_{50}$	.05 (.02)	5.13		2.16 (3,628) *
For PEN slope, β_6				
Intercept 2, $\hat{\gamma}_{60}$	-.06 (.03)	-5.82		-2.30 (3,628) *
Primary Concept of a Problem				
For POT slope, β_7				
Intercept 2, $\hat{\gamma}_{70}$	-.07 (.02)	-6.76		-2.79 (3,628) *
For NFL slope, β_8				
Intercept 2, $\hat{\gamma}_{80}$	-.02 (.03)	-1.98		-.85 (3,628)
For COM slope, β_9				
Intercept 2, $\hat{\gamma}_{90}$	-.33 (.05)	-27.39		-6.69 (3,628) **
For ECD slope, β_{10}				
Intercept 2, $\hat{\gamma}_{10\,0}$	.29 (.09)	33.64		3.27 (3,628) *

Note. * $p < .05$, ** $p < .001$. SE = standard error, GM = game mechanics difficulty, PU = physics understanding difficulty, NONE = the problem was unsolved, RAMP = ramp was used, SB = springboard, PEN = pendulum, ECT = energy can transfer, POT = properties of torque, NFL = Newton's first law, COM = conservation of momentum, ECD = energy can dissipate.

Interpretation of the Effects

Note that the outcome variable has been log-transformed, and thus was not on the original metric. To meaningfully interpret the obtained results, we used the following formula (van Garderen & Shah, 2002) to show the effects of the predictors in terms of the change in percentages of the outcome:

$$p_j = 100 \times \left(\exp\ \{c_j\} - 1\right)$$

where p_j is the percentage change (shown as "% Change" in Table 4) in the predicted outcome by one unit change in the predictor X_j, and c_j is the raw coefficient of the predictor X_j (shown as "Raw Coef." in Table 4). Finally, to compare the coefficients for the continuous predictors (i.e., GOLDS, SILVERS, GM, and PU), we used the following formula (Hox, Moerbeek, Schoot, Moerbeek, & Schoot, 2017) to standardize the coefficients (shown as "Stnd. Coef." in Table 4):

$$standardized\ coef. = \frac{unstandardized\ coef. \times\ standard\ deviation\ of\ explanatory\ var.}{standard\ deviation\ of\ outcome\ var.}$$

where $unstandardized\ coef.$ is the change in the predicted outcome variable (shown as "Raw Coef." in Table 4), and the exploratory predictors are GOLDS, SILVER, GM, and PU. As shown in Table 4, the effect of GOLDS (i.e., the number of gold trophies student *j* earns in the game) on the random intercept was statistically significant: $\hat{\gamma}_{01}$ = -.01 (*t* (134) = -4.13, *p* < .001). This finding means that by one unit increase in GOLDS, the intercept is expected to decrease by 1% on average controlling for all other predictors. Also, the effect of SILVERS (i.e., the overall number of silver trophies collected by the student j in the game) on the random intercept was statistically significant: $\hat{\gamma}_{02}$ = -.004 (*t* (134) = -2.40, *p* < .05). This result means that by earning one more silver trophy, the intercept is expected to decrease by .4% on average controlling for all other predictors. In terms of the standardized coefficients (shown as "Stnd. Coef." in Table 4), the effect of GOLDS (-.09) was minimally greater on average than SILVERS (-.06) on student persistence across all the students. That is, students persisted more in problems they solved with silver trophies than gold ones—hence, most of the persistence time was generated from the unsolved problems and in problems which were solved with silver trophies. Note that to achieve a gold trophy in a problem, one must solve the problem with an elegant solution (e.g., with drawing only one object). Therefore, gold solutions might be accomplished in a shorter amount of time than silver solutions.

The average effect of GM difficulty on persistence was significant: $\hat{\gamma}_{10}$ = .03 (*t* (3,628) = 2.27, *p* < .05). This finding means that with one unit increase in the problem's GM, students' persistence increases by 3% on average controlling for all other predictors. The average effect of the PU difficulty on persistence was also significant: $\hat{\gamma}_{10}$ = .10 (*t* (3,628) = 10.39, *p* < .001). This result means that with one unit increase in the problem's PU, students' persistence increases by 11% on average controlling for all other predictors. Comparing the standardized coefficients, PU (.19) is more effective than GM (.04) in predicting students' persistence.

For coefficients of the dummy variables representing five categories for the simple machine used to solve a problem (i.e., ramp, springboard, pendulum, lever, or none), because we selected the lever category as the reference group, the coefficient for each category is the mean difference between that category and the lever group controlling for all other predictors. Again, these coefficients should be interpreted as percentages of the mean difference. The mean differences between all the four categories and the reference group were statistically significant. The categories unsolved problems (NONE), solved by a ramp (RAMP), and solved by a pendulum (PEN) had 43%, 16 %, and 6% smaller persistence means than the solved by a lever (LEVER) group respectively, and only solved by a springboard (SB) group has a 5% larger persistence mean than the LEVER group, holding all other predictors constant. Therefore, we can rank the simple machines in terms of the average persistence across students with regard to the simple machine used for solving a problem as NONE, RAMP, PEN, LEVER, and SB, from low to high. This indicates that the type of simple machine used to solve a problem can predict students' persistence in *Physics Playground*.

Finally, we compared means of students' persistence regarding primary concept, controlling for all other predictors. We selected ECT (energy can transfer) as the reference group. The mean differences of students' persistence in problems with POT (properties of torque), COM (conservation of momentum), and ECD (energy can dissipate) and problems with ECT (energy can transfer) as their primary concept were statistically significant. More specifically, an average student playing problems with POT and COM as the primary concept was predicted to show 7% and 27% *lower* persistence respectively than playing problems with ECT as the primary concept, holding all other predictors constant. Further, an average student playing problems with ECD as the primary concept was predicted to show 34% *higher* persistence on average than students playing problems with ECT as their primary concept, controlling for all other predictors.

The mean difference between ECT and NFL (Newton's First Law) was only 2% and was not statistically significant. Therefore, we can rank the concepts in terms of the average persistence shown across students when solving problems as COM, POT, NFL, ECT, and ECD, from low to high. This indicates that the primary physics concept of a problem has predictive power on students' persistence in *Physics Playground*. This finding might provide some evidence about the difficulty of these concepts when used in the context of a digital game like *PP* (as there are more than physics concept at play per problem—e.g., a secondary concept changing the conceptual difficulty of a problem). However, when taught in isolation in a classroom, the difficulty of these concepts might be different.

Model Comparison

Finally, the residual variance for the random intercept was statistically significant (χ^2 (134) = 365.73, $p < .001$), meaning that there was left-over variability across students for the random intercept. To examine how much of the within-school variability was explained by adding the level-1 predictors, we calculated Pseudo R-squared, using the following formula (Raudenbush & Bryk, 2002):

$$\text{Pseudo } R^2 = \frac{\hat{\sigma}^2_{\text{unconditional}} - \hat{\sigma}^2_{\text{with predictors}}}{\hat{\sigma}^2_{\text{unconditional}}}$$

where $\hat{\sigma}^2_{\text{unconditional}}$ is the level-1 residual variance of the fully unconditional model (.28) and $\hat{\sigma}^2_{\text{with predictors}}$ is the level-1 residual variance with level-1 predictors (.23). The added level-1 predictors helped explain 18% of the variance in the outcome variable compared to the fully unconditional model. Next, we discuss our findings, limitations, and future research.

Discussion and Conclusion

Persistence is a valuable attribute (Duckworth et al., 2007) and other researchers and organizations have emphasized how important persistence is for being successful—in school or generally in life (Eisenberger, 1992; Eisenberger & Leonard, 1980; Moore & Shute, 2017; Partnership for 21st Century Learning, 2015) Well-designed digital games have great potential for both assessing and fostering such personal variables. Stealth assessment (Shute, 2011) provides an infrastructure with an evidence-based methodology that can use digital games as an assessment vehicle to assess and support hard-to-measure constructs like persistence. Investigating the effects of particular features of digital games (e.g., GM difficulty and PU difficulty) on learners' persistence can help educators design games that can improve student persistence.

From our findings, we can say that (1) based on the analysis of ICC, the individual differences within students' gameplay could not be ignored in the analysis of this data, which means on average students showed differences in persistence; and (2) both GM and PU had significantly predicted student persistence. Specifically, the more difficult the problem, the more students persisted. This finding aligns with the definition of persistence mentioned earlier (i.e., Feather, 1962). The findings also tended to validate our difficulty indices. Specifically, PU had a greater predictive power for predicting student persistence than GM. Figuring out the physics concept behind the best solution for a problem in *Physics Playground* is very important for students to find the best strategy (simple machine) to solve the problem.

On the other hand, when learners become familiar with the game, the GM difficulty of a problem becomes relatively less impactful on students' performance after playing some problems. Therefore, the power of PU in predicting students' persistence seems reasonable. These findings show that our hypotheses about game problems difficulty and persistence are met.

It is worth noting that persistence is different from the difficulty of game problems. One might argue that more difficult problems can take more time to solve—which is reasonable. However, in our study, we included the times on unsolved problems and problems which were solved above the average time. The time coming from such conditions can be seen as persistence and not just difficulty (although there is some effect of difficulty involved, and we investigated this effect in this study). Consider this example: an unskilled player may spend a long time (above average duration) solving an easy problem, or quit solving a difficult problem after just a brief amount of time. These conditions existed in the data we used and coded as persistence in our study (the former problem was easy, and the later problem was difficult). In other words, the persistence time included in our study had to do more with persistence (mostly when the problems were not solved) than difficulty.

We also investigated how simple machine usage affected student persistence. We ranked simple machines in terms of the average persistence produced by students when solving a problem using the simple machines as NONE (for an unsolved problem), RAMP, PEN (pendulum), LEVER, and SB (springboard) from low to high. These findings make sense because the complexity of drawing a springboard is substantially greater than the complexity of creating a ramp. This ranking of simple machines again supports and validates our GM difficulty index supporting our hypothesis about simple machines used in the game problems.

Regarding the effects of a game problem's primary concept on student persistence, we were able to rank order physics topics relative to students' average persistence when solving a problem. The ranking of concepts from low to high was: conservation of momentum, properties of torque, Newton's first law, energy can transfer, and energy can dissipate. Note that this ranking does *not* align perfectly with the conceptual ordering of these topics based on our physics experts' input, who noted that Newton's laws of motion should be taught earlier than momentum, energy, and torque. However, our physics experts included both primary and secondary physics concepts of the game problems in the PU rubric. Our current analyses examined the problems' primary concept, which may have resulted in this discrepancy between our PU difficulty index and the ranking of concepts by our physics experts. Moreover, we had an unbalanced number of problems for each concept (see Table 3), making this ranking unstable. Our hypothesis regarding the primary concept of the game problems was not met. More detailed research is needed to answer the *why* question for this ranking.

As we hypothesized, the number of gold and silver trophies collected by students is significant predictors of student persistence (level-2 predictors)—with a negative relationship. When they collect more trophies, the time spent on the problems (our measure of persistence) decreases. In other words, according to the persistence measure used in this study, high-performing students had less chance to show their persistence, mainly because the "tasks" were relatively easy for them. One implication of this finding is that to train students (both low and high-performing) to become more persistent, educational games can include a stealth assessment of persistence and use that information to provide adaptive challenges for students. In that case, both types of students can face challenges that can help them get enough training to be persistent. It is also essential to have challenges with the right level of difficulty so that students can overcome challenges with a reasonable amount of effort. That is, too difficult problems may lead to frustration and quitting behaviors (Karumbaiah et al., 2018), while too easy problems may be boring.

Several student-level (level-2) predictors, such as gaming frequency, gender, physics knowledge, and post-test, were not significant predictors of persistence, which could be due to this study's limitations (discussed in the next section). These findings are against our hypotheses we made about student-related variables. The non-significant gender difference is in alignment with Jenson and de Castell's (2011) study showing that gender differences in digital gameplay can be diminished when girls are given more autonomy. Discussing the conditions our study provided for girls, which led to girls showing the same persistence level as boys, is not in this study's scope.

In conclusion, our study's findings suggest that ordering game problems based on their difficulty indices can support the development of student persistence. With persistence, the odds of learning from the game increases. The findings of this study could help game-based learning and psychology researchers to (a) accurately measure constructs such as persistence using digital games and assessment methods such as stealth assessment, and (b) consider both the effects of conceptual and game mechanics difficulty on students' persistence when designing digital games.

In the future, research of this nature (exploring predictors of students' persistence in educational games) can help shape the next generation of educational games. Those games could include real-time assessment of constructs such as persistence and detect behaviors that indicate a lack of persistence. Such games can use the assessment of persistence to adapt their challenges and game environment to students' persistence level to help students persist more when faced with game challenges. The ultimate goal would be to help students practice being more persistent using educational games and transfer that skill in their real-life situations. More persistent students can stick to their goals and achieve them rather than quitting when faced with problems (Duckworth, 2007). Further research is needed to see if our findings can be generalized to other educational games, educational settings, student demographics, and subject matter (e.g., mathematics). Next, we discuss the limitation of our study and suggest some future related research areas.

Limitations and Future Research

In this study, we operationalized persistence as the time spent on unsolved problems as well as above-average times spent on solved problems in *Physics Playground*. An improvement to this study could be including other measures of persistence derived from the log data. For example, the number of times a student revisited a problem to either get a trophy or change a silver trophy to a gold one are additional indicators of persistence. When we investigated the number of revisits students made to each problem, we found a strong, negative correlation (r = -.80) between time spent on the problem (our current persistence measure) and the number of revisits. Therefore, we decided not to include the number of revisits for this study. We could conduct another analysis using the number of revisits as the only measure of persistence, but we chose not to do this to keep the current study simple and clear.

Moreover, future studies could use data mining techniques to distinguish between productive and counter-productive persistence. For example, Owen et al. (2019) referred to counter-productive persistence as "wheel-spinning" and defined it as "spending considerable time on a topic without achieving mastery" (p. 378). It would be very informative to identify what variables can predict productive vs. counter-productive persistence. Knowing the factors which can lead to either type of persistence can help researchers and game designers create situations for productive persistence (which can lead to learning, according to Owen et al., 2019) and avoid students falling into the trap of counter-productive persistence by providing supports or asking students to leave the current game problem, play another problem and revisit the current one at another time.

Another improvement to our model could be the inclusion of additional level-2 (student-level) predictors—for example, students' game satisfaction, physics self-efficacy, academic performance (e.g., GPA), motivation, and other possible predictors that can affect students' persistence based on the literature. Since the data were collected in the past, we did not have any means to obtain any information about these predictors. We encourage other researchers to design studies (including level-1—game-related—, and level-2—student-related predictors) with a specific goal of identifying significant predictors of persistence in games. Finally, we encourage other researchers to use other educational games covering other subjects (e.g., math) to investigate the student-level and game-level predictors of students' persistence. Such research can lead to games that can assess and enhance students'' persistence—an essential skill our children need to have to be successful in life.

Acknowledgments

This work was supported by the US National Science Foundation [award number #037988] and the US Department of Education [award number #039019]. We also would like to acknowledge Russell Almond for his feedback on this paper.

References

Almond, R. G., Kim, Y. J., Velasquez, G., & Shute, V. J. (2014). How task features impact evidence from assessments embedded in simulations and games. *Measurement: Interdisciplinary Research and Perspectives*, *12*(1–2), 1–33. https://doi.org/10.1080/15366367.2014.910060

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. New York, NY: Springer-Verlag.

Andrews, G. G. (2008). Gameplay, gender, and socioeconomic status in two American high schools. *E-learning and Digital Media*, *5*(2), 199-213.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Benoit, K. (2011). *Linear regression models with logarithmic transformations*. *22*(1), 23–36. London, UK: London School of Economics.

Croxton, R. A. (2014). The role of interactivity in student satisfaction and persistence in online learning. *Journal of Online Learning and Teaching*, *10*(2), 314.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Journal of Educational Technology & Society*, *17*(1), 17-28.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Eisenberger, R. (1992). Learned industriousness. *Psychological Review*, *99*(2), 248. https://doi.org/10.1037/0033-295X.99.2.248

Eisenberger, R., & Leonard, J. M. (1980). Effects of conceptual task difficulty on generalized persistence. *The American Journal of Psychology*, *93*(2), 285–298. https://doi.org/10.2307/1422233

Elliot, A. J., & Dweck, C. S. (2013). *Handbook of competence and motivation* (1st ed.). New York, London: Guilford Publications.

Feather, N. T. (1962). The study of persistence. *Psychological Bulletin*, *59*(2), 94–115. https://doi.org/10.1037/h0042645

Garson, G. D. (2013). Introductory guide to HLM with HLM 7 software. *Hierarchical linear modeling: Guide and applications*, 55-96.

Gee, J. P. (2005). *Learning by design: Good video games as learning machines*. 12.

Granic, I., Lobel, A., & Engels, R. C. M. E. (2013). The benefits of playing video games. *American Psychologist*, 13.

Hox, J. J., Moerbeek, M., Schoot, R. van de, Moerbeek, M., & Schoot, R. van de. (2017). *Multilevel Analysis: Techniques and Applications* (3rd ed.). New York: Routledge, https://doi.org/10.4324/9781315650982

IBM Corp. (2017). IBM SPSS statistics (Version 25). Retrieved from https://www.ibm.com/products/spss-statistics

Israel-Fishelson, R., & Hershkovitz, A. (2019). Persistence in a game-based learning environment: The case of elementary school students learning computational thinking. *Journal of Educational Computing Research*, *58*(5), 891-918.

Jenson, J., & de Castell, S. (2011). Girls@ Play: An ethnographic study of gender and digital gameplay. *Feminist Media Studies*, *11*(2), 167-179.

Jackson, L. A., Gardner, P. D., & Sullivan, L. A. (1993). Engineering persistence: Past, present, and future factors and gender differences. *Higher Education*, *26*(2), 227-246.

Kai, S., Almeda, M. V., Baker, R. S., Heffernan, C., & Heffernan, N. (2018). Decision tree modeling of wheel-spinning and productive persistence in skill builders. *JEDM| Journal of Educational Data Mining*, *10*(1), 36-71.

Karumbaiah, S., Rahimi, S., Baker, R. S., Shute, V., & D'Mello, S. K. (2018). Is student frustration in learning games more associated with game mechanics or conceptual understanding? In J. Kay, R. Luckin, M. Mavrikis, & K. Porayska-Pomsta (Eds.), *International Conference of Learning Sciences* (pp. 1–2). London, UK.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development*, *10*(3), 2.

Kiefer, A., & Shih, M. (2006). Gender differences in persistence and attributions in stereotype relevant contexts. *Sex Roles*, *54*(11-12), 859-868.

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

Loup, G., Serna, A., Iksal, S., & George, S. (2016). Immersion and persistence: Improving learners' engagement in authentic learning situations. In *European conference on technology enhanced learning* (pp. 410-415). Springer, Cham.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86. https://doi.org/10.1027/1614-2241.1.3.86

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81. https://doi.org/10.1037/0022-3514.52.1.81

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Moore, G. R., & Shute, V. J. (2017). Improving learning through stealth assessment of Conscientiousness. In A.

Marcus-Quinn & T. Hourigan (Eds.), *Handbook on digital learning for K-12 schools* (pp. 355–368). https://doi.org/10.1007/978-3-319-33808-8_21

Owen, V. E., Roy, M. H., Thai, K. P., Burnett, V., Jacobs, D., Keylor, E., & Baker, R. S. (2019). Detecting Wheel-Spinning and Productive Persistence in Educational Games. *International Educational Data Mining Society*.

Partnership for 21st Century Learning. (2015). Framework for 21st-century learning. Retrieved from http://www.p21.org/news-events/p21blog/2312-data-points-toward-effective-practices

Perez-Felkner, L., McDonald, S. K., & Schneider, B. L. (2014). What happens to high-achieving females after high school? Gender and persistence on the postsecondary STEM pipeline. *Gender differences in aspirations and attainment: A life course perspective*, 285-320.

Qian, M., & Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, *63*, 50-58.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Reigeluth, C., & Karnopp, J. (2013). *Reinventing schools: It's time to break the mold*. R&L Education.

Richard, G. T. (2013). Gender and gameplay: Research and future directions. *Playing with virtuality: Theories and methods of computer game studies*, 269-284.

Rotherham, A. J., & Willingham, D. T. (2010). "21st-Century" skills-not new, but a worthy challenge. *American Educator*, *17*, 17–20.

RStudio Team. (2016). RStudio: Integrated Development Environment for R (Version 1.0.136). Retrieved from http://www.rstudio.com/

Ryans, D. G. (1939). The measurement of persistence: An historical review. *Psychological Bulletin*, *36*(9), 715. https://doi.org/10.1037/h0060780

Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 139-187). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group.

Shute, V., Leighton, J. P., Jang, E. E., & Chu, M.-W. (2016). Advances in the science of assessment. *Educational Assessment*, *21*(1), 34–59. https://doi.org/10.1080/10627197.2015.1127752

Shute, V., Almond, R., & Rahimi, S. (2019). *Physics Playground* (Version 1.3) [Computer software]. https://pluto.coe.fsu.edu/ppteam/pp-links/

Shute, V., & Wang, L. (2016). Assessing and supporting hard-to-measure constructs in video games. In A. A. Rupp & J. P. Leighton (Eds.), *The handbook of cognition and assessment* (pp. 535–562). https://doi.org/10.1002/9781118956588.ch22

Shute, V. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age Publishers.

Shute, V., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., … Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, *86*, 224–235. https://doi.org/10.1016/j.compedu.2015.08.001

Shute, V., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 43–58). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-3546-4_4

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: The MIT Press.

Sümer, M., & Aydın, C. H. (2018). Gamification in open and distance learning: A systematic review. *Learning, Design, and Technology: An international compendium of theory, research, practice, and policy*, 1-16.

van Garderen, K. J., & Shah, C. (2002). Exact interpretation of dummy variables in semilogarithmic equations. *The Econometrics Journal*, *5*(1), 149–159. Retrieved from JSTOR.

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, *29*(6), 2568–2572. https://doi.org/10.1016/j.chb.2013.06.033

Ventura, M., Shute, V. J., & Small, M. (2014). Assessing persistence in educational games. In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for adaptive intelligent tutoring systems: Learner modeling, Volume 2.* (pp. 93-101). Orlando, FL: U.S. Army Research Laboratory.

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education*, *60*(1), 52–58. https://doi.org/10.1016/j.compedu.2012.07.003

Wang, L. (2017). *To help, or not to help? An investigation of the effects of learning support in Physic Playground* [Doctoral dissertation]. Florida State University, USA.

Author Information

Seyedahmad Rahimi
https://orcid.org/0000-0001-9266-758X
Florida State University
College of Education. 3212 Stone Building.
Department of Educational Psychology and Learning Systems.
1114 W. Call Street
Tallahassee, FL, 32306
USA
Contact e-mail: *srahimi@fsu.edu*

Valerie Shute
https://orcid.org/0000-0002-9179-017X
Florida State University
College of Education. 3212 Stone Building.
Department of Educational Psychology and Learning Systems.
1114 W. Call Street
Tallahassee, FL, 32306
USA

Qian Zhang
https://orcid.org/0000-0003-4396-9375
College of Education. 3212 Stone Building.
Department of Educational Psychology and Learning Systems.
1114 W. Call Street
Tallahassee, FL, 32306
USA