Received: 10 March 2020 | Revised: 25 May 2020 | Accepted: 20 June 2020
DOI: 10.1111/jcal.12473

ARTICLE

Journal of Computer Assisted Learning WILEY

Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games

Valerie Shute | Seyedahmad Rahimi | Ginny Smith | Fengfeng Ke | Russell Almond | Chih-Pu Dai | Renata Kuba | Zhichun Liu | Xiaotong Yang | Chen Sun

Department of Educational Psychology and Learning Systems, College of Education, Florida State University, Tallahassee, Florida

Correspondence
Valerie Shute, Department of Educational Psychology and Learning Systems, College of Education, Florida State University, 3205G Stone Building, 1114 W. Call Street, Tallahassee, FL 32306.
Email: vshute@fsu.edu

Funding information
Institute of Education Sciences, Grant/Award Number: #039019; National Science Foundation, Grant/Award Number: #037988

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12473.

Abstract
In this study, we investigated the validity of a stealth assessment of physics understanding in an educational game, as well as the effectiveness of different game-level delivery methods and various in-game supports on learning. Using a game called *Physics Playground*, we randomly assigned 263 ninth- to eleventh-grade students into four groups: adaptive, linear, free choice and no-treatment control. Each condition had access to the same in-game learning supports during gameplay. Results showed that: (a) the stealth assessment estimates of physics understanding were valid—significantly correlating with the external physics test scores; (b) there was no significant effect of game-level delivery method on students' learning; and (c) physics animations were the most effective (among eight supports tested) in predicting both learning outcome and in-game performance (e.g. number of game levels solved). We included student enjoyment, gender and ethnicity in our analyses as moderators to further investigate the research questions.

KEYWORDS
adaptivity, game-based learning, learning supports, stealth assessment, STEM education

1 | INTRODUCTION

STEM-related education has, and will continue to have, a significant impact on people's lives around the world (Hasanah & Tsutaoka, 2019; Kelley & Knowles, 2016). That is, as we grow into a global society, becoming increasingly interconnected and technologically dependent, the workforce will have a high demand for people with STEM-related knowledge and skills. However, most educational systems are not currently producing enough STEM graduates to fill this need. For a specific example relevant to this paper, the number of students studying physics is not only low but also lacks diversity. According to the most recent 2015 report of the Trends in International Mathematics and Science Study (Tofig, 2017), only 4.8% of 18-year-old students in the United States enroll in physics courses, and more than half (61%) of those students are male. Moreover, the American Physical Society (2020) used 2013–2017 data collected by the Integrated Postsecondary Education Data System, along with ethnicity data from the United States Census, and reported that the percentage of physics bachelor's degrees awarded to Hispanics was only 8%, less than 3% to African Americans, just 18% to females and 82% to males.

To increase the workforce possessing STEM-related knowledge and skills, educators need to increase the number and diversity of students entering STEM areas, especially in physics. The presence of diverse perspectives (including gender and ethnic diversity) in most fields leads to more innovative solutions. Although this problem can be approached in various ways, the current research focuses on two related goals that we believe can have the largest impact. The first goal is to get more children, particularly females and certain underrepresented minorities (e.g. Black and Hispanic children), excited about and interested in physics. Recognizing that interest alone is not enough, our second goal is to identify ways to facilitate and deepen physics learning. Well-designed digital games represent a promising vehicle for meeting both goals: capturing children's interest in physics, and supporting active, contextualized learning across all children.

1.1 | Well-designed educational games as vehicles for assessment and learning

As children grow up in today's technology-rich world, one commonality is that they love to play digital games. According to the Entertainment Software Association's report (2019), 75% of Americans have at least one gamer in their household. In 2018, Americans spent \$43.4 billion to purchase various types of digital games. Moreover, playing digital games is prevalent across all gender, ethnic and socioeconomic groups (Entertainment Software Association, 2019).

This huge popularity of digital games has spawned interest in examining the effectiveness of such games on assessing and improving various competencies (i.e. knowledge, skills and other attributes). These games are suitable for assessment because designers can embed *continuous* measures of learning therein (Shute, Ventura, & Ke, 2015), overcoming problems associated with disruptive and limited assessments evidenced by multiple-choice types of tests (Shute, Ke, Almond, Rahimi, Smith, & Lu, 2019; Shute & Ventura, 2013). For instance, researchers have used games to assess various competencies, such as problem solving (e.g. Yang, 2012); computational thinking (Zhao & Shute, 2019); emotional regulation (Spann, Shute, Rahimi, & D'Mello, 2019); visual–spatial abilities and attention (e.g. Green & Bavelier, 2007, 2012; Shute et al., 2015); persistence (Ventura, Shute, & Zhao, 2013); creativity (Jackson et al., 2012); and civic engagement (Ferguson & Garza, 2011). Collectively, the findings show that well-designed digital games can be promising vehicles for assessment and learning (Shute & Ke, 2012; Shute, Leighton, Jang, & Chu, 2016).

There are several reasons why well-designed digital games are suitable vehicles for learning. These games provide ongoing and timely feedback, interactivity and active participation (Gee, 2003; Ifenthaler, Eseryel, & Ge, 2012; Shute, Ke, & Wang, 2017a). In addition, adaptive challenges (i.e. tasks matched to players' abilities; Vygotsky, 1978) and dynamic feedback in well-designed digital games can facilitate the state of flow (Csikszentmihalyi, 1990). The flow state happens when one fully engages in a task, loses track of time and experiences a deep feeling of enjoyment. Experiencing engagement while playing educational games (i.e. digital games with learning purposes) is important for learning to occur (Shute & Ke, 2012).

Despite years of research in designing game-based learning, there is still a lot we do not know. We *do* know that an engaging game capable of supporting learning for a broad range of students must accurately, and in real time, assess the competencies it aims to support. Then, adaptive challenges, tailored feedback and other learning supports can be delivered to fit students' individual needs (Conati, 2002; Shute & Zapata-Rivera, 2012). An educational game with accurate and ongoing assessment (e.g. stealth assessment), as well as adaptive delivery of challenges, feedback and learning supports, can potentially help students attain and sustain the flow state, matching gameplay with skill level. However, we *do not* currently know the psychometric qualities (i.e. reliability, validity and fairness) of these in-game assessments, nor do we know the added value of adaptivity in games relative to learning (e.g. Clark, Virk, Barnes, & Adams, 2016; Leemkuil & de Jong, 2012; Sampayo-Vargas, Cope, He, & Byrne, 2013). Finally, incorporating learning supports into a game to maximize learning without losing the fun of gameplay has been an important and unresolved challenge in game-based learning research for over a decade (Shute, Almond, & Rahimi, 2019; Shute, Ke, et al., 2019; Squire, 2006).

There are thus three main aims of the current study to address these gaps. First, we want to establish the psychometric qualities of a stealth assessment methodology used in the game – *Physics Playground* (*PP*; Shute, Almond, & Rahimi, 2019). Second, we plan to evaluate the effects of three different conditions of the game (i.e. adaptive, linear and free choice) on student learning of physics. Finally, we want to test the effectiveness of the in-game learning supports relative to students' acquisition of physics knowledge and skills. Towards that end, we address the following research questions:

1. Is our stealth assessment of physics understanding a reliable, valid and fair measure?
2. Which delivery method of game levels (i.e. adaptive, linear or free choice) is more effective for improving students' physics understanding when controlling for incoming knowledge?
3. Which type of embedded learning support most effectively enhances learning and game performance?

For research questions 2 and 3, above, we also examined these variables (i.e. delivery method and learning support) relative to student enjoyment, and included gender and ethnicity as moderators in the equations.

2 | BACKGROUND

2.1 | Stealth assessment

To adapt game challenges and learning supports to students, successfully, in a well-designed digital game, we need to accurately assess and track students' knowledge and skill development during gameplay. Unlike assessments in traditional educational settings that occur external to the learning activities, educational games provide an avenue for assessing learning while it occurs. To do so, we need a type of assessment that is unobtrusive and ongoing within a well-designed digital educational game (Shute, Rahimi, & Chen 2017b). Creating such an assessment is complex and challenging. Stealth assessment in game-based learning environments offers a possible solution (e.g. Georgiadis, van Lankveld, Bahreini, & Westera, 2018; Min et al., 2019; Shute, 2011).

Stealth assessment is based on an assessment design framework called evidence-centred design (ECD; Almond, Mislevy, Steinberg, Yan, & Williamson, 2015; Mislevy, Steinberg, & Almond, 2003). ECD's primary purpose is to structure the collection of evidence needed to make valid claims about the level of students' competencies. ECD defines a framework of conceptual and computational models that work in harmony. The four core ECD models are: (a) the competency model (CM)—operationalizing the construct we want to assess (e.g. conceptual physics understanding) and defining the claims to be made about students' competencies; (b) the evidence model (EM)—automatically scoring and accumulating valid evidence (i.e. observables) of a claim about students' competencies (i.e. unobservables); (c) the task model (TM)—detailing the nature and form of the tasks (e.g. game levels) that will elicit the evidence needed for the EM; and (d) the assembly model (AM) which specifies the number, types and sequencing of tasks.

In stealth assessment, specific gameplay behaviours are dynamically linked to the CM. As students interact with the game environment, they generate a continuous stream of data captured in the game's log files. The stealth assessment filters through and analyses that data—in real-time—to identify and extract evidence related to the CM. This is the evidence identification (EI) process. The EI's output is the input data (e.g. scores, tallies) for the evidence accumulation (EA) process, which statistically updates the claims about relevant competencies in the CM (e.g. the probability of a student being low, medium or high on a given competency). The more evidence a student generates during gameplay, the more accurate the estimates of competency levels. Competency-level estimates can be used for various purposes (e.g. adaptive delivery of game levels, feedback and learning supports). *PP* uses stealth assessment of conceptual physics understanding (discussed in detail in Section 3) as the basis for developing adaptive delivery of its game levels.

2.2 | Adaptivity in digital games

Facilitating and maintaining the flow state in any learning activity requires clear goals, achievable challenges and immediate feedback (Csikszentmihalyi, 1997). Adaptive balancing of challenges in digital games can facilitate and sustain the flow state (see Vandewaetere, Cornillie, Clarebout, & Desmet, 2013). Theoretically, such adaptation leads to higher levels of enjoyment, and eventually to higher levels of learning when playing well-designed digital games (e.g. Bontchev & Georgieva, 2018; Jagušt, Botički, & So, 2018). Adaptive sequencing of game levels based on students' performance and game level difficulty is consistent with the theory of the zone of proximal development (ZPD; Vygotsky, 1978). ZPD indicates that learning is optimized when students are given learning activities at the edge of their abilities.

Limited, and somewhat conflicted research exists, comparing adaptive sequencing to non-adaptive sequencing (i.e. linear or free choice) within digital games. In one study, Sampayo-Vargas et al. (2013) found that students who played the adaptive version of their game scored significantly higher on the post-test compared to students who played linearly and students who completed a related non-game activity. In contrast, adaptivity may not actually generate better learning relative to non-adaptive conditions. Other researchers have found that learning environments offering students choices to navigate within a learning environment can be more effective than environments that make those decisions for students (e.g. Black & Deci, 2000; Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004). For instance, Vanbecelaere et al. (2019) compared the effectiveness of an adaptive versus a non-adaptive educational game on kindergarten students' cognitive and non-cognitive gains. Results showed significant learning gains in general but no difference between conditions. The findings of our study will add to the dialog on adaptivity in digital educational games.

2.3 | Learning supports in digital games

There is growing consensus and weight of evidence suggesting that well-designed learning supports embedded in digital games can promote students' learning outcomes (e.g. Chen & Law, 2016; Moyer-Packenham et al., 2019; Sun, Chen, & Chu, 2018; Wouters & van Oostendorp, 2013; Young et al., 2012; Zeglen & Rosendale, 2018). For instance, Tsai, Kinzer, Hung, Chen, and Hsu (2013) tested the effectiveness of in-game, content-related learning supports designed to promote learning the principles of projectile motion. Targeted multiple-choice questions were included after each game level. Results showed a significant correlation ($r = .44$, $p < .05$) between students' average time spent with the learning supports and their post-test scores. It is important to note, however, that providing content-related learning supports that disrupt gameplay can disrupt enjoyment.

Kao, Chiang, and Sun (2017) used a game called *Crayon Physics Deluxe* to investigate the effectiveness of two types of learning supports: hints (i.e. minimal guidance) and worked examples (i.e. full expert solutions of a game level) in an experiment. Their study consisted of four groups in a pretest–post-test design—three gameplay groups (hints, worked examples and no supports) and one control group. While all students who played the game scored significantly higher on the post-test than those who did not play the game, the hint group scored significantly higher than all other groups on the post-test controlling for incoming knowledge. Students in the worked-example group and students in the no-support group showed no significant difference on physics knowledge acquisition. The authors speculated that students in the worked example group may have only tried to replicate what they saw instead of thinking deeply about the physics behind the levels. These results show the type of learning support and the degree of information provided within the learning support can affect students' learning.

It is important to foster motivation in educational games because in the absence of motivation, learning suffers (e.g. Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). Including in-game learning supports can potentially motivate students to do more advanced activities, and to accomplish more than when learning supports are not included (Gee, 2003; Huang & Oh, 2018; Johnson, 2019; Khamparia & Pandey, 2018; Wouters & van Oostendorp, 2013). Again, the big challenge is to include learning supports in games that help students learn the content knowledge deeply, but do not disrupt enjoyment while students are immersed in gameplay. With this in mind, we designed multiple learning supports in *PP* to maximize both students' learning and game performance. Next, we discuss the details of our study.

3 | METHOD

3.1 | Participants

Our sample consisted of 280 9th–11th grade students in a large K-12 school in the southeastern United States. We included the data from 263 students who completed both the pretest and posttest, submitted their parental consent forms and signed the assent form in this study. We had the same number of students self-identify as male (n = 128) and female (n = 128), with a wide range of ethnicities. Self-reported ethnicities representing more than 1% of the respondents were: Asian (n = 8), Black or African American (n = 77), Hispanic (n = 23), White (n = 114), Other (n = 7), Black or African American and White (n = 6), Black or African American and Hispanic (n = 3), and Hispanic and White (n = 9).

Students participated in the research activities during their science classes. After the no-treatment control group was selected, we randomly assigned participants into one of three experimental conditions—adaptive, linear or free choice. All students completed the pretest and posttest. Students in the experimental conditions played *Physics Playground* for 4 hr across six sessions. Students in all conditions were compensated with a $30 gift card after completing the post-test.

3.2 | Materials

All experimental materials were administered online. Students accessed the materials on school laptops, desktops, Chromebooks and Surface Pros.

3.2.1 | Educational game

PP is a computer-based game designed to help middle and high school students learn conceptual physics related to Newton's laws of force and motion, linear momentum, energy and torque. Using a mouse- or stylus-driven interaction, the goal is to move a green ball to hit a red balloon. The game includes two task types: sketching and manipulation. In the sketching levels, students can draw simple machines (i.e. ramp, lever, pendulum and springboard) that interact with the two-dimensional environment according to Newtonian mechanics (Figure 1). In the manipulation levels, students can change various physics parameters in the environment (i.e. gravity, air resistance, mass and bounciness of the ball), and manipulate external forces (i.e. puffers and blowers) (Figure 2). Each level is designed to be solved by specific physics parameters or agents.

Two physics experts helped us create our competency model (Figure 3). They also identified the primary and secondary physics concepts associated with each game level. This was represented using a Bayesian network (BN; Almond et al., 2015).

Game levels in *Physics Playground* differ in terms of difficulty. We created two sets of rubrics to determine a level's difficulty—one concerning its game mechanics, and the other related to the underlying physics. Game mechanics difficulty (which ranged from 1 to 5 per level) was based on factors such as the relative positions of the ball and the balloon, the number of obstacles present, the novelty of the problem and the number of objects or parameters required to solve the level. Physics difficulty (which also ranged from 1 to 5 per level) was based on a rubric created by our physics experts. The rubric considered the primary and secondary physics concepts (see nodes in the middle section of Figure 3) that were associated with each level. Scoring for a level's primary concept was: force and motion = 0; momentum and energy = 1; torque = 2. Another difficulty point was added if the level: (a) required the balancing of forces (i.e. equilibrium or Newton's third law); (b) involved conservation or the transfer of energy; or (c) consisted of both primary and secondary concepts that came from two different 'parents' (e.g. Newton's first law and energy can transfer). Each level was scored by two raters on both dimensions and any disagreements were resolved in consultation with our physics experts.

Based on the difficulty ratings, we established a par value (like in golf) per level related to the minimum number of objects (in the sketching levels), or attempts (in the manipulation levels) needed for an elegant/efficient solution. Pars determine if a person receives a silver or

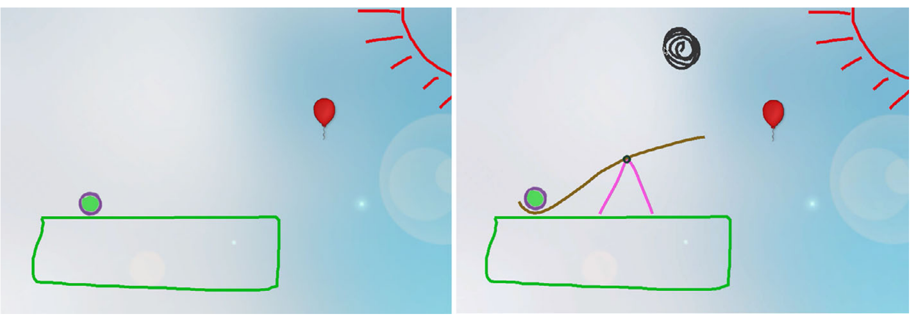


FIGURE 1 Sketching level in *PP*—to solve the level, students can draw a lever and drop a weight on the other side of the lever to lift the ball to the balloon. *PP*, *Physics Playground* [Colour figure can be viewed at wileyonlinelibrary.com]

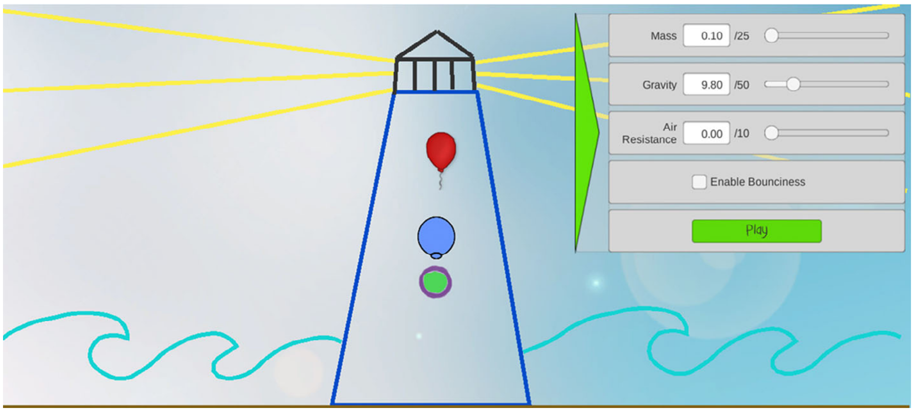


FIGURE 2 Manipulation level in *PP*—to solve the level, students have to manipulate the gravity slider and enable the bounciness function. *PP*, *Physics Playground* [Colour figure can be viewed at wileyonlinelibrary.com]

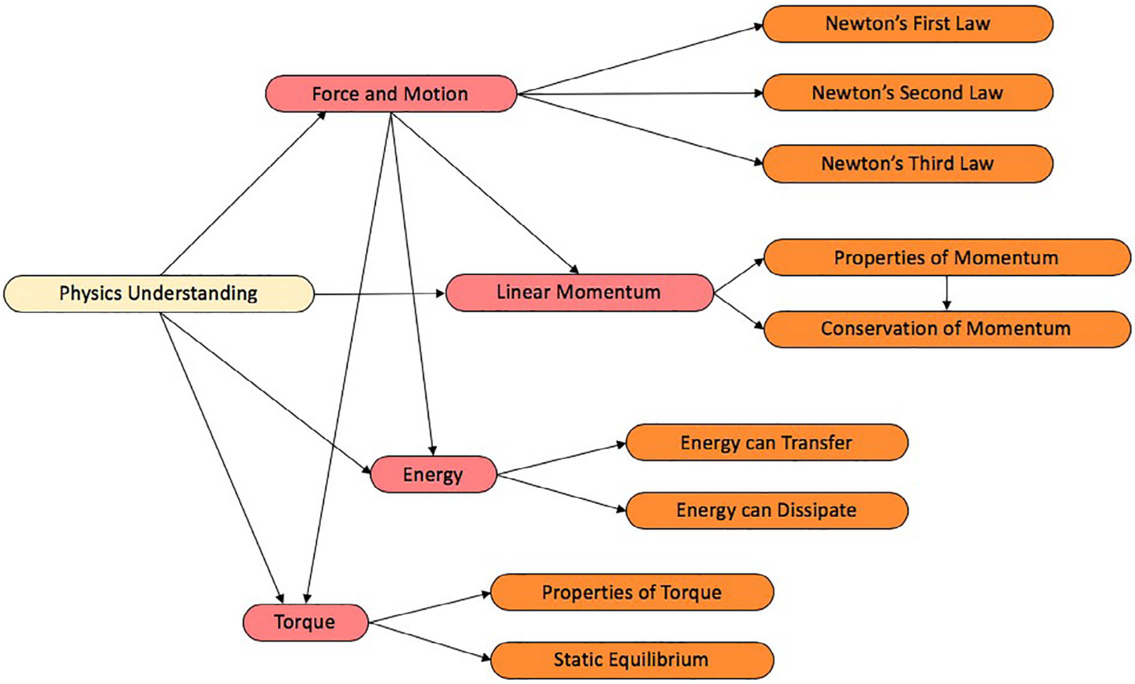


FIGURE 3 Physics understanding competency model for *PP*. *PP*, *Physics Playground* [Colour figure can be viewed at wileyonlinelibrary.com]

gold coin for their solution (i.e. if the student solves a level under par they receive a gold coin, otherwise silver). The adaptive algorithm (discussed later) uses pars to deliver easy, medium or hard levels in *PP*.

Students learn the game mechanics at the beginning of the game by playing tutorial levels for both task types (i.e. sketching and manipulation). Along with the tutorial levels, *PP* includes learning supports that can be accessed at any time during gameplay. A help button resides in the lower-right corner of the screen in each level. When students click on the help button, a popup menu shows three options: *Show me the Physics* (see Figure 4); *Show me a solution or a Hint*; and *Show me Game Tips*. Each support is briefly described below, but for more details and illustrations of the supports, see Shute, Almond, and Rahimi (2019) and Shute, Ke, et al. (2019).

Show me the Physics leads students to a screen containing the following physics supports: Animation, 'Definition', Formula (when applicable), Hewitt video and Glossary.

- The physics animation support (shown as 'Animation' in Figure 4) contains videos presenting physics concepts (e.g. properties of torque) in the game environment relevant to a student's current game level (e.g., see: https://bit.ly/38L8vHz).
- The definition support includes physics terms applicable to the game's content. Students watch a short animation illustrating the term in the game environment (e.g. gravitational potential energy) and complete the term's definition through a fill-in-the-blank, drag-and-drop interaction.
- The formula support presents the physics concept's formula and defines the associated variables.
- The Hewitt video support contains cartoon animations explaining various physics concepts, originally developed by Paul Hewitt, and edited with permission to present a targeted competency in the game.
- The glossary support contains brief explanations of a set of physics terms relevant to the game.

FIGURE 4 Learning supports available in Show me the Physics (top) and example of the learning support for an interactive definition (bottom) [Colour figure can be viewed at wileyonlinelibrary.com]

Show me a solution or a Hint presents two buttons: 'Show me a hint' and 'Show me a solution.' If students select the former, they will see a written hint (e.g. *Try drawing a pendulum*). If they select a solution, they will watch a short expert video solution (i.e. worked example) of the current level. *Show me Game Tips* presents snapshots of the game tutorials, a review of the game mechanics, and an illustrative key of the components in *My Backpack* (discussed in more detail below).

The game also includes an incentive system. As previously discussed, students earn either a silver coin (worth $10 game money) or a gold coin (worth $20) for solving a level based on the efficiency/elegance of their solution. Students can earn additional game money by accessing the physics learning supports. Conversely, students must pay $60 to view solutions (i.e. worked examples). Accessing hints and game tip supports do not cost or earn money for the student.

Upon finishing each level, students can choose to replay the level or proceed to the next level. Students can also access *My Backpack* (Figure 5) where they can check their gameplay progress, game money balance and levels of physics understanding. In addition, they can select the store to customize the type of ball (e.g. beach ball or volleyball) or change the background image and background music. Purchasing game customizations is another way students can spend the game money they earn.

The version of *PP* used in this study consisted of 10 tutorial levels and 81 game levels covering the nine physics competencies (i.e. children nodes—on the far right in the competency model) arranged in the following sequence: Newton's first law, energy can transfer, energy can dissipate, properties of momentum, conservation of momentum, properties of torque, equilibrium, Newton's second law and Newton's third law.

3.2.2 | Physics understanding test

We created 36 illustrative multiple-choice items covering the nine physics competencies in the game, counterbalanced between two equivalent forms for a pretest and posttest (pretest = 18 items, α = .77; posttest = 18 items, α = .82). Each form includes two items per competency. The items were (a) designed in the context of *PP* (i.e. including a video or an image from the game environment), (b)

FIGURE 5 My backpack—students can check their gameplay progress, money balance and physics understanding (top), as well as access the game store (bottom) [Colour figure can be viewed at wileyonlinelibrary.com]

developed with the help of two physics experts and (c) subjected to several pilot tests before administration in the current study (Figure 6).

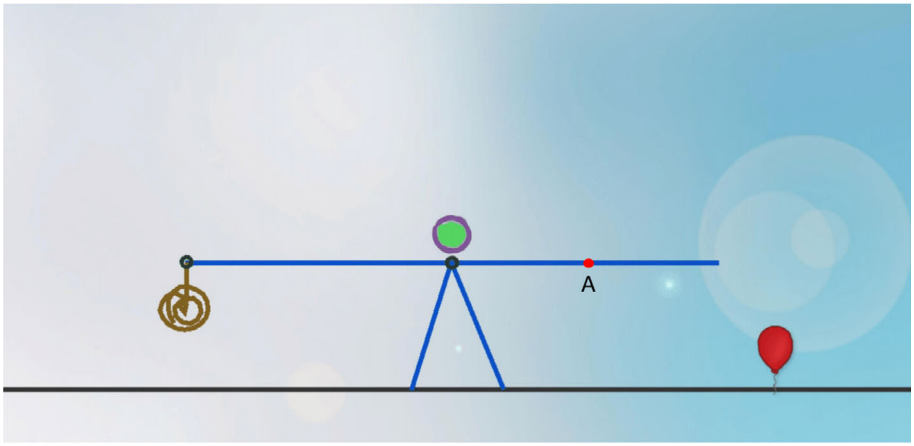
3.2.3 | Game and learning support satisfaction questionnaire

To evaluate students' satisfaction with both the game and the learning supports, we developed and used a 16-item questionnaire with two subscales: (a) game satisfaction, with 10 items, Cronbach's α = .86 (e.g. 'I enjoyed the game very much'), and (b) learning support satisfaction, with 6 items, Cronbach's α = .73 (e.g. 'The supports helped me understand the physics'). Students responded on a five-point Likert scale, from strongly disagree (1) to strongly agree (5). All items from the game and learning support satisfaction questionnaire are included in Table A1 in the Appendix, along with the descriptive statistics, per item.

3.3 | Research design

One purpose of the study was to examine the psychometric properties (reliability and validity) of our stealth assessment estimates of physics understanding. Then, using the valid in-game measures, we could develop and analyse the impact of adaptive sequencing of game levels on learning, and the effectiveness of the learning supports. We used a between groups repeated measure design with four conditions: adaptive sequencing (n = 64), linear sequencing (n = 68), free choice (n = 67) and control (n = 64). A few classes were selected to form the control group. The remaining participants were randomly assigned at the student level into one of the three treatment conditions on Day 1.

The *adaptive* condition used an algorithm to determine the order to present the levels to best fit a student's needs. We assessed students' physics competencies using a Bayesian network (BN) approach. The network is split into a core student model (corresponding to the CM, but student specific) and a number of Bayesian network fragments, corresponding to the EMs as applicable to that level. The EM



If an identical brown weight is added to the other side of the lever at point A, what will happen to the ball?
a) It will stay where it is.
b) It will roll to the right.
c) It will roll to the left.
d) It depends on the mass of the ball.

FIGURE 6 Example of an item from the physics understanding test. The correct answer is C [Colour figure can be viewed at wileyonlinelibrary.com]

fragments join to the student model at the competencies identified by the experts, and describe how observable outcomes from the game level (identified by the experts) relate to the competencies. At the end of each level, the events logged by the game engine associated with each game level are processed to set values for observables. The EM fragment is joined to the core student model and used to update the current state of knowledge about the competencies (Almond et al., 2015). Then, if the estimate of a student's current level of understanding about a competency (e.g. Newton's first law) was satisfactory (e.g. ~0.50), the adaptive algorithm would deliver more difficult levels associated with that competency. If the competency estimate fell too low (e.g. <0.33), the algorithm would pull an associated physics learning support video and present it to the student before delivering the next level. Once the competency estimate became sufficiently high (e.g. >0.83), the student would 'graduate' to the next competency. When a student graduated from all competencies, they would enter an endgame mode where the level selection process would select unplayed levels from each competency. Finally, in the case where the student was estimated as being 'high' on all nine competencies, and had completed all 81 levels, the algorithm provided levels for which the student received only a silver coin.

In the *linear* condition, students followed a predetermined sequence of levels, with limited ability to skip a level. Students could choose to go to the next or previous level without the need to solve the current level by pressing the 'Escape' button from the keyboard and then clicking either the 'Next' or 'Previous' button. However, in this condition, students could not freely choose which level to play. The researchers and physics experts arranged the sequence of the levels based on the progression of physics topics and associated level difficulty.

Finally, in the *free-choice* condition, students were provided with the same level sequence as the linear condition, but they could move back and forth through all of the levels in the game and select which one they wanted to play. Students in all three conditions had equal access to the learning supports during gameplay through the help button in the bottom-right corner of the screen.

3.4 | Procedure

The experiment spanned 6 days of class time, with six sessions per classroom in total. Each session was 50 min. On the first day, participants completed a demographic survey and an online pretest of physics knowledge (18 items), followed by an introduction to *PP* gameplay. Sessions two through five consisted of gameplay for the duration of the class. Students played the game independently with headphones, and were monitored by the members of our research team. The final session consisted of gameplay followed by the online post-test, the game and learning support satisfaction survey, and receipt of the gift card.

4 | RESULTS

Our results are presented in line with the research questions related to the psychometric qualities of the stealth assessment estimates, the impact of different game conditions (adaptive, linear and free choice) on learning and enjoyment, and the impact of various learning supports on learning and enjoyment. We hypothesized that (a) the stealth assessment measures would be valid, (b) learning from, and enjoyment of gameplay would occur—perhaps with an advantage to the adaptive condition and (c) our learning supports would further boost learning and enjoyment, beyond that from simple gameplay.

4.1 | Validation of stealth assessment estimates

To validate our stealth assessment estimates of physics understanding, we first established the reliability of our external pretest and

TABLE 1 Fine-grained validation of stealth assessment estimates

Stealth assessment estimates	Pretest	Post-test
Force and motion	0.29**	0.30**
Linear momentum	0.27**	0.27**
Energy	0.22**	0.35**
Torque	0.14*	0.18**

*$p < .05$.
**$p < .01$.

post-test measures. As described earlier, each test consisted of 18 items, measuring all nine focal competencies in duplicate. These were matched forms, with multiple-choice formatted items accompanied by pictures. The tests have been revised across 2 years of testing, and the current reliabilities (Cronbach's α values) were: pretest = .77; posttest = .82; n = 263. Next, we computed the correlation between our overall 'physics understanding' estimate from the game (see far left side of Figure 3) with our external test scores. Results showed that both the pretest ($r = .36$, $p < .01$) and posttest ($r = .40$, $p < .01$) scores significantly correlated with the overall stealth assessment estimate.

We were also able to test specific correlations involving each one of our four competency estimates (see the middle of Figure 3, e.g. Torque) with the score of relevant pretest and post-test items (e.g. two item scores summed for Torque on pretest and post-test to create a Torque pretest and Torque post-test score). As with the general physics estimate, the more fine-grained stealth assessment estimates significantly correlated with their associated external measures both on pretest and post-test (see Table 1).

In summary, it appears that our stealth assessment estimates are measuring the constructs that we intended them to measure, overall as well as at a more granular, diagnostic level. Note that the BN scores are currently based only on the experts' original estimates. Refining the model using the data from the field test should yield even better measures of physics competency.

4.2 | Learning physics from the game

4.2.1 | Overall learning

The most general question related to learning addresses whether students, regardless of game condition, learned any physics from gameplay. We conducted a paired sample t test including all students in the three game conditions (adaptive, linear and free choice). As predicted, results showed that students scored significantly higher on the post-test (M = 12.46, SD = 3.86) than the pretest (M = 11.82, SD = 3.53), $t(198) = 3.10$, $p = .002$, $d = .17$, 95% CI [0.23, 1.04]. Because we had a no-treatment control group (i.e. students who did not play the game but completed the pretest, then a week later completed the post-test), we could see any possible 'test effects'. Results from a paired sample t test including just the subjects in the control group showed that they scored the same on the pretest (M = 11.61, SD = 3.65) and post-test (M = 11.59, SD = 4.19); $t(63) = -0.04$, $p = .97$, $d = .005$, 95% CI [−0.77, 0.74].

4.2.2 | Learning by condition

After establishing that learning occurred as a function of gameplay, and that there were no pretest influences on post-test scores, we then examined physics learning as a function of game condition (adaptive, linear and free choice). We computed an ANCOVA with post-test as the dependent variable, condition as the independent variable and pretest serving as the covariate. Surprisingly, results showed no significant outcome differences by condition, holding pretest constant ($F[2,195] = 0.34$; $p = .71$, partial $\eta^2 = 0.003$). Table 2 presents all pretest and posttest data across all four conditions. Note that there were no significant differences in pretest score by condition, so the subjects were randomized well.

4.2.3 | Learning by gender

Another question we addressed, touching on the fairness/equity issue discussed earlier, concerned physics learning from the game as a function of gender. We computed an ANCOVA, with post-test score as the dependent variable, gender [i.e. males (n = 105) and females (n = 94)] as the independent variable, and pretest score as the covariate. Results showed no significant outcome differences by gender holding pretest constant [$F(1, 196)] = 0.07$; $p = .80$, partial $\eta^2 < 0.001$].

4.2.4 | Learning by ethnicity

Similarly, we tested differences in learning physics as a function of ethnicity. An ANCOVA was computed, with post-test as the

TABLE 2 Descriptive statistics of pretest and post-test by condition

Condition	Pretest *M (SD)*	Post-test *M (SD)*	Gain *M (SD)*
Adaptive (n = 64)	11.77 (3.40)	12.23 (3.67)	0.47 (2.73)
Linear (n = 68)	11.82 (3.40)	12.41 (4.04)	0.59 (2.86)
Free choice (n = 67)	11.88 (3.81)	12.72 (3.94)	0.84 (3.11)
Control (n = 64)	11.61 (3.65)	11.59 (4.19)	−0.02 (3.03)
Total (N = 263)	11.77 (3.55)	12.25 (3.96)	0.48 (2.93)

dependent variable, ethnicity [i.e. White [n = 83], Black/African American [n = 62] or Hispanic [n = 15]) as the independent variable, and pretest as the covariate. Results showed no significant outcome differences by ethnicity holding pretest constant (F[2,156] = 1.54; p = .22, partial η^2 = 0.02).

4.3 | Game enjoyment

4.3.1 | Overall enjoyment

Having established that students' learning of physics improved via playing the game, we next wanted to ensure that this was not at the expense of game enjoyment. We defined game enjoyment as the combination of two game-survey items (i.e. '*I enjoyed the game very much*', and '*I'd like to play this game again*'). These statements were each rated on a scale of 1 (strongly disagree) to 5 (strongly agree). We computed the average of the two enjoyment items, which ranged from 1 to 5. Overall, students really enjoyed the game (M = 3.90, SD = 1.01, n = 195).

4.3.2 | Enjoyment by condition

We reported earlier that there were no learning differences as a function of game condition. But did students in each of the three conditions differ in terms of their enjoyment of the game? The means of enjoyment across all three conditions were strikingly similar (3.90 for adaptive and free choice, and 3.84 for linear), and the results of an ANOVA showed that there was no significant difference in enjoyment by condition [F(2, 192) = 0.08; p = .92, *partial* η^2 = 0.001].

4.3.3 | Enjoyment by gender

To test whether males and females differentially enjoyed the game, we computed an ANCOVA predicting game enjoyment by gender, holding pretest scores constant because males showed slightly higher pretest scores (M = 12.68, SD = 3.40) compared to females (M = 10.84, SD = 3.49). Results from the ANCOVA showed that both females (M = 3.90, SD = 0.99, n = 92) and males (M = 3.90, SD = 1.04, n = 103) equally enjoyed the game: [F(1, 192) = 0.36; p = .55, partial η^2 = .002].

4.3.4 | Enjoyment by ethnicity

We also tested for differential enjoyment as a function of ethnicity. We computed an ANCOVA with game enjoyment as the dependent variable, ethnicity (i.e. those self-identifying as White, Black/African American or Hispanic) as the independent variable and pretest as the covariate. As with the gender analysis, we included pretest score as a covariate, given some differences in incoming knowledge by ethnicity. The pretest scores by ethnicity were: White (M = 12.55, SD = 3.30, n = 83); Black/African American (M = 10.48, SD = 3.73, n = 62); and Hispanic (M = 12.13, SD = 2.67, n = 15). Results of the ANCOVA showed no significant enjoyment differences by ethnicity holding pretest constant [F(2, 154) = 0.78; p = .46, *partial* η^2 = .01]. The final enjoyment scores by ethnicity were: White (M = 3.87, SD = 1.03, n = 82); Black/African American (M = 3.77, SD = 0.96, n = 61); and Hispanic (M = 4.17, SD = 0.75, n = 15).

4.4 | Learning support effectiveness

As described in Section 3, we designed and developed eight different types of supports—five related to physics (i.e. physics animations, interactive definitions, Hewitt videos, glossary and formulas), and three related to solving game levels (i.e. worked examples, hints and tips on game mechanics). Students, regardless of condition, voluntarily accessed the learning supports in the current study via a 'Help' button in the bottom-right corner of the screen. Based on responses to the learning support satisfaction questionnaire, students found the learning supports to be helpful and liked having them in the game (see Appendix).

The first learning support question we examined was: which of the supports were accessed most frequently? The three most-frequently accessed supports included: *hints* (M = 5.10, SD = 5.41); *physics animations* (M = 3.94, SD = 4.57); and *worked examples* (M = 2.60, SD = 2.84). Hints and worked examples provided only game-level support. The physics animations targeted the underlying physics, thus we hypothesized that the physics animations would be the only support that correlated with our physics pretest and post-test scores. The correlations of the top three supports with our pretest and post-test scores confirmed this hypothesis: (a) *hints*—no correlation with pretest (r = .06) or post-test (r = .07); (b) *physics animations*—significant correlation with pretest (r = .32, p < .01) and posttest (r = .33, p < .01); and (c) *worked examples*—no correlation with pretest (r = .10) or posttest (r = .06).

We also computed a stepwise regression analysis predicting post-test score by pretest score, and the frequency data of all eight supports (i.e. physics animations, worked examples, hints, glossary, interactive definitions, formulas, Hewitt videos and game tools). The results showed a Multiple R = .71 (R^2 = .50); with just two predictors: pretest and physics animations in the model [F(2, 198) = 97.46; p < .001]—Pretest β = .66; t = 12.27, p < .001; Physics animations β = .11; t = 2.11, p = .04.

4.4.1 | Effects of physics animations on game performance

In addition to our more formal outcome measure of physics learning (assessed via an external pretest and post-test), another important measure relates to students' in-game performance. Towards that end, we examined data related to the following: total number of levels solved, total number of gold coins earned and total number of silver coins earned. We computed three regression analyses. Results showed that *Physics Animations* significantly predicted the levels completed [β = .43, $F(1, 197)$ = 45.15, p < .001, R^2 = .18], gold coins earned [β = .35, $F(1, 197)$ = 27.84, p < .001, R^2 = .12] and silver coins earned [β = .31, $F(1, 197)$ = 21.42, p < .001, R^2 = .10]. That is, for each 1SD change in *Physics Animations* frequency, the number of (a) levels completed increases by 0.43 SD, (b) gold coins increases by 0.35 SD and (c) silver coins increases by 0.31 SD. In summary, students watching more *Physics Animations* completed more levels and earned more gold and silver coins than those watching fewer *Physics Animations*.

4.4.2 | Effects of physics animations on game enjoyment

Viewing physics animations boosts both learning and game performance. But how does watching physics animations relate to overall enjoyment of the game? To address this question we computed another regression analysis with game enjoyment as the dependent variable and *Physics Animations* as the independent variable. Results showed that *Physics Animations* significantly predicted game enjoyment [β = .15, $F(1, 193)$ = 4.53, p = .04, R^2 = .02]. That is, for each one SD change in *Physics Animations* frequency, game enjoyment increases by 0.15 SD.

4.4.3 | Interaction of learning supports and game level difficulty

For our final analysis, we wanted to explore any interactions between viewing learning supports and game levels—specifically game level difficulty. We expected that students would tend to access learning supports (particularly physics animations, hints and worked examples) when playing the more difficult game levels. To test this hypothesis, we computed the average time spent playing each game level (i.e. average duration), which significantly correlated (r = .56, p < .001) with our composite difficulty index (i.e. the sum of game mechanics and physics difficulty indices, ranging from 1 to 10). On average, it took students longer to solve harder than easier levels. The average duration of playing game levels significantly correlated with accessing learning supports—i.e. physics animations (r = .29, p = .01), hints (r = .83, p < .001) and worked examples (r = .86, p < .001). These correlations suggest that students were, in fact, accessing more learning supports (both content- and game-related) in more difficult levels.

5 | SUMMARY AND DISCUSSION

One main goal of our research was to design and study a digital educational game that dynamically measures and supports the development of conceptual physics understanding for a diverse set of learners. The results of our study indicate that we met that challenge. That is, our findings showed significant overall improvement of students' conceptual physics understanding after approximately 4–5 hr of gameplay. Moreover, participants reported that they enjoyed playing the game. Importantly, there were no gender or ethnicity effects on the participants' learning or enjoyment of the game. These findings suggest that an educational game like *PP* can act as a versatile learning tool to enhance conceptual physics understanding for a diverse learner group, without sacrificing the fun of gameplay. One limitation of the game is that it currently presents the content in English.

5.1 | Stealth assessment in games for assessment and learning

An accurate and ongoing assessment (e.g. stealth assessment) is a prerequisite element to drive adaptive challenges and dynamic learning support in a digital educational game. The current findings provided empirical support for the validity of using stealth assessment to measure domain-specific competence, such as conceptual physics understanding, in a game-based learning setting. Overall, we found that the stealth assessment estimate of physics understanding, as well as the finer-grained estimates of the four specific competencies (i.e. force and motion, linear momentum, energy and torque), significantly and positively correlated with the external physics test scores. These findings support the argument that real-time stealth assessment can act not only as an authentic assessment for active learning, but also function diagnostically to guide the presentation of adaptive challenges and/or the delivery of dynamic learning supports during game-based learning (Shute, 2011; Shute, Ke, & Wang, 2017). Thus, our findings support prior research and the design conjecture that we can use well-designed digital games as vehicles for both assessment and learning (Shute et al., 2016; Shute & Ke, 2012).

5.2 | Adaptivity in level navigation

According to the theory of the zone of proximal development (Vygotsky, 1978), adaptive sequencing of game levels based on players' in-game performance can lead to higher levels of enjoyment and/or engagement, and eventually to higher levels of learning (Bontchev & Georgieva, 2018; Jagušt et al., 2018). However, the current study does *not* indicate a significant effect of adapting game levels to fit learners' needs relative to either game enjoyment or test performance. This result is consistent with the findings of Vanbecelaere et al. (2019) who similarly reported no difference in game-based learning as a function of adaptivity. It also supports other findings that a computer-based learning environment that makes navigation decisions for students (i.e. adaptive or linear) is not necessarily better than one that offers students choices to navigate (e.g. Black & Deci, 2000; Vansteenkiste et al., 2004). On the other hand, this finding does not support the argument for the benefit of computer-controlled adaptive level navigation in digital gaming (Sampayo-Vargas et al., 2013).

The finding suggests that the relative effectiveness on learning of adaptive versus non-adaptive sequencing (e.g. linear and free choice) in an interactive learning environment warrants further research. Note that, in the current study, adaptive sequencing was driven by the physics competency estimates per student. However, such estimates may, in addition, correlate with students' self-efficacy related to different game levels. In other words, students' non-adaptive gameplay sequence (especially in the free-choice condition) may somewhat concur with the computer-controlled adaptive gameplay sequence, thus making the conditional difference in game-based learning sequence minimized. Moreover, the linear sequencing condition in this study still allowed the students to navigate between levels, which again reduces the conditional differences in the game-based learning path. Future research should investigate different mechanisms in planning adaptive sequencing in gameplay, such as integrating individual learners' enjoyment state or frustration level as an additional index in the adaptivity algorithm.

It is also worth noting that there are multiple approaches to building adaptive control in a hypermedia learning environment (Kelly, 2008). The advantage of adaptively delivered versus learner control in a game-based learning environment should be further examined via a future study that examines alternative adaptivity strategies, such as adaptive presentation of learning supports or types of game tasks (e.g. sketching and manipulation game levels) in response to learners' competency and/or degree of enjoyment.

5.3 | Effective in-game learning supports

A large challenge of using games as a vehicle for domain-specific learning is to include content-related learning supports in games while not disrupting enjoyment (Ke, 2016; Tsai et al., 2013). Our exploration of the relative effectiveness of different types of in-game learning supports indicated that hints, physics animations and worked examples were the most frequently accessed. Among these three supports, only physics animations correlated with the post-test scores. Specifically, the various levels of physics animation usage significantly predicted physics post-test scores as well as various in-game measures (e.g. gold coins received). Our finding on the effective learning supports differs from Kao et al.' work (2017); in that, we did *not* find hints to be a particularly effective in-game learning support—particularly compared to the effects of the physics animations. But it does support their argument that the type of learning support and the degree of information provided within the learning support can affect students' learning.

Some possible reasons why the physics animations were such effective supports include the following. First, the physics animations were designed to have the same look and feel as the game (i.e. the videos were produced with *PP*'s level editor). Second, the physics animations included both physics knowledge and gameplay mechanics. They also included a common failed attempt followed by a successful attempt. This could help students with both their physics understanding as well as learning the game mechanics. The physics animations were also based on the tutorial levels (i.e. illustrating simple machines), comprising the primary means of solving the game levels. Finally, each physics animation was applicable across multiple levels that were linked to the video in terms of the shared competency. In summary, this learning support was designed to be content and gameplay related, and was able to successfully enhance students' physics understanding. This finding aligns with previous discussions on the importance of intrinsic integration of learning and gameplay in the design of game mechanics and support features (Ke, 2016).

Another goal of the current work was to figure out how to maximize learning with a game using various learning supports, without ruining the fun of gameplay. As mentioned earlier, this has been an ongoing challenge in educational games. Our results showed that the best learning support (i.e. physics animations) predicted physics post-test scores, but not at the expense of enjoyment. The number of physics animations viewed significantly predicted our composite measure of game enjoyment.

Finally, we found that students typically accessed more learning supports (both physics- and game-related supports) in the more difficult game levels. These findings show that our learning supports were designed and embedded in the game appropriately. That is, we expected students to access the supports when they really needed help, and that is what students did.

6 | CONCLUSIONS AND IMPLICATIONS

Overall, the current study has provided empirical evidence supporting the validity of game-based stealth assessment and the feasibility of using a digital educational game as an engaging and effective tool for both assessment and learning with a diverse set of learners. Participating students, regardless of gender or ethnicity, developed an increased understanding of physics and self-reported enjoyment with playing *Physics Playground*.

The best way to sequence levels or provide adaptive challenges in a digital game to optimize learning, however, is still unclear. In the current study, adaptivity was provided mainly through computerized-level sequencing based on the stealth assessment of a student's physics competency state. The future design of adaptive sequencing of game levels should be further investigated to include multiple facets of game-based states (e.g. cognitive and affective). Future research on adaptive challenges can also examine a micro-level personalization approach in which particular types of in-game learning supports are adapted to the students' game-based states. Designing in-game learning supports to improve both enjoyment and learning is another area ripe for investigation. Our research suggests that integrating subject-matter content and gameplay mechanics within learning supports will assist learning and promote learner-support interactions.

ACKNOWLEDGEMENTS

This work was supported by the US National Science Foundation (award number #037988) and the US Department of Education (award number #039019). We would also like to thank the team members who helped in this project—Adam LaMee, Seyfullah Tinger and Jiabei Xu.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Valerie Shute https://orcid.org/0000-0002-9179-017X
Fengfeng Ke https://orcid.org/0000-0003-4203-1203

REFERENCES

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.

American Physical Society. (2000). *Physics graphs & statistics*. Retrieved from https://www.aps.org/programs/education/statistics/degreesbyrace.cfm

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*, 740–756.

Bontchev, B., & Georgieva, O. (2018). Playing style recognition through an adaptive video. *Computers in Human Behavior*, *82*, 136–147. https://doi.org/10.1016/j.chb.2017.12.040

Chen, C.-H., & Law, V. (2016). Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. *Computers in Human Behavior*, *55*, 1201–1212. https://doi.org/10.1016/j.chb.2015.03.010

Clark, D. B., Virk, S. S., Barnes, J., & Adams, D. M. (2016). Self-explanation and digital games: Adaptively increasing abstraction. *Computers & Education*, *103*, 28–43. https://doi.org/10.1016/j.compedu.2016.09.010

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*(7–8), 555–575. https://doi.org/10.1080/08839510290030390

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

Csikszentmihalyi, M. (1997). *Finding flow*. New York, NY. Basic Books.

Entertainment Software Association. (2019). *2019 essential facts about the computer and video game industry*. Retrieved from https://www.theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/

Ferguson, C. J., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior*, *27*(2), 770–775. https://doi.org/10.1016/j.chb.2010.10.026

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*, London, UK: Palgrave Macmillan.

Georgiadis, K., van Lankveld, G., Bahreini, K., & Westera, W. (2018). *Accommodating stealth assessment in serious games: Towards developing a generic tool*. 2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games) (pp. 1–4). https://doi.org/10.1109/VS-Games.2018.8493409

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, *18*(1), 88–94. https://doi.org/10.1111/j.1467-9280.2007.01853.x

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology*, *22*(6), R197–R206. https://doi.org/10.1016/j.cub.2012.02.012

Hasanah, U., & Tsutaoka, T. (2019). An outline of worldwide barriers in science, technology, engineering and mathematics (STEM) education. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 193–200.

Huang, W. D., & Oh, E. G. (2018). Motivational support from digital game-based learning environments (DGBLEs) for scientific topics designed by novice end users. *Educational Media International*, *55*(2), 123–136. https://doi.org/10.1080/09523987.2018.1484043

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 1–8). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-3546-4_1

Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., Von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the children and technology project. *Computers in Human Behavior*, *28*(2), 370–376. https://doi.org/10.1016/j.chb.2011.10.006

Jagušt, T., Botički, I., & So, H.-J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education*, *125*, 444–457. https://doi.org/10.1016/j.compedu.2018.06.022

Johnson, E. K. (2019). Waves: Scaffolding self-regulated learning to teach science in a whole-body educational game. *Journal of Science Education and Technology*, *28*(2), 133–151. https://doi.org/10.1007/s10956-018-9753-1

Kao, G. Y.-M., Chiang, C.-H., & Sun, C.-T. (2017). Customizing scaffolds for game-based learning in physics: Impacts on knowledge acquisition and game design creativity. *Computers and Education*, *113*, 294–312. https://doi.org/10.1016/j.compedu.2017.05.022

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, *64*(2), 219–244. https://doi.org/10.1007/s11423-015-9418-1

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Kelly, D. (2008). Adaptive versus learner control in a multiple intelligence learning environment. *Journal of Educational Multimedia and Hypermedia*, *17*(3), 307–336.

Khamparia, A., & Pandey, B. (2018). Effects of visual map embedded approach on students learning performance using Briggs–Myers learning style in word puzzle gaming course. *Computers & Electrical Engineering*, *66*, 531–540. https://doi.org/10.1016/j.compeleceng.2017.12.041

Leemkuil, H., & De Jong, T. (2012). Adaptive advice in learning with a computer-based knowledge management simulation game. *Academy of Management Learning & Education*, *11*(4), 653–666. https://doi.org/10.5465/amle.2010.0141.

Min, W., Frankosky, M., Mott, B. W., Rowe, J., Smith, P. A. M., Wiebe, E., … Lester, J. (2019). DeepStealth: Game-based learning stealth assessment with deep neural networks. *IEEE Transactions on Learning Technologies*, *13*, 1–1, 325. https://doi.org/10.1109/TLT.2019.2922356

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

Moyer-Packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E. K., Roxburgh, A. L., … Clarke-Midura, J. (2019). How design features in digital math games support learning and mathematics connections. *Computers in Human Behavior*, *91*, 316–332. https://doi.org/10.1016/j.chb.2018.09.036

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education*, *69*, 452–462. https://doi.org/10.1016/j.compedu.2013.07.004

Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 59–78). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-39298-1_4

Shute, V. J. (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.

Shute, V. J., Almond, R. G., & Rahimi, S. (2019). *Physics playground* (Version 1.3) [Computer software]. Tallahassee, FL. Retrieved from https://pluto.coe.fsu.edu/ppteam/pp-links/

Shute, V. J., & 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. J., Ke, F., Almond, R. G., Rahimi, S., Smith, G., & Lu, X. (2019). How to increase learning while not decreasing the fun in educational games. In R. Feldman (Ed.), *Learning science: Theory, research, and practice* (pp. 327–357). New York, NY: McGraw Hill.

Shute, V. J., 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. J., Rahimi, S., & Chen, S. (2017). Measuring and supporting learning in educational games). In M. F. Young & S. T. Slota (Eds.), *Exploding the castle: Rethinking how video games and game mechanics can shape the future of education* (pp. 201–220). Charlotte, NC: Information Age.

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

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of portal 2 and lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 7–27). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139049580.004

Spann, C. A., Shute, V. J., Rahimi, S., & D'Mello, S. K. (2019). The productive role of cognitive reappraisal in regulating affect during game-based learning. *Computers in Human Behavior*, *100*, 358–369. https://doi.org/10.1016/j.chb.2019.03.002

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, *35*(8), 19–29.

Sun, C., Chen, L., & Chu, H. (2018). Associations among scaffold presentation, reward mechanisms and problem-solving behaviors in game play. *Computers & Education*, *119*, 95–111. https://doi.org/10.1016/j.compedu.2018.01.001

Tofig, D. (2017). *America's advanced mathematics and physics students in a global context*. National Center for Education Statistics Blog. Retrieved from https://nces.ed.gov/blogs/nces/post/america-s-advanced-mathematics-and-physics-students-in-a-global-context

Tsai, F. H., Kinzer, C., Hung, K. H., Chen, C. L. A., & Hsu, I. Y. (2013). The importance and use of targeted content knowledge with scaffolding aid in educational simulation games. *Interactive Learning Environments*, *21*(2), 116–128. https://doi.org/10.1080/10494820.2012.705852

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2019). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, 1–12. https://doi.org/10.1111/jcal.12416

Vandewaetere, M., Cornillie, F., Clarebout, G., & Desmet, P. (2013). Adaptivity in educational games: Including player and gameplay characteristics. *International Journal of Higher Education*, *2*(2), 106–114.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic role of intrinsic goals and autonomy-support. *Journal of Personality and Social Psychology*, *87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Ventura, M., Shute, V. J., & 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

Vygotsky, L. (1978). *Mind in society: The development of higher psychological process*, Cambridge, MA: Harvard University 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*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018

Yang, Y.-T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. *Computers & Education*, *59*(2), 365–377. https://doi.org/10.1016/j.compedu.2012.01.012

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, *82*, 61–89. https://doi.org/10.3102/0034654312436980

Zeglen, E., & Rosendale, J. A. (2018). Increasing online information retention: Analysing the effects of visual hints and feedback in educational games. *Journal of Open, Flexible and Distance Learning*, *22*(1), 22–33.

Zhao, W., & Shute, V. J. (2019). Can playing a video game foster computational thinking skills? *Computers & Education*, *141*, 103633. https://doi.org/10.1016/j.compedu.2019.103633

How to cite this article: Shute V, Rahimi S, Smith G, et al. Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *J Comput Assist Learn*. 2020;1–15. https://doi.org/10.1111/jcal.12473

APPENDIX A.

Game and learning support satisfaction questionnaire

Students responded to the 10-item game satisfaction and the 6-item learning support satisfaction questionnaires. Individual items and descriptive statistics are shown in the following table. Ratings per item were on a 1 (strongly disagree) to 5 (strongly agree) scale, and overall revealed student satisfaction with both the game and the supports.

TABLE A1 Game and learning supports satisfaction questionnaire

Game satisfaction	*M*	*SD*
I enjoyed the game very much	4.03	1.02
I thought the game was boring [R]	3.62	1.16
The game did not hold my attention [R]	3.53	1.16
I thought I performed well in the game	3.91	0.94
I was pretty skilled at playing the game	3.69	1.00
I put a lot of effort into solving levels	4.08	0.91
The game helped me learn some physics	3.96	0.91
Physics is fun and interesting	3.75	1.09
I would like to play this game again	3.72	1.16
I would recommend this game to my friends	3.43	1.18
Scale	3.77	0.70
Learning supports satisfaction		
The 'show solutions' helped me solve the levels	3.82	1.25
The 'show physics' helped me learn physics	3.60	1.09
The supports were generally annoying [R]	3.46	1.08
The supports were pretty easy to use	3.78	0.83
The supports did not help me at all [R]	3.56	1.09
I would rather solve levels without supports [R]	3.28	1.28
Scale	3.58	0.72

Note: [R] items were reverse coded for analysis. For example, the mean for the item '*I thought the game was boring*' is 3.62. Students generally disagreed with this item and thought the game was not boring.