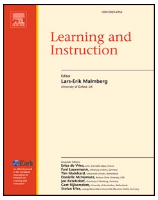
ARTICLE IN PRESS

Learning and Instruction xxx (xxxx) xxx

Contents lists available at ScienceDirect

Learning and Instruction

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



Does embedding learning supports enhance transfer during game-based learning?

Katie Bainbridge [a,*], Valerie Shute [b], Seyedahmad Rahimi [d], Zhichun Liu [b], Stefan Slater [c], Ryan S. Baker [c], Sidney K. D'Mello [a]

[a] Institute of Cognitive Science, University of Colorado, Boulder, USA
[b] Department of Educational Psychology and Learning Systems, Florida State University, USA
[c] Penn Center for Learning Analytics, University of Pennsylvania, USA
[d] School of Teaching and Learning, College of Education, University of Florida, USA

ARTICLE INFO

Keywords:
game-Based learning
Educational games
Learning supports
Physics playground
Animations

ABSTRACT

Educational video games are hypothesized to be good environments for promoting learning; however, research on conceptual learning from games is mixed. We tested whether embedding a learning support in the form of short animations illustrating physics concepts that can be used to aid gameplay improved learning. Ninety-six 7th to 11th grade students were randomly assigned to play *Physics Playground* with or without the learning supports over a 4-day period. Results indicate that students who played a version of the game with embedded learning supports showed more improvement on a far- ($d = 0.36$), but not on a near-transfer physics assessment ($d = 0.17$) compared to those who played without the supports. The learning supports did not affect students' enjoyment with the game. We conclude that the game-embedded animations were effective at promoting conceptual learning without sacrificing the fun of game-based learning.

Video games show great potential for learning and may one day be a valuable addition to classroom instruction (Gee, 2007; Kim & Shute, 2015; Mayer, 2014; Shute, Ke, et al., 2019, Young et al., 2012). However, educational game research is haunted by the following paradox: increasing the learning decreases the fun, and increasing the fun decreases the learning (Shute, Ke, et al., 2019; Wright, 2017). Learning outcomes for educational video games relative to traditional instruction are highly mixed and vary widely depending on the characteristics of the game (Mayer, 2014; Young et al., 2012). One way of improving the learning from an educational game is to add *learning supports,* which aim to scaffold players' learning and help them transfer knowledge and skills acquired in the game. However, many studies that successfully improve learning outcomes of an educational video game with learning supports utilize non-game elements (e.g. worksheets, vocabulary lists, advance organizers, reflective summaries) that players are unlikely to voluntarily engage with outside of a research context (Fiorella & Mayer, 2012; Parong & Mayer, 2018). The current study aims to address this problem by developing and testing a learning support that is seamlessly embedded within an educational game, feels like a natural part of the game, can aid with gameplay, and improves transfer of knowledge outside of the game context.

1. What is transfer?

Learning in a game does not benefit students in the real world unless they can apply what they learned in the game to a new context. This ability to apply knowledge outside of the context in which it was learned it is called transfer, and it can be surprisingly elusive to achieve (Singley & Anderson, 1989). There are multiple potential levels of transfer: specific, mixed, and general (Mayer, 2003). Specific transfer is limited to contexts that are very similar to the one in which the learning occurred, e.g., if the skills acquired playing chess on the computer only improves the ability to play chess in chess club. General transfer is the opposite extreme, in which learning in one context applies to all other related contexts, e.g., if playing chess improves IQ. Mixed transfer (sometimes called 'the specific transfer of general skills') is the middle ground between these two extremes; the generalizable skills used to play chess, such as thinking multiple steps ahead and keeping various outcomes in mind, can transfer to completely different contexts that use those skills in a similar way (Mayer, 2003). For example, perhaps making a schedule

---

* Corresponding author.
E-mail address: kbrainbridge@gmail.com (K. Bainbridge).

https://doi.org/10.1016/j.learninstruc.2021.101547
Received 11 April 2020; Received in revised form 25 August 2021; Accepted 20 September 2021

Please cite this article as: Katie Bainbridge, *Learning and Instruction*, https://doi.org/10.1016/j.learninstruc.2021.101547

for multiple employees (which requires mentally tracking multiple peoples' availability and comparing alternative arrangements) would be easier for a person who plays a lot of chess.

General transfer is largely agreed to be a myth, as previous attempts to demonstrate it have been unsuccessful or debunked (Thorndike & Woodworth, 1901; Sala & Gobet, 2017). Specific transfer is the easiest to demonstrate experimentally, but if specific transfer is as far as learning can extend, that implies that most classroom instruction will not be useful outside of the school context, which is a pessimistic view of education. It also defies intuition that learning is only ever applied in nearly identical contexts, and some researchers have suggested that the imbalance of evidence supporting specific transfer over mixed transfer may be the result of an inability to recognize transfer (Bransford et al., 1999). Mixed transfer, while difficult to elicit under experimental settings, does occur (Singley & Anderson, 1989), and is generally the goal for most multimedia learning environments such as game-based learning (Mayer, 2002). Accordingly, learning game designers and researchers need to demonstrate that the learning that occurs in the game will meaningfully transfer on assessments outside of the game. This is the goal of the present study, where we further differentiate between "near" mixed transfer and "far" mixed transfer (Bainbridge, 2019). If the new context in which the learning is applied is very similar to the game context, perhaps even bordering on specific transfer, that would be an example of near transfer. If the new context is sufficiently dissimilar to the game context, that would be an example of far transfer.

2. Theoretical and empirical foundations of game-based learning

Developmental psychologists have long suggested that play is a primary mode of learning for humans (Piaget, 1972), a framework that has inspired countless play and game-based curricula over the years (Montessori, 1976). However, Huizinga (1980) warned that focusing too much on education during play can hurt both. The most essential feature of play is pleasure, and once the learning becomes more important than the pleasure, the benefits of play are lost (Huizinga, 1980). Learning game designers must balance between ensuring their games promote learning without ruining the enjoyment that makes game learning an appealing pedagogical tool.

An important step towards this goal is to adopt a theory-based approach to game design. According to Mayer (2014), a game has five definable features. Games are: (1) rule-based – there is a knowable set of rules and/or causal systems; (2) responsive – the environment reacts to player actions; (3) challenging – they provide opportunities for success at progressively difficult tasks; (4) cumulative – the environment reflects the player's past actions and allows for goal setting; and (5) inviting – the environment is interesting and fun for the player.

Video games are media that apply these features to a computer or digital setting. Learning games (also called "educational games" or "serious games") are video games meant to promote learning, as opposed to a game that primarily aims to entertain. It is easy to see why learning games would lend themselves to an educational goal, as they share many elements with good learning environments (Shute, Rieber, & Van Eck, 2011). In particular, the *responsive* elements of games encourage active learning (Bransford, Brown, & Cocking, 2000; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001) and can provide immediate, individualized feedback (Shute, Ventura, & Kim, 2013; Vollmeyer & Rheinberg, 2005). The *inviting* elements of games stimulate interest and motivation (Gee, 2007; Rotgans & Schmidt, 2011). The *challenging* and *cumulative* elements of games provide adaptive difficulty so that players keep working at the upper limit of their ability (Vygotsky, 1987). In addition, challenges and dynamic feedback in games helps create an environment that can foster the sense of flow (Csikszentmihá lyi, 1990) and potentially cultivate the mindset that generates persistence despite failure (Dweck, 2008). Thus, the argument that educational video games should benefit learning outcomes has strong theoretical foundations. The evidence, however, is not so clear cut when it comes to learning valued academic content and skills assessed outside of the game environment (i.e., mixed transfer).

3. Do learning games support academic learning?

Whether video games are effective learning tools in practice is still up for debate and largely depends on the nature of the game (Mayer, 2014), and review papers reveal mixed outcomes (e.g., Hanus & Fox, 2015; Young et al., 2012). In terms of positive outcomes, well-designed learning games have been successful in promoting learning in domains ranging from visual-spatial abilities and attention (Green & Bavelier, 2007, 2012; Shute, Ventura, & Ke, 2015), persistence (Ventura, Shute, & Zhao, 2013), creativity (Blanco-Herrera, Gentile, & Rokkum, 2019; Jackson & Games, 2015), civic engagement (Ferguson & Garza, 2011), and academic content and skills (Coller & Scott, 2009; DeRouin-Jessen, 2008; Habgood & Ainsworth, 2011). Despite these promising results, in a review of nearly 300 articles, Young et al. (2012) found that only a handful of studies demonstrated a learning benefit for a group that played a STEM educational game relative to a group that received traditional instruction. They conclude there is little support for the academic value of educational video games for math and science. n some instances, participants in the educational game condition performed worse than those who received traditional instruction (Adams, Mayer, MacNamara, Koenig, & Wainess, 2012; Harris, 2008; Parchman, Ellis, Christinaz, & Vogel, 2000; Swaak, De Jong, & Van Joolingen, 2004). However, many of these cases can be attributed to method, rather than to the medium of games (Mayer, 2014).

Rather than speaking to whether video games are effective learning tools altogether, studies tell us that the *way* the game groups and control groups were implemented *in that instance* largely determined the outcomes, both good and bad. We can learn from the failures and successes of prior educational game interventions what elements do and do not aid learning.

For example, in a study comparing a discovery learning inspired physics game to expository text instruction, the hypothesis that the pure discovery environment of the game would enhance intuitive physics knowledge was not supported; the group that read expository text performed better than the game group on both definitional and intuitive physics questions with an overall effect size of $d = 0.43$ (Swaak et al., 2004). This indicates that while the opportunity for discovery learning *is* an appealing aspect of educational video games, there are caveats to how it should be implemented. Pure discovery learning, with minimal guidance and instruction, has been shown in multiple contexts to mostly benefit high-achieving learners with high prior knowledge, and to lead to worse outcomes for other learners (Kittel, 1957; Mayer, 2004; Papert, 1980). It is likely that the players needed more structure and guidance than Swaak et al. (2004) incorporated into their physics game for effective learning to occur.

In a study by Parchman and colleagues (2009), navy electronic technician trainees were taught basic electricity and electronics fundamentals with one of four methods: computerized drill and practice (digital flashcards), enhanced computer-based instruction (a lesson given digitally with text, narration, models, animations, and quizzes), an educational video game, and traditional classroom instruction. Trainees who were taught with the educational video game improved less from pretest to posttest than the flashcard and digital lesson conditions ($d = -0.79$). In this case, it is likely that the game put too much extraneous load on the players as it was a role-playing game that was heavy on narrative content. It is very possible that the educational content in this game (i.e., fundamentals of electricity) was not as memorable as the narrative content, and that players developed their schemas around the narrative rather than around the principles of electricity.

Along this vein, multiple studies have shown that adding too much irrelevant content in the pursuit of improving interest or fun will usually backfire (Harp & Mayer, 1997). Moreno and Mayer (2000) refer to this as *the coherence principle of multimedia learning*, which posits that as many elements as possible within the learning environment should be relevant to the learning goal, minimizing extraneous load. Indeed, there is converging evidence that a greater amount of narrative interest in an educational game (at least when the content is not narrative in nature as it would be in a history game) is detrimental to good learning outcomes in games (Pilegard & Mayer, 2016).

A study by Adams et al. (2012) corroborates the need to be wary of both discovery learning and excessive narrative in educational video games. They compared a discovery learning game about pathogens to a content-matched slideshow lesson. Participants in the game condition performed worse than the participants in the slideshow condition on assessments of retention ($d = 1.37$) and transfer ($d = 0.57$). In a second experiment they found that participants in a narrative-heavy learning game about electricity performed worse than participants in a content-matched slideshow condition on assessments of the electricity content ($d = -0.31$). The message of these studies is that learning games should provide players with sufficient guidance and avoid superfluous narrative.

One of the most common pitfalls in all genres of game-based learning is conflating improvement in the game for learning (Melby-Lervag, Redick & Hulme, 2016; Owen et al., 2010; Redick et al., 2013; Simons et al., 2016). In order for learning within the game to be meaningful, the player must be able to transfer that learning outside of the game. One of the biggest culprits of this conflation is the "brain-training" game platform *Lumosity*. The developers' research shows consistent gains for regular players on an internal measure of cognitive ability, the Lumosity Brain Index (LBI), which is heavily based on the training games themselves (Hardy et al., 2015). However, independent researchers find no evidence of transfer on independent measures (e.g., visual attention) (Bainbridge & Mayer, 2018; Schwartz, Chase, Oppezzo, & Chin, 2011). In order to demonstrate effectiveness, an educational game must be studied using measures of transfer, and ideally mixed transfer in an environment dissimilar enough to be considered far transfer.

In summary, the inconsistent outcomes in the game-based learning literature suggest that more work needs to be done to establish under what circumstances educational games successfully improve learning, especially when that learning is assessed outside of the game. The theoretical similarities between video game environments and good learning environments, along with some experimental evidence showing educational games can be effective, are encouraging and warrant continued research on incorporating learning games into classroom instruction. One promising line of research involves the use of learning supports within games.

4. Theoretical and empirical foundations of learning supports within games

Much as failures in the literature can teach us how not to implement video games for learning, successes in the research can help us identify what features do result in effective game learning outcomes. Research that aims to isolate exactly what features are effective is sometimes referred to as the *value added approach* (Mayer, 2014). In this approach, a feature that should theoretically improve an outcome of interest (e.g., learning, grit, motivation, interest) is added to an educational game. Researchers then compare the outcome of interest in players who play the standard version of the game to players who received the added feature.

Learning supports are features added to an educational game meant to improve learning outcomes. Overall, a meta-analysis of value-added studies in the educational game literature revealed that adding learning supports to educational games resulted in a $d = 0.37$ (sigma) effect size advantage for players who received supports over those who played a standard version of the game (Clark, Tanner-Smith, & Killingsworth, 2016). Multiple frameworks and categories of learning supports for educational games exist in the literature (for notable examples see Mayer, 2014 and Wouters & van Oostendorp, 2013). Here we focus on learning supports from Melby-Lervåg, Redick, and Hulme's (2016) framework that are particularly well suited to improving academic outcomes: pretraining and coaching.

Both pretraining and coaching use the principles of cognitive load theory (Paas & Sweller, 2014), which proposes that the amount of information an individual can simultaneously process in working memory is inherently limited. If more information than the learner can handle is presented at once, no learning (or worse, incorrect learning) will occur. There are three types of cognitive load: extraneous, essential and generative. Extraneous cognitive load is cognitive processing that does not serve the learning goal (e.g., an unrelated graphic). Essential cognitive load is the minimum processing that is required to represent the information in working memory. The more complex the topic, the more essential cognitive load is necessary, and the more processing capacity will be used. Generative cognitive load is processing required for deeper understanding or insight. Generative processing can't happen if the learner's capacity is taken up by extraneous or essential cognitive load. According to cognitive load theory, learning supports should work to minimize extraneous load and manage the amount of essential load in order to free up room for generative cognitive load. Both pretraining and coaching do just that.

Pretraining refers to prior instruction meant to enhance the player's ability to learn from the game. Complicated game mechanics can overwhelm learners with too much extraneous load, leaving few cognitive resources available for the essential processing needed to understand the content underlying the game (Paas & Sweller, 2014). This can be mitigated by acquainting players with the educational content prior to game play. Pretraining can also activate existing prior knowledge, which likewise improves learning outcomes (Mayer, 1983). Coaching is similar to pretraining, but occurs throughout the gaming experience. It also aims to reduce extraneous load, but is an ongoing process that continually encourages players to relate their game actions to the learning content rather than relying on the activation of prior knowledge, as in pretraining. This encourages players to keep the majority of their cognitive load on the essential processing rather than the extraneous elements of the game environment that don't relate to the learning goal.

Both types of learning supports can theoretically mitigate some of the failures of the educational games described above. Pretraining and coaching provide structure and guidance, which avoid the pitfalls of pure discovery learning (Adams et al., 2012; Swaak et al., 2004). Likewise, by highlighting the underlying educational content in the game, coaching and pretraining may reduce the likelihood that the player will pay more attention to the narrative or other superfluous details than to the learning content/objectives.

Several studies show advantages for adding pretraining to an educational game. For example, in a study on a farming simulation game, Leutner (1993) randomly assigned players to receive (a) a tutorial on farming concepts before they started playing, or (b) just play the game. The pretrained group performed better than the control group on a posttest of applied farming concepts with an effect size of $d = 0.55$. Similarly, Fiorella and Mayer (2012) found that players who received a worksheet listing electricity principles before they played a game in which they manipulated electrical conduction to solve puzzles scored better on a posttest of electricity concepts than players who did not receive the worksheet ($d = 0.77$).

Coaching has also been shown to be effective at improving learning outcomes in educational games. Using the same farming simulation game, Leutner (1993) reported that players randomly assigned to be connected to an online advice forum – provided warnings, corrections, and comments in response to their choices in the game – performed better on an applied posttest of farming concepts than those who played a base version of the game ($d = 0.85$). In a study by Cameron and Dwyer (2005), a group that receive explanatory feedback after every choice in a cardiophysiology simulation game performed better ($d = 0.57$) on a delayed assessment of cardiophysiology relative to the control group that did not receive explanations.

ter Vrugte et al., (2017) provide a notable example of adding coaching-style learning support to an educational game. They compared a group of students who played a proportional reasoning math game to a group that played the game with faded worked examples. They found that students who played the game with faded worked examples improved more than the control group on transfer measures with an effect size of $d = 0.60$. However, Vrugte and colleagues (2016) did not measure participants' subjective experience of the game, and forcing students to see solutions to the puzzles they were solving before they had a chance to figure it out for themselves could have made the game less fun to play. Educational game studies need to take enjoyment into account, because if the students are not enjoying the game, why use a game at all?

5. Embedded animations as learning supports

Existing approaches to embedding learning supports in games, despite being effective, have one major limitation. Many of the supports reviewed above took the form of a paper worksheet, a list of terms, or an interruption to the gameplay, such as prompts to summarize or reflect to promote generative processing. The majority of these interventions are likely only possible under supervised, controlled conditions like a laboratory study. The effectiveness of video games in promoting learning is greatly reduced if the likelihood of the student engaging with the very element (the learning support) that makes the game effective is minimal. Moreover, interrupting a fun game with academic-like activities (e.g., worksheet) reduces one of the biggest appeals of video games: the fun (Shute, Ke, et al., 2019).

Our goal was thus to incorporate aspects of pretraining and coaching without relying on elements outside of the game (like a worksheet) and without significantly disrupting the game experience (like a reflection). We selected game-based animations which combine elements of pretraining and coaching, depending on how and when the player uses it. The supports are short – about a minute long – and are directly embedded in the gameplay, so they should not feel disruptive or tedious, nor should they strain cognitive load. They are also designed to look and feel like a natural part of the game, and can give players insights into how to improve gameplay or reflect on their gameplay, while connecting what they do in the game with the underlying content.

Animations have been shown to be effective learning supports in many learning contexts, not just games. There is still some debate over whether animations are more effective than simply providing a static diagram (Mayer, Hegarty Mayer, & Campbell, 2005); it could be that the continuous motion adds more extraneous load than necessary, when a static image would suffice. However, a meta-analysis of 61 studies of learning from animations found an overall positive effect of $g = 0.23$ in favor of learning from animations over learning from static graphics (Berney and Bé trancourt, 2016). This effect size is increased when the animations feature narration ($g = 0.34$) and when the animations did not include redundant narration and text ($g = 0.88$) Both of these findings fit theories of multimedia learning (Mayer, 2009), which recommend taking advantage of both auditory and visual channels of information processing to minimize extraneous load and manage essential load (Paas & Sweller, 2014). Our animations were heavily informed by these multimedia principles. Table 3 details how these principles were incorporated into our animation design.

6. Current study

The aim of the current study is to test game-embedded animations as effective learning supports in a game. We implemented the supports in the context of *Physics Playground* (*PP*; Shute, Almond, & Rahimi, 2019), a video game shown to benefit secondary school students' intuitive, informal understanding of physics principles (e.g., predicting simple physical events). We conducted a classroom study to investigate whether adding a learning support intended to explicitly link students' gameplay experiences with physics concepts improves learning of the underlying physics concepts on *out-of-game assessments*. Students were randomly assigned to either play a version of *Physics Playground* with no supports or a version with the learning supports. We hypothesized that (1) students who received the learning supports would have higher scores on a physics transfer posttest than students who played the game with no supports. The learning supports were designed to feel more like a natural part of the game than in the hope that this would help the educational content of the game to feel like gameplay instead of like schoolwork and thus avoid the traditional trade-off between increasing learning and decreasing fun (Shute, Ke, et al., 2019). We additionally hypothesized that (2) the learning supports would not harm players' experience of the game, such that students who play *Physics Playground* with learning supports would not report lower enjoyment of the game than students who play a base version of *Physics Playground* with no supports. If the support investigated here is effective it could inspire future educational games to incorporate similar learning supports that embed elements of pretraining and/or coaching into the gameplay itself.

7. Method

7.1. Participants and design

Our sample consisted of 96 students from two different K-12 schools. School A was a private school located in Pennsylvania. and the students generally were of a relatively high socioeconomic status. The participants in School A ($n = 51$) were primarily white (80%) with a balanced gender distribution (49% female and 51% male). They were mostly enrolled in ninth and 10th grade. School B was a public school located in the southeastern U.S. The participants in School B ($n = 46$) were more diverse (31% White, 46% Black), and also had a balanced gender distribution (42% female, 51% male). Most of the participants in School B were enrolled in grades 7–8. Twenty-eight percent of the students at school B were eligible for free or reduced-price lunch, an indicator of socio-economic status.

All participants were randomly assigned to one of two conditions (School A: 25 to treatment and 26 to control; School B: 24 to treatment and 21 to control). They were compensated with a $30 gift card after completing the experiment. See Table 1 for more detail on the participants. IRB approval was obtained from both the University of

Table 1
Participant demographic information (number of participants in each cell).

	School A	School B
Sex		
Male	25	24
Female	26	19
Decline	0	2
Grade		
7th grade	0	21
8th grade	0	20
9th grade	38	4
10th grade	12	0
11th grade	1	0
Ethnicity		
Black	1	21
White	41	14
Asian	5	2
Hispanic	1	0
Mixed Ethnicity	4	5
Condition		
Control	26	21
Treatment	25	24

Pennsylvania and Florida State University; these institutions led the data collection. Parental permission was obtained with a signature and returned directly to the researchers. Student assent information was provided verbally in class before the study began and was also provided in a written form, which students signed and returned directly to the researcher shortly before the study began.

Our sample size was determined by availability of students and participating teachers for the four-day testing period. It also involved members of the research team traveling to data collection sites, which also restricted the data collection window.

7.2. Materials

Physics Playground. *Physics Playground* (*PP*) is a 2-dimensional, web-based computer 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 (Shute, Ke, et al., 2019). The game dynamically responds to players' interactions with the game. This responsiveness is accomplished via detailed formal simulation of a virtual physics "world" using actual, accurate physics formulas and calculations to account for mass, gravity, friction, momentum, and many other physics concepts.

The goal of the game is to move a green ball to hit a red balloon among a variety of obstacles. It is meant to be solved by players drawing simple machines (i.e., ramp, lever, pendulum, and springboard), that interact with the two-dimensional environment according to Newtonian mechanics. To illustrate, Fig. 1a shows screenshots of the level *Double Hoppy*. One solution, shown in Fig. 1a, is to draw an object above the ball and drop the object on the right side of the springboard to launch the ball to the balloon. The game has over 100 levels, but a subset of 28 levels was used for the current study. The first 5 levels are tutorial levels which introduce the participant to essential game actions such as nudging the ball and drawing simple machines like levers, pendulums, and springboards. After the tutorial, the levels differ in difficulty according to a variety of factors such as the distance between the ball and the balloon, number of obstacles presented, novelty of the problem, and the number of objects or parameters required to solve the level. The levels are organized into playgrounds, and while they are presented in a linear order, students can freely navigate among the various levels and playgrounds. Students earn gold or silver coins upon solving a level. Gold coins are earned if the student draws the minimum number of objects needed to solve the level. For example, in the *Double Hoppy* example above, if the student draws one object for the lever and one object for the ball/weight, and nothing else, they would receive a gold coin. If they used more than the minimum number of objects for a level, they received a silver coin.

Game elements in *PP* include ongoing feedback, interactive problem solving, and adaptive challenges. Moreover, because there is not just one correct "answer" to a problem in *PP*, and the game allows players to create their solutions by drawing objects that come alive, these features foster curiosity and discovery, which is not typically present in more traditional learning environments.

Learning Supports. The Learning Supports are short physics animation videos that illustrate a relevant physics concept (e.g., properties of torque) in the *Physics Playground* environment, accompanied by narration. They illustrate the physics principle underlying that level within the *PP* environment using formal physics terminology.

Players can use the targeted principle from the Learning Support to figure out how best to solve the level or to reflect on the physics principles underlying a successful or unsuccessful solution to a level. For example, consider the level *Double Hoppy* (see Fig. 1a), where the solution is to drop an object on the right side of a springboard to launch the ball into the air. On this level, the Learning Support would show a different *PP* level environment that demonstrates how bending a springboard can launch the ball into the air (see Fig. 1b). This Learning Support is intended to communicate that transferring energy to the ball with a springboard will help them solve the level, while also presenting the formal physics terminology that corresponds to that concept (i.e., elastic potential energy can transfer to kinetic energy).

Each Learning Support illustrates the specified physics concept using the relevant simple machine and appropriate physics terms. A narrator speaks over the animation, explaining how what is illustrated relates to the underlying physics concept. Keywords are printed over the animation for emphasis. See Fig. 1b for a frame-by frame depiction of the *EcT: Springboard* PA, including narration transcription.

The Learning Support design was informed by theories of transfer and multimedia learning (see Table 3). They are meant to use elements of pre-training and coaching to draw attention to the educational content and connect it with the gameplay. They combine narration and visualization in accordance with the multimedia, modality, and redundancy principles. The utilize signaling and segmenting to manage essential load and minimize extraneous load. Thus, the Learning Supports should be effective at improving learning outcomes in physics, which is what we test here.

We developed seven Learning Supports relevant to the levels selected for this study. Each is about a minute long. They fall under two overarching categories: *Energy can Transfer* (EcT) and *Properties of Torque* (PoT). In the EcT category there are four videos: *Lever*, *Pendulum*, *Ramp*, and *Springboard*. In the PoT category there are three videos: *Lever: Distance*, *Lever: Mass*, and *Springboard*. All seven videos are different, although they may cover related topics. To ensure each video was seen at least once, three of the videos were triggered to play at the beginning of a relevant level for the treatment group (e.g., when players in the support condition entered the level "Tetris", they would first be asked to watch the video "Energy can Transfer: Ramp" to see an abstracted explanation of the solution), and four of the videos were triggered to play at the end of a relevant level (e.g., when players solved the level "Uphill Battle" they were prompted to watch the video "Energy can Transfer: Pendulum" before they moved on). In addition, the treatment group had access to a Learning Support relevant to the solution of every level through the "Help" button in the lower right-hand corner of the screen.

Other game supports. Although the control condition did not have access to Learning Supports, either triggered automatically or available

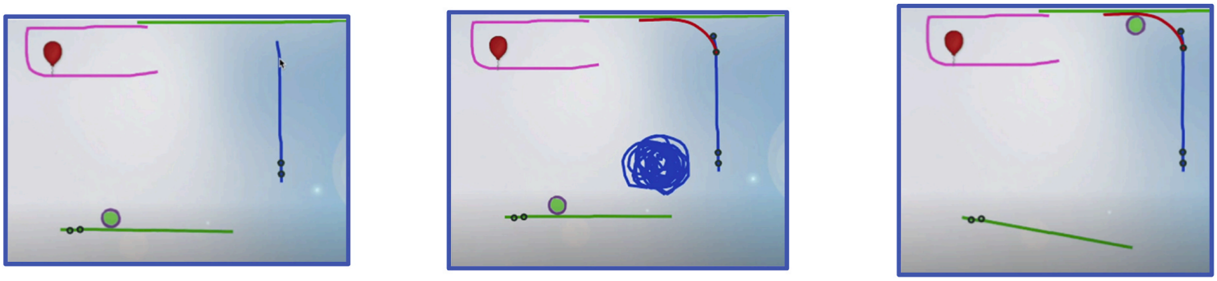
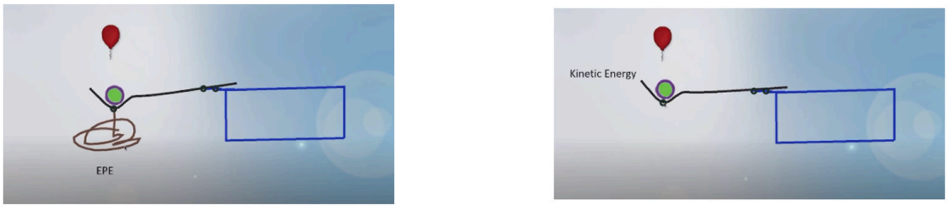


Fig. 1a. The level "Double Hoppy" can be solved by drawing an object and dropping that object on a springboard, which transfers the energy from the falling object to the ball so it can reach the balloon.

Narration: "Here you are going to see how to transfer energy to the ball using a springboard."

Narration: "You can store Elastic Potential Energy in a springboard by making it bend"



Narration: "Releasing the springboard transfers that energy to the ball."

Narration: "Then the ball will have Kinetic Energy, the energy of motion"."

Narration: "Making the springboard bend more will give it more Elastic Potential Energy"

Narration: "This time releasing the springboard will give the ball a greater amount of Kinetic Energy, making the ball go higher."

Fig. 1b. The Learning Support for "Double Hoppy" shows how a springboard can be used to solve a different level in Physics Playground. It illustrates that dropping an object on a springboard will transfer energy from the object to the ball, which will move the ball to the balloon. All Learning Supports can be seen on the Physics Playground *Youtube channel.*

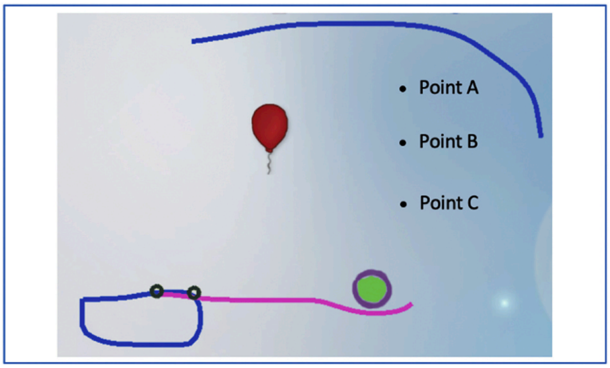


Fig. 1c. The near transfer items that assess transferring energy with springboards present the question in a Physics Playground level, using the same simple machine participants draw to solve levels. The goal on these assessments is the same as it is in the game: how best to get the ball to the balloon?.

through the help menu, they did have access to other resources when they were stuck. Participants in both conditions could click a "Help" button at the bottom right of the screen to gain access to Solution Videos (SVs) through a "show solutions" button or receive reminders of the controls and simple machines they learned about during the tutorial levels through a "Show Game Tips" button. SVs showed a video of the solution to that specific level with no voice over and no physics terms. If there was more than one solution to a level, more than one solution video was available.

Physics Knowledge Assessment. We created two comparable 14-item physics test forms for pretest and posttest, each of which included ten near-transfer test items and four far-transfer test items that were developed with the help of two physics experts. The near transfer items were designed with two principles: (1) reduction and simplification of the items and their choices, and (2) providing the items in the context of PP (i.e., including a video or an image from the PP environment (Fig. 1c). The similarity to the learning environment is why these items are considered "near" transfer. The far-transfer items were a subset (only EcT and PoT) from a conceptual physics test based on the Force Concept Inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992) that has been used in previous studies. These items do not have the look and feel of PP (Fig. 1d), nor did they have the look and feel of the Learning Supports. In fact, the Learning Supports are more similar to the near transfer items because they use the *Physics Playground* environment.

Game and Learning Supports Satisfaction Survey. We used a researcher-created 16-item, 5-point Likert scale questionnaire (1 = "low" to 5 = "high") to measure game enjoyment with ten items (e.g., "I enjoyed the game very much"), and satisfaction with the learning supports using six items (e.g., "The supports helped me understand the physics"). The questionnaire was identical between conditions with the exception of one question ("The 'Show Physics' option helped me learn physics") which was only provided to those in the treatment condition because the control group did not receive this option.

7.3. Procedure

The experiment in both locations consisted of four days of class time,

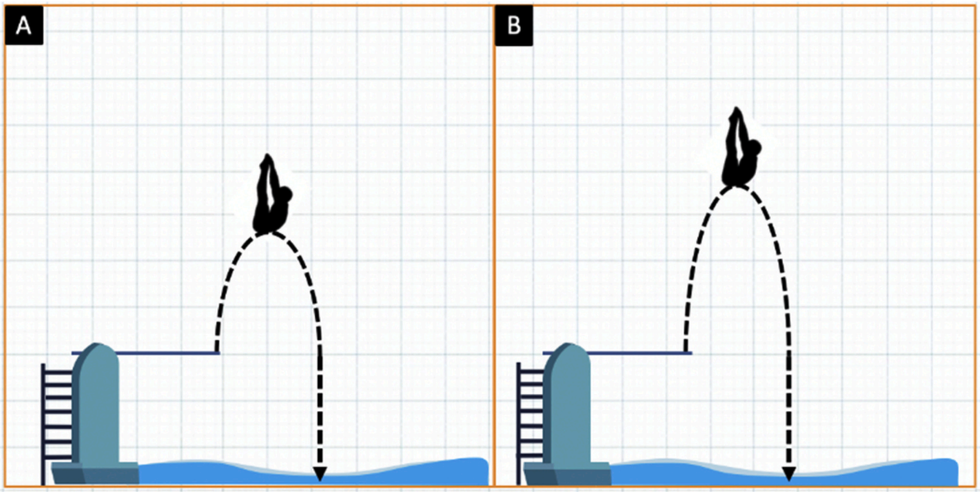


Fig. 1d. The far transfer items that involve springboards present the question in a new environment, and the springboards in question are not the same as the ones drawn in the game. The context of the questions is more natural, and the application of the physics concepts is more varied than in the near transfer items.

with four 45–50 min sessions per class. Participants at School A used MacBook laptop computers provided by the school. Participants at School B used Surface Pros provided by the researchers. All students were provided with a mouse and headphones, but if they preferred to use their own headphones that was permitted. All pretest and posttest measures were completed online using Qualtrics.

Day 1 consisted of participants completing a demographic survey and the physics knowledge pretest. This was followed by an introduction to PP tutorial that consisted of five “tutorial” levels that explicitly walked the player through how to use the game controls and draw the simple machines needed to solve levels. Participants then began the 28 gameplay levels targeting the concepts of Energy and Transfer and Properties of Torque. The levels were presented in a pre-determined order, but players could jump around to different levels or return to levels if they chose. Days 2 and 3 consisted of continued gameplay of those 28 levels. Day 4 consisted of some gameplay followed by the completion of the online posttest, game satisfaction survey, and compensation for participation.

Table 2
Means, standard deviations, and ranges for assessment performance by group.

Variable	Group	Mean	SD	Observed Range
Pretest Total (Range: 0, 14)	Control	8.00	2.46	2, 13
	Treatment	7.35	2.33	3, 13
Pretest Near Transfer (Range: 1, 10)	Control	5.83	2.05	1, 9
	Treatment	5.53	1.76	2, 9
Pretest Far Transfer (Range: 0, 4)	Control	2.17	0.82	1, 4
	Treatment	1.82	0.93	0, 4
Posttest Total (Range: 0, 14)	Control	9.57	1.60	6, 12
	Treatment	9.65	1.95	5, 14
Posttest Near Transfer (Range: 1, 10)	Control	6.64	1.24	4, 9
	Treatment	6.69	1.48	3, 10
Posttest Far Transfer (Range: 0, 4)	Control	2.94	0.85	1, 4
	Treatment	2.96	0.82	1, 4
Total Gain	Control	1.57	0.85	−3, 8
	Treatment	2.31	0.82	−3, 10
Near Transfer Gain	Control	0.81	2.08	−3, 6
	Treatment	1.16	1.97	−3, 7
Far Transfer Gain	Control	0.77	1.11	−2, 3
	Treatment	1.14	0.98	−1, 3
Enjoyment (Range: 1, 5)	Control	3.88	0.81	1.80, 5
	Treatment	3.94	0.86	2.20, 5

8. Results

Descriptive statistics for all measures are shown in Table 2. To ensure that pre-existing differences between subjects did not influence our results, we analyzed pretest scores across schools and conditions. A *t*-test revealed that there was a significant difference between School A and School B, $t(94) = 4.03$, $p < 0.01$, with School A performing significantly better ($M = 8.53$, $SD = 2.34$) on the pretest than School B ($M = 6.69$, $SD = 2.10$). While SES differences between the schools likely play a role, age likely underlies much of the performance gap between the schools, as the average grade level of a participant at School A was 2 years above the average grade level of participants in School B. A Mann-Whitney U

Table 3
Multimedia learning principles and how they relate to learning from physics animations (PAs).

Multimedia Learning Principle	Definition	Animation characteristic
Multimedia Principle	People learn better from words and pictures than from words alone.	PAs are narrated over a visualization of the concept.
Modality Principle, Redundancy Principle	People learn better from graphics and narrations than from animation and on-screen text. But graphics, narration AND on-screen text isn't helpful either. Graphics and narration is the best combination.	PAs do not have subtitles, but are instead a combination of narration and a dynamic visualization of that narration.
Signaling Principle	People learn better when cues that highlight the organization of the essential material are added.	Keywords are displayed on the screen at relevant times in relevant places. Arrows and other signaling tools aid in highlighting how the narration relates to the animation
Segmenting Principle	People learn better from a multimedia lesson is presented in smaller segments rather than as a continuous unit.	PAs are about a minute long and cover only one concept at a time
Coherence principle	People learn better when extraneous words, pictures and sounds are excluded rather than included.	PAs are designed to be minimal and to induce as little extraneous load as possible

Citation: Melby-Lervåg, Redick, & Hulme, 2016 indicated that the students at School B were significantly younger than the students at School A ($U = 76$, $p < 0.001$. Fortunately a Chi-Square test revealed that random assignment succeeded, and that School A and School B were equally represented between the two conditions, X^2 (1, $N = 96$) = 0.18, $p = 0.67$, so differences between schools should not influence the effectiveness of the intervention. Nevertheless, we included school as a covariate to serve as a proxy for both SES and age in subsequent analyses to control for these differences. We separated the assessment items into near, far, and overall (combination of near and far transfer items) scores (see Fig. 2).

The treatment group received an average of 5.53 ($SD = 1.71$; range 3–10) Learning Supports over the course of the experiment. Only Learning Supports that were engaged with (i.e. the participant didn't close it when it was triggered) are included in this count. Participants had the option to close a triggered Learning Support at any time.

Hypothesis 1. The support group will outperform the control group on the posttest

We first tested whether students showed improvement from pretest to posttest. Paired-sample t-tests indicated that both the support and the no-support groups improved significantly from pretest to posttest on both near, t(95) = 4.79, $p < 0.01$, and far, t(95) = 8.90, $p < 0.01$, transfer items. When the conditions are analyzed separately, t-tests show that both the control condition [$t(46) = -4.33$, $p < 0.01$] and the support condition [$t(48) = -6.70$, $p < 0.01$] improved significantly from pretest to posttest.

Next, we tested whether gain scores varied as a function of condition. We conducted an ANCOVAs using gain scores (near and far) as the dependent variable, group assignment as the independent variable, and school and posttest duration as covariates. Posttest duration was included as a covariate because there was considerable variability among participants in regards to how long they spent on the posttest (M = 484.40s, SD = 165.25s, Min = 287s, Max = 1165s). The posttest occurred on the fourth day of the study and was subject to participant fatigue and other motivation effects. A *t*-test showed that these potential motivational differences were not a function of what condition students were assigned to [$t(94) = -1.59$, $p = 0.12$], so we chose to control for this.

There was not a significant effect of condition on near transfer gain scores, F(1,92) = 0.77, p = 0.38, d = 0.17. However, there was a significant effect of condition on far transfer gain scores F(1,92) = 4.26, p = 0.04, d = 0.36 whereby the treatment group (M = 1.14, SD = 0.98) outperformed the control group (M = 0.77, SD = 1.11).

As a follow-up analysis, we correlated total duration of exposure to Learning Supports in minutes with gain scores for the support group only. The Learning Supports were triggered automatically, but were voluntary; a player could exit out of the Learning Support at any time. About 55% of the players exited at least one of the supports without watching, making number of Learning Supports viewed a continuous variable. Pearson correlations confirmed total duration of engagement with the Learning Supports was marginally correlated with overall gain (r = 0.25, p = 0.09) and significantly correlated with far transfer gain (r = 0.29, p = 0.04), but not correlated with near gain (r = 0.16, p = 0.28). Thus, exposure to the Learning Supports was indeed associated with greater improvement on physics knowledge assessments applicable outside of the game environment.

Hypothesis 2. Enjoyment will not be influenced by Learning Supports

A major drawback of adding learning supports to educational video games is the risk that they will decrease player enjoyment. To assess whether this was the case with the physics animation learning supports (PAs), we compared the groups on the enjoyment sub-items of the Game and Learning Supports Satisfaction questionnaire (Cronbach's alpha = 0.85) using an ANOVA with group as a between-subjects' factor and self-reported enjoyment as a dependent variable. The results revealed that players in the treatment condition (M = 3.94, SD = 0.86) reported comparable enjoyment to participants in the control condition (M = 3.88, SD = 0.81) who did not receive the Learning Supports, F(1,94) = 0.12, p = 0.74. Thus, adding the Learning Supports to gameplay did not decrease players' enjoyment of the game.

We also compared the groups on the satisfaction with supports sub items of the Game and Learning Supports Satisfaction Questionnaire. These supports include Game Tips and SVs for the control group and Game Tips, SVs and Learning Supports in the treatment condition. Players in the treatment condition reported comparable satisfaction with the supports (M = 3.88, SD = 0.78) to those in the control condition

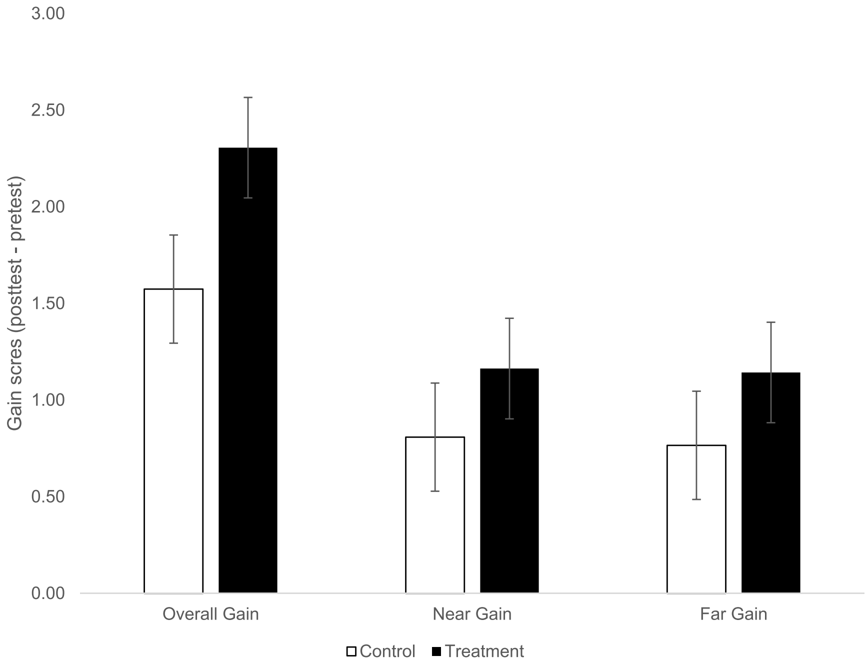


Fig. 2. Simple learning gains by condition. The error bars depict standard error.

(M = 3.81, SD = 0.81), F(1,94) = 0.22, p = 0.64. The presence of Learning Supports did not influence players' perception of the game or the help available to them.

9. Discussion

The present study sought to improve upon existing learning supports for educational video games. We developed animations illustrating physics concepts (PAs) and embedded them within *Physics Playground*, both as hints available through the help menu and as useful tips that were occasionally triggered on relevant levels. We hypothesized that participants who played a version of the game *Physics Playground* with Learning Supports would improve more on a measure of physics understanding than participants who played a standard version of the game. This hypothesis was partially supported. After controlling for school affiliation and time spent on the posttest, the group that played *Physics Playground* with physics animation Learning Supports demonstrated greater improvement on far transfer measures of physics knowledge with a medium effect size of $d = 0.36$. We also hypothesized that adding the embedded Learning Supports to *Physics Playground* would not decrease player enjoyment. This hypothesis was supported; participants who played the game with Learning Supports reported the same level of enjoyment as participants who played the game without Learning Supports.

These results suggest that the embedded physics animation learning supports developed for this game were effective at improving meaningful learning that was transferred outside of the game without the use of a more traditional academic support like a worksheet. The learning that the control group acquired was mostly applicable within the game environment, as evidenced by the lack of significant difference between the groups on near gain scores. The Learning Supports helped the treatment condition to transfer their experience to a situation distinct from *Physics Playground*. This is evidence of the elusive mixed transfer; these participants are more likely to be able to use knowledge obtained during gameplay on a different, more academic-like, context. This boost to the player's ability to use and apply the academic content underlying the game did not come at the cost of their enjoyment of the game.

Like all studies, ours has limitations. For one, we followed a *value-added* model rather than a *media comparison* model (Mayer, 2014) meaning that neither game condition was compared to another form of physics instruction, such as a classroom lesson. Although both the control and support conditions improved from pretest to posttest on the physics knowledge assessment, we cannot say whether this improvement was greater than it would have been had the students spent a comparable amount of time learning about the same concepts from a lecture, textbook, or some other form of instructional support.

The voluntary, "opt-in" nature of the supports as currently implemented limits the power of the comparison between groups, as over half of the treatment group (55.1%) watched fewer than five Learning Supports throughout the 4-day intervention. These participants' experience of the game would have been functionally very similar to that of control participants, muddying the waters of the group comparison. In addition, greater exposure to the Learning Supports may reflect underlying individual differences in that students who are more conscientious or more intrinsically interested in the content are more likely to choose to watch the supports. An experimental study taking away the element of choice, or a study that included measures for conscientiousness and interest and controlled for those factors would address this limitation.

Our sample size was informed by the practical limitations of gathering data in a school setting. The lack of significant difference between the support and control conditions on near and overall gain scores could be due to an inadequate sample. However, the effect size for near transfer was so low in our model that a power analysis could not be done ($\eta_p^2 = 0.00$), indicating that the insignificant difference between groups found in this study was not due to power. Similarly, the difference in enjoyment between groups produced an effect size of $d = 0.07$, which a power analysis conducted in G*Power ($\alpha = 0.05$, $1\text{-}\beta = 0.80$) suggested would require 4796 participants in order to detect a difference with a two-tailed test. Even if that many students *could* be recruited and a significant effect found, the difference would not be meaningful; the benefits to learning would still outweigh the small risk to enjoyment. A power analysis conducted in G*Power on far transfer scores ($\eta_p^2 = 0.043$, $\alpha = 0.05$, $1\text{-}\beta = 0.80$) recommended a total sample size of 182 students in order to detect a difference between the support and control group on far gain with a two-tailed test. That the difference between groups on far transfer gain was robust enough to be found, despite a power analysis suggesting we would need 86 more participants than were recruited, is noteworthy. Overall, while the sample size *is* limited, we have no reason to believe that more participants would have changed our results.

To address the limitations of the current study, future studies could include two control groups: one that plays a standard version of the game, and one that receives content-matched traditional instruction for a comparable amount of time. With this design, one can make both value-added and media comparison inferences from the results. A future study could also investigate the value of making the supports voluntary vs. involuntary. If transfer of learning really is tied to exposure to the learning supports, shouldn't a game with a learning goal *require* interaction with these features? Alternatively, removing the sense of choice and control over the learning experience could potentially reduce learner motivation, in accordance with Self-Determination Theory (Deci & Ryan, 2012; Patall, Cooper, & Wynn, 2010; Schrader & Nett, 2018). A design that compared a standard version of the game with no supports, a version of the game with voluntary supports, and a version of the game with mandatory supports, and then used both physics understanding and motivation/effort as dependent variables would address this question. Finally, replication with a larger sample is warranted.

The current study implemented the Learning Supports in three ways: before a relevant level, after a relevant level was solved, or as "hints" given during the level at the player's whim; however, *when* a player sees a learning support may have a profound impact on how and what the player learns. Evidence from the Advance Organizer literature (Fensham & West, 1976; Mayer, 1983), the Adjunct Question literature (Rothkopf, 1966; Rickards & Di Vesta, 1974), and the Productive Failure literature (Kapur, 2016; Schwartz, Chase, Oppezzo, & Chin, 2011) all suggest varying beneficial outcomes with different theoretical mechanisms for placing learning support material before vs. after an active learning exercise (of which an educational game is an example). A study design comparing the learning outcomes when relevant learning supports are given before vs. after a level would address which mechanism is best suited to learning in this circumstance.

n summary, the current study aimed to develop and test a learning support embedded in an educational game that was organic to the game environment, did not disrupt game play, and felt more like part of the game than learning supports of the same genre (pretraining and coaching) had in previous interventions. The results indicate that this aim was successful. Adding a physics animation video illustrating the underlying physics principles relevant to the solution for the current level improved far transfer learning outcomes in 7th-10th grade students who played *Physics Playground* for approximately 3 h, without decreasing their enjoyment of the game. The implications for future educational game design await further research, such as whether or not such learning supports should be voluntary and when the learning supports should be presented. However, the current study does suggest that embedded learning supports administered within the game environment may be a fruitful direction for other educational games to explore.

CRediT authorship contribution statement

Katie Bainbridge: Writing – original draft, Formal analysis, review & editing. **Valerie Shute:** Funding acquisition, Conceptualization,

Methodology, Formal analysis, review & editing. **Seyedahmad Rahimi:** Conceptualization, Methodology, Data collection, Software, review & editing. **Zhichun Liu:** Conceptualization, Methodology, review & editing. **Stefan Slater:** Conceptualization, Data collection, review & editing. **Ryan S. Baker:** Funding acquisition, Conceptualization, review & editing. **Sidney K. D'Mello:** Funding acquisition, Conceptualization, Formal analysis, review & editing.

Acknowledgments

We wish to express our gratitude for the funding by the US Department of Education (IES R305A170376) and the US National Science Foundation (NSF #037988) for generously supporting this research. We also want to thank Ginny Smith, Renata Kuba, Chih-Phu Dai, and Yeyu Wang for their help with this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2021.101547.

References

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235.

Bainbridge, K. (2019). *The role of variety in the transfer of cognitive skills*. UC Santa Barbara): Doctoral dissertation.

Bainbridge, K., & Mayer, R. E. (2018). Shining the light of research on Lumosity. *Journal of Cognitive Enhancement, 2*(1), 43–62.

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education, 101*, 150–167.

Blanco-Herrera, J. A., Gentile, D. A., & Rokkum, J. N. (2019). Video games can increase creativity, but with caveats. *Creativity Research Journal, 31*(2), 119–131. https://doi.org/10.1080/10400419.2019.1594524

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research In Education, 24*, 61–100.

Cameron, B., & Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research, 16*(3), 243–258.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86* (1), 79–122.

Coller, B. D., & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*(3), 900–912.

Csikszentmihályi, M. (1990). The domain of creativity. In M. A. Runco, & R. S. Albert (Eds.), *Theories of creativity: Vol. 115. Sage focus editions* (pp. 190–212). Sage Publications, Inc.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd.

DeRouin-Jessen, R. (2008). Electronic theses and dissertations. Unpublished manuscript. In *Game on: The impact of game features in computer-based training*. Orlando, FL: University of Central Florida.

Dweck, C. S. (2008). *Mindsets and math/science achievement*. New York, NY: Carnegie Corp. of New York–Institute for Advanced Study Commission on Mathematics and Science Education.

Fensham, P. J., & West, L. H. T. (1976). Prior knowledge or advance organizers as effective variables in chemical learning. *Journal of Research In Science Teaching, 13* (4), 297–306.

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.

Fiorella, L., & Mayer, R. E. (2012). Paper-based aids for learning with a computer-based game. *Journal of Educational Psychology, 104*(4), 1074–1082.

Gee, J. P. (2007). *Good video games+ good learning: Collected essays on video games, learning, and literacy*. Peter Lang.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88–94.

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology, 22*(6), R197–R206.

Habgood, M. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal Of the Learning Sciences, 20*(2), 169–206.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & education, 80*, 152–161.

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, et al. (2015). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. *PloS One, 10*(9).

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92.

Harris, D. (2008). *A comparative study of the effect of collaborative problem-solving in a Massively Multiplayer Online Game (MMOG) on individual achievement* (pp. 1–199). University of San Francisco.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141–151.

Huizinga, J. H. (1980). *Homo ludens. Routledge*.

Jackson, L. A., & Games, A. I. (2015). Video games and creativity. *Video Games and creativity* (pp. 3–38). Academic Press.

ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2017). Computer game-based mathematics education: Embedded faded worked examples facilitate knowledge acquisition. *Learning and Instruction, 50*, 44–53.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.

Kim, Y. J., & Shute, V. J. (2015). Opportunities and challenges in assessing and supporting creativity in video games. In G. Green, & J. Kaufman (Eds.), *Research frontiers in creativity* (pp. 100–121). San Diego, CA: Academic Press.

Kittel, J. E. (1957). An experimental study of the effect of external direction during learning on transfer and retention of principles. *Journal of Educational Psychology, 48*, 391–405.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3* (2), 113–132.

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology, 75*(1), 40.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press.

Mayer, R. E. (2003). *Learning and instruction*. Prentice Hall.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*(4), 256.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512–534.

Montessori, M. M., Jr. (1976). *Education for human development: Understanding montessori*. New York: Schocken Books.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117.

Owen, A. M., Hampshire, A., Grahn, J. a, Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature, 465*(7299), 775–778. https://doi.org/10.1038/nature09042

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning, 27*, 27–42.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*(1), 73–87.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal Of Educational Psychology, 110*(6), 785.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102*(4), 896.

Piaget, J. (1972). *Development and learning. Readings on the development of children*.

Pilegard, C., & Mayer, R. E. (2016). Improving academic learning from computer-based narrative games. *Contemporary Educational Psychology, 44*, 12–20.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology. General, 142*(2), 359–379. https://doi.org/10.1037/a0029082

Rickards, J. P., & Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*(3), 354–362.

Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction, 21*(1), 58–67.

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3*(4), 241–249.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science, 26*(6), 515–520.

Schrader, C., & Nett, U. (2018). The perception of control as a predictor of emotional trends during gameplay. *Learning and Instruction, 54*, 62–72.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*, 759–775.

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

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. *Learning Science: Theory, Research, and Practice, 327*.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games and learning. *Trends and issues in instructional design and technology, 3*, 1–31.

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.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's playground. *The Journal of Educational Research, 106*(6), 423–430.

Simons, D. J., Boot, W. R., Charbness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., et al. (2016). Do brain training programs work? *Psychological Science in the Public Interest, 17*(3), 103–186.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill (No. 9)*. Harvard University Press.

Swaak, J., De Jong, T., & Van Joolingen, W. R. (2004). The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge. *Journal of Computer Assisted Learning, 20*(4), 225–234.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *Psychological Review, 8*(4), 384.

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.

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning And Instruction, 15*(6), 589–602.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 11*(4–5), 381–419.

Vygotsky, L. S. (1987). The development of scientific concepts in childhood. *The collected works of LS Vygotsky, 1*, 167–241.

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.

Wright, M. T. (2017). *Paradox of the learning game: The promise and plight of video games and learning*. Philadelphia, PA: Design Meets Mind.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., et al. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89.