Computers & Education 190 (2022) 104600

Contents lists available at ScienceDirect

Computers & Education

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

Timing of learning supports in educational games can impact students' outcomes

Seyedahmad Rahimi [a,*], Valerie J. Shute [b], Curt Fulwider [b], Katie Bainbridge [c], Renata Kuba [b], Xiaotong Yang [b], Ginny Smith [b], Ryan S. Baker [d], Sidney K. D'Mello [c]

[a] *University of Florida, School of Teaching and Learning, Gainesville, USA*
[b] *Florida State University, Educational Psychology & Learning Systems, Tallahassee, USA*
[c] *University of Colorado Boulder, Department of Computer Science, USA*
[d] *University of Pennsylvania, Graduate School of Education, Philadelphia, USA*

A R T I C L E I N F O

Keywords:
Game-based learning
Educational games
Physics playground
Learning supports

A B S T R A C T

Learning does not automatically occur by playing educational games; instead, learning opportunities should be carefully designed in such games. For instance, research has indicated the importance of embedding learning supports within educational games to promote learning and other outcomes (e.g., enjoyment). However, more research is needed to determine when it is best to provide the supports—before or after attempting a game level? We investigated this question in a game called *Physics Playground* where we randomly assigned 149 students (M_{age} = 14, *SD* = 0.96) to receive learning supports—short videos—either immediately *Before* (n = 50) or *After* (*n* = 46) students worked on solving game levels. We also included a no-support *Control* (*n* = 53) condition. We found that students assigned to the two treatment conditions visited fewer game levels, but spent more time per level, and reported lower frustration levels than those assigned to the *Control* condition. And although students in the *After* condition had lower in-game performance measures than those in the *Control* condition, they achieved higher near- and far-transfer scores on the posttest after controlling for gameplay success and pretest scores. Thus, there appears to be some tradeoffs with respect to the inclusion and timing of learning supports. There were no major differences between the treatment conditions regarding learning and subjective measures. The findings of this study can help advance the design of educational games that are intended to enhance students' learning.

1. Introduction

There is ample evidence that well-designed digital games can improve learning (e.g., Byun & Joung, 2018; Clark et al., 2016; Lei et al., 2022; Shute et al., 2020a; Tsai & Tsai, 2018, 2020). Well-designed games include clear goals, incremental or adaptive challenges, as well as ongoing feedback, which combine to potentially induce a state of *flow*, or engagement so strong that one loses track of time (Csikszentmihalyi, 1990; Shute & Ke, 2012). Learning does not automatically occur as a byproduct of playing well-designed educational games; instead, learning opportunities are intricately designed into the game and are guided by relevant theories. For example,

* Corresponding author. 0501D, University of Florida, College of Education, School of Teaching and Learning, 2821 Norman Hall, PO Box 117048, Gainesville, FL, 32611, USA.

E-mail address: srahimi@ufl.edu (S. Rahimi).

https://doi.org/10.1016/j.compedu.2022.104600
Received 28 April 2022; Received in revised form 12 May 2022; Accepted 22 July 2022
Available online 8 August 2022
0360-1315/

self-determination theory (SDT) states that people become intrinsically motivated when their psychological needs for competence (i.e., that one can succeed), autonomy (i.e., ownership of one's behavior), and relatedness (i.e., feeling connected to others) are satisfied (Black & Deci, 2000). One way to nurture students' intrinsic motivation when playing educational games is by including learning supports—i.e., external instructional techniques that facilitate students' cognitive processes when playing educational games (Wouters & van Oostendorp, 2013). Well-designed supports in well-designed educational games can enhance learners' sense of competence and, in turn, their motivation (according to SDT). In-game learning supports can help learners acquire the content knowledge and, as a result of learning, show better progress in the game (Kuba et al., 2021b; Rahimi et al., 2021).

Accordingly, research shows that including learning supports in educational games can help facilitate learning (e.g., Cai et al., 2022; Vrugte et al., 2015; Wouters & van Oostendorp, 2013). In their meta-analysis of learning supports in games (e.g., video solutions to game problems, advice/hints, and worked examples), Wouters and van Oostendorp (2013) found that supports aided the transfer of learning from educational games to real-world contexts with an average effect size of $d = 0.34$ across 29 studies. Similarly, Cai et al. (2022) conducted a meta-analysis based on 49 studies and found that adding instructional support is an effective way to improve students' learning outcomes when playing digital learning games (effect size $g = 0.43$). Moreover, Cai and colleagues reported that the effects of the supports could be moderated by grade level of the learners and the type of game they played. The design features of instructional support (e.g., when to deliver the supports), however, have not been taken into consideration as possible moderating variables.

That said, not all supports are created equally—as poorly-designed supports can interrupt the flow of gameplay and reduce a learner's sense of autonomy (Adams & Clark, 2014; Habgood & Ainsworth, 2011; Habgood et al., 2005; Rahimi et al., 2021). Recently, researchers have been testing different design choices for learning supports intended to maintain flow and improve learning (e.g., Kuba et al., 2021a; Shute et al., 2020b). One design issue is whether the supports are tightly integrated within the learning game (e.g., Bainbridge et al., 2022) or are external to the game (e.g., a worksheet). Another issue is whether the supports are voluntarily accessed by students (e.g., Shute et al., 2020a) or delivered by the game when deemed appropriate (Cloude et al., 2021). In this study, we test a third significant design issue that has yet to be appropriately addressed—that is, the timing of supports; we ask when is the best moment to deliver learning supports—before or after playing a game level?

2. Delivery timing of learning supports

As reviewed below, the case can be made for providing learning supports either before or after engaging a learning activity (e.g., solving a problem or playing a game level). Supports can also be delivered within a game level, but that is outside the scope of this study as that disrupts gameplay.

2.1. Providing learning supports before playing a game level

Multiple lines of educational research, including the literature on adjunct questions and advance organizers (Rickards & Di Vesta, 1974; Rothkopf, 1966), suggest that placing learning supports before the start of the level will aid learning. Learning supports presented before learners engage in a learning task can serve as advance organizers (Barzilai & Blau, 2014; Liao et al., 2019), which refer to presenting a concrete model of principles to be learned before a learning activity (e.g., a diagram, key terms, or an outline of main points). In one relevant study, Barzilai and Blau (2014) compared the effects of three conditions in an educational math game: receiving external learning supports (i.e., an online tool outside of the game) before playing, after playing, and gameplay without supports. Results revealed that students who received external supports before playing scored higher on the posttest than students in the other conditions. The authors argued that the external learning supports presented before gameplay helped students integrate the new material (i.e., game problems) with their prior knowledge schemas.

Relatedly, the adjunct question literature suggests that placing learning supports before a game level may cue "forward effects." That is, the player pays attention to the structural features of what they are about to do or see (e.g., how the level relates to pendulums in a physics game) and ignores the incidental features of the level (e.g., how to get around a tricky obstacle) (Rickards & Di Vesta, 1974; Rothkopf, 1966). Specifically, receiving a preview of the main idea prior to learning will boost learners' recall for conceptual information and transfer, as well as reduce verbatim recognition (Fensham & West, 1976).

Along the same lines, Tsai et al. (2013) examined the effects of an educational game called TANK-Q on students' learning about the principles of projectile motion. The authors compared posttest scores between three conditions: receiving learning supports (i.e., explanations of the concepts of projectile motion with pictures, text, and formulas) before and during gameplay, during gameplay only, and no supports. They found that students who received learning supports before and during gameplay performed better on the posttest and had higher levels of accuracy than students who only received supports during gameplay or received no learning supports. Moreover, results showed that receiving learning supports before and during gameplay did not interfere with game enjoyment. Also, students who received supports before and during gameplay tended to spend more time using the in-game supports at the beginning of gameplay and almost none at the end as they no longer needed the supports. In contrast, students who received supports only during gameplay showed an increasing trend in support usage (i.e., using more supports as the game went on). The authors argued that spending more time using supports before gameplay could promote accuracy in answers about the target content. When integrating into other curricular activities, gameplay tended to be arranged after instruction (Pando Cerra et al., 2022; Ristanto et al., 2022). That is, students first learn new knowledge from direct instructions such as lecture, then practice and apply such knowledge in gameplay. Pando Cerra et al. (2022) have argued that doing so might boost students' motivation as they acknowledge the relevance of learned knowledge in solving game problems. It is important to note that the studies above investigated the provision of learning supports either before and during vs. during (Tsai et al., 2013) or before vs. after a gameplay session (Barzilai & Blau, 2014), or only before a gameplay session (Pando Cerra et al., 2022), and the supports in both studies were external to the game environment (e.g., reviewing learning content in an online learning system).

In the current study, we focused on providing supports internally (i.e., integrated into the game environment) rather than externally to the game. In addition, the supports were delivered either before or after playing a game *level* as opposed to a complete session of gameplay, thus we can gain insights from prior studies while acknowledging the differences between the earlier designs and the current study. For instance, Gresalfi and Barnes (2016) compared students' performance in a math game when receiving learning supports either after their final solutions or in the early stages while generating their initial guesses, but before engaging with the problems. Students receiving supports before engaging with the math problems demonstrated higher learning gains and a more accurate understanding of mathematical solutions (e.g., using math to justify their decision) than students who received supports after their solution. Feedback after the solution was likely viewed as the "end" of the activity, and no further action was needed. Thus, the early supports served as a resource for supporting problem-solving by indicating what students should do next.

From the perspective of cognitive load theory (Sweller, 2011), providing supports before a learning activity can reduce extraneous processing by guiding learners to select appropriate information and apply it to the problem at hand (Liao et al., 2019). Liao and colleagues investigated the effects of watching an instructional video—which explained key underlying content before gameplay—on students' cognitive load when playing a digital physics game. Results from a self-reported survey on cognitive load indicated that supports significantly increased students' germane cognitive load (i.e., cognitive capacity used in coding new information to the current schema), suggesting the supports may help students effectively manage their cognitive resources.

Although providing supports before a game level sounds promising in relation to promoting learning and performance, Gee (2005) has argued that students may try to evade such supports when they are focused on playing and winning the game. In addition, Fensham and West (1976) noted that advanced organizers are most effective for learners who lack prior knowledge. So, for students who score relatively high on pretests, the benefits of advance organizers may be reduced or eliminated. Advance organizers may also deprive high-achieving students from the beneficial effects of discovery learning (Bruner, 1971) and productive failure (Kapur, 2012).

2.2. Providing learning supports after playing a game level

Providing learning supports after the student attempts a level draws on literature related to the adjunct question (Rickards & Di Vesta, 1974; Rothkopf, 1966), formative feedback (Shute, 2008), and productive failure (Kapur, 2016; Kapur & Bielaczyc, 2012; Schwartz et al., 2011; Westera, 2022). According to the adjunct question literature (which supports both the before and after placement of supports for different outcomes), providing supports after a learning event (e.g., when a learner solves a game level) may cue "backward effects." That is, the supports can prompt the student to reflect on their actions in light of the new information (from the supports), prompting integration (Rickards, 1979). Students receiving supports after a game level have been found to have a greater recall for structural or conceptual content relative to those without such supports (Rickards & Di Vesta, 1974; Rothkopf, 1966).

Some researchers have found that learning supports designed to prompt students' reflection on their gameplay (e.g., "Please pause for a moment and write an explanation about what you did in this game level") might be necessary for a meaningful game-based learning experience (Cloude et al., 2021; Kiili, 2005; Moreno & Mayer, 2002). For example, Cloude et al. (2021) examined the effects of students' reflections on solving game levels in a science game. Students were prompted to reflect on the topics they learned as they progressed in the game, which improved problem-solving outcomes. However, Vrugte et al. (2015) found no significant effect of adding reflection prompts on students' learning when playing an educational game focused on ratios and proportional reasoning. Providing learning supports after playing a game level might produce reflection, without the need to prompt the students to reflect on their solution.

The type of learning support provided after each game level might also influence game performance. For instance, O'Neil et al. (2014) found that in a math game, only prompts designed to promote essential and generative processing after each level enhanced game performance. These prompts aimed to help students connect the game elements to the mathematical elements. The authors argued that the prompts that were too simple (e.g., a recall question to capture learners' attention) created minimal cognitive processing, whereas those that were too complex (e.g., an abstract reasoning question) produced extraneous processing. Similarly, Kuba et al. (2021a) found that only learning supports that connected the game mechanics to the underlying target knowledge and were designed to reduce extraneous load while promoting generative processing, predicted posttest scores and game performance.

Presenting supports after learners complete a task (either successfully solved or not) can also serve as valuable feedback (Conati et al., 2013; Gresalfi & Barnes, 2016). If an attempt was unsuccessful, the learning support could be corrective feedback (i.e., providing the correct answer or solution), and if the attempt was successful, the learning support could be simply verification feedback (i.e., noting the solution was correct)—both of which have been shown to be effective in improving learning across a wide range of environments (Shute, 2008).

The productive failure literature similarly argues that it is beneficial to deliver supports after a learning activity (Kapur, 2012; 2016; Schwartz, 2011). This research generally suggests that students start by engaging in solving problems involving concepts they have yet to learn, followed by consolidation and instruction on the targeted concept. With delayed supports, students might experience difficulties or frustration; however, they gain the opportunity to explore different possibilities on their own (Kapur, 2012). Delayed supports should prepare students to learn better from subsequent experiences than those students receiving instruction from the beginning (Kapur, 2012). While Kapur and colleagues (2012) found that providing supports after students' attempted solutions did not facilitate their problem-solving performance during learning, it did enhance their conceptual understanding and knowledge transfer on the posttest compared to the students who received learning supports before solving problems. Therefore, it is important to distinguish students' in-game performance from their learning when playing educational games, and provide adequate feedback (i.e., proper learning supports) on students' unsuccessful game performance to help them learn from their mistakes (Westera, 2022).

2.3. Current study

The studies reviewed above show that when learning supports should be delivered within educational games is an open question. We investigated this issue with 7th to 11th graders who played *Physics Playground* (*PP*; Shute et al., 2019), a two-dimensional educational game on Newtonian physics (Fig. 1). In collaboration with physics experts, we previously developed various learning supports in *PP* (see Shute et al., 2020a). Design variations included the type of learning support (e.g., expert video solutions, hints, formulas, definition minigame, and short animation physics videos; see Wouters & van Oostendorp, 2013 for more details about the types of learning supports) and media (e.g., videos or text-only). After several usability and experimental studies (Bainbridge et al., 2022; Kuba et al., 2021a; Shute et al., 2020a,b), the short animated physics videos appeared to be the most effective form of support for learning. These supports were created using multimedia learning principles (Mayer, 1997) and the first principles of instruction (Merrill, 2002). They use the look and feel of the game to demonstrate the underlying physics related to a game level such as the use of an appropriate simple machine (e.g., ramp, lever, springboard) to solve the level (see Fig. 3).

In a previous study, Bainbridge et al., (2022) reported that students who had access to the physics videos while playing *PP* performed better on far-transfer test items than students without supports, but there was no significant difference on near-transfer items. Far-transfer refers to adapting knowledge to a new context outside of the context in which the knowledge was acquired (e.g., when students learn percentage in class and can determine the price of an item after a percentage discount). Near-transfer refers to applying knowledge in similar contexts in which the knowledge was developed (e.g., when students solve problems on an exam that are similar to problems on their homework) (Boldbaatar & Şendurur, 2019). In Bainbridge et al., (2022) study, the support timing was intermixed for the treatment condition with some supports presented before playing certain levels while others were presented after attempting certain levels. Students also had access to the supports during gameplay via a "Help" button. Therefore, the optimal time to present learning supports is still unclear.

Accordingly, in the current study, we examined the effects of learning supports' delivery timing on students' in-game performance, learning, and subjective perceptions. In a between-subjects design, the learning support videos were either delivered before or after students engaged in a game level; there was also a no-support control condition. Our specific research questions were as follows: What are the effects of learning supports' delivery timing (before or after playing a game level vs. no-support control) on *(RQ1) gameplay and in-game performance, (RQ2) learning (near- and far-transfer) controlling for incoming knowledge, and (RQ3) subjective perceptions of gameplay (i.e., enjoyment, competence, frustration, and value)?*

As the literature about the learning supports' delivery timing in educational games showed mixed results, we did not have clear hypotheses about our research questions. As described earlier, providing supports before playing a game level can work as an advance organizer (Barzilai & Blau, 2014; Liao et al., 2019), reducing students' cognitive load (Liao et al., 2019; Mayer et al., 2002), and cueing students for success in the game. Specifically, the physics video supports demonstrate the use of a relevant simple machine (i.e., lever, ramp, springboard, and pendulum) to solve a level (although levels can often be solved by multiple machines) while concurrently discussing the physics behind the level. Hence, students who see the most relevant simple machine before solving a level might be cued to use the appropriate machine from the start rather than getting frustrated by trial and error (Karumbaiah et al., 2018). Thus, students in the *Before* condition might experience more successful gameplay, resulting in a more positive subjective experience and learning (previous research has demonstrated that in-game performance predicted students' scores on the posttest controlling for the pretest; see Shute et al., 2020a). Conversely, it may be argued that any trial and error, exploration, and reflection that the students in the *After* condition experience can help learning because the supports may serve as a form of formative feedback. Further, viewing the supports after engaging in a level can prepare students for success in subsequent levels, thereby supporting the generalizability of knowledge and more robust learning. Therefore, a tentative hypothesis is that providing supports after a level might result in higher learning, though not necessarily better in-game performance. Conversely, learning supports delivered before one plays a game level could improve students' in-game performance but might not yield better learning outcomes. Finally, we hypothesize that students assigned to the treatment conditions will have a more positive subjective experience about the game compared to students in the control condition, perhaps more for those in the *Before* condition since the provision of supports could lead to greater in-game success.

The literature reviewed on effective educational games and the timing of learning supports focused on the use of external supports

Fig. 1. The game level in *PP* (left) and its solution using a springboard (right).

provided before or after entire gameplay sessions. To our knowledge, no other studies have investigated the effects of providing in-game learning supports before or after playing individual game levels. Therefore, the results of this study add to the discourse on learning supports in educational games with a focus on in-game learning supports instead of learning supports that are external to gameplay (i.e., instructor-led sessions). The current study can also improve our understanding of how the timing of in-game learning supports impacts students' game performance, learning, and other outcomes (e.g., enjoyment and frustration). As a result, stakeholders (i.e., game-based learning researchers, educators, and educational game developers) can use the findings of this study when designing and developing computer-based, educational games. As the design literature on effective educational games continues to grow, higher-quality resources can be developed that follow well-established theories and that help advance educational outcomes for learners.

4. Method

4.1. Participants

Due to the constraints of the COVID-19 pandemic, all research was conducted online. We collected data from five cohorts of middle and high school students between the Spring and Fall of 2020. Data were collected from 204 participants, of which 149 are analyzed here (see Data Exclusion). The first cohort ($n = 12$) participated as a remote after-school program (students were from across the US), the second ($n = 16$), and fourth ($n = 12$) cohorts participated as a remote activity related to their class (i.e., students of this cohort were classmates), the third ($n = 69$) cohort's students were also classmates and had an option to participate remotely from home or come to their class wearing masks (everyone had to login via Zoom; researchers collected the data remotely with the help of teachers), and the fifth cohort ($n = 40$) participated as a remote summer camp program.

Students' ages ranged from 12 to 17 years, and the sample was diverse with respect to ethnicity and gender (see Table 1). Students who completed the study received a certificate of participation in addition to other performance-based incentives (see below).

5. Research design

We used a between-subjects pretest-posttest design. Students were randomly assigned into one of three conditions: (1) *Before* ($n = 50$): in which learning supports were delivered before students started a level; (2) *After* ($n = 46$): in which learning supports were delivered after students solved or quit a level; and (3) *Control* ($n = 53$): in which students did not receive any learning supports during gameplay. There were no condition effects on participant inclusion in the study (χ^2 (3) = 2.72, $p = .44$).

6. Materials

6.1. Educational game

Physics Playground (*PP*; Shute et al., 2019) is a 2-dimensional game created for 7th to 11th graders targeting Newtonian physics understanding (e.g., energy can transfer and properties of torque). The goal is to hit a red balloon with a green ball achieved by drawing simple machines (i.e., ramp, springboard, lever, and pendulum) and objects (e.g., weights); everything in the game obeys the laws of physics. Fig. 1 shows a level that can be solved using a springboard. First, students create a springboard and draw and attach a weight to the springboard. Then, they delete the weight once the springboard is bent to release the springboard and exert the needed

Table 1
Participants' demographics (n = 149).

Age (*M* (*SD*))	14.33 (0.96)
Source (%)	
Cohort 1	12 (8.1)
Cohort 2	16 (10.7)
Cohort 3	69 (46.3)
Cohort 4	12 (8.1)
Cohort 5	40 (26.8)
Gender (%)	
Female	68 (45.6)
Male	56 (37.6)
NA	25 (16.8)
Ethnicity (%)	
Asian	14 (9.4)
Black or African American	32 (21.5)
Hispanic	9 (6.0)
Mixed	18 (12.1)
Other	3 (2.0)
White	49 (32.9)
Prefer not to Answer	1 (0.7)
NA	23 (15.4)

Note. NA = missing value.

force to launch the ball high enough to hit the balloon.

The difficulty of a level is determined by the complexity of game mechanics and physics understanding needed. Together, these determine the number of objects needed for an efficient solution (referred to as "par" in the game, as in golf). Pars range from 1 (easy levels) to 5 (hard levels) and determine if a person receives a silver or gold coin for their solution (e.g., if the student solves a level on or under par, they receive a gold coin, otherwise silver). The game includes a button called *Coins* with information about the number of objects needed to obtain a gold coin (e.g., "Your solution needs to have less than 3 objects to earn a gold coin").

The game version for this study included five tutorial levels, two warmup levels (easy levels), and 28 game levels with various difficulties targeting the physics concepts of *energy can transfer* (ECT) and *properties of torque* (POT)—in total, 35 levels were all grouped within one large playground. Students played through these levels during individual gameplay sessions. Students could freely navigate through the game levels, however, based on our observations across multiple studies in the past (e.g., Kim & Shute, 2015), students usually choose to navigate the game levels linearly (as the game levels are presented to them in a predetermined order).

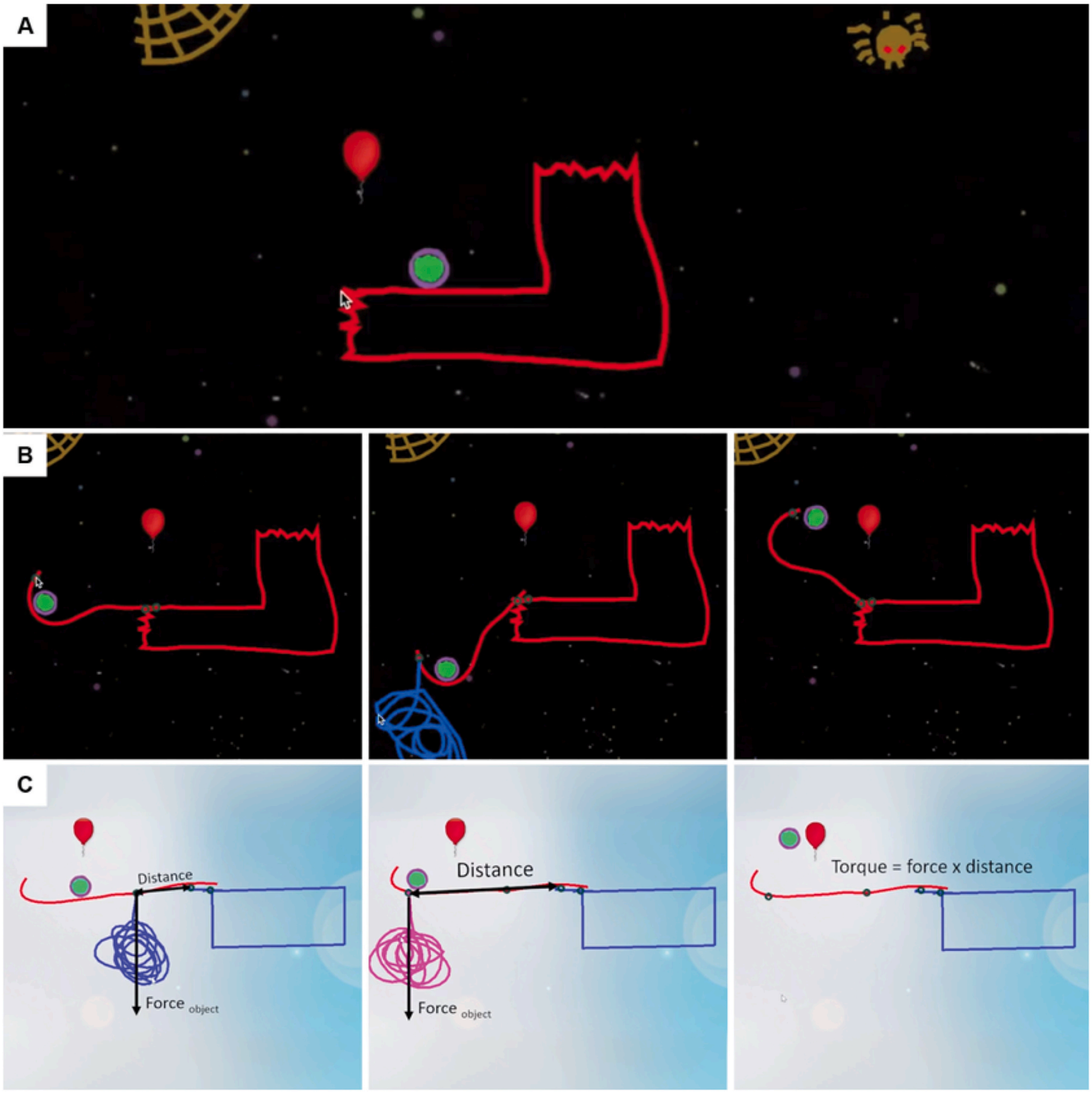

Fig. 2. (A) An example of an ECT level in *PP*; (B) the springboard solution; and (C) POT Springboard physics video.

6.2. Learning Supports—Physics videos

The physics videos present physics concepts (e.g., ECT, and POT) in the game environment relevant to a student's current game level. For example, in a level designed to target properties of torque (POT), the physics videos would present the connection between torque and relevant game mechanics or solutions. This alignment was intended to help students learn the concept behind a game level and direct them to the right solution. The physics videos were created with the following characteristics. First, we created the videos based on Merrill's first principles of instruction (e.g., learning will be promoted when new information is demonstrated to the learners, and when learners apply the new knowledge in solving real-world problems) (Merrill, 2002), and Mayer's principles of multimedia (Mayer, 1997) (see Kuba et al., 2021a). Second, the videos were created in the game environment using a level editor which is an interactive environment that allows non-technical users to create game levels by placing the ball and the balloon anywhere on the screen, drawing lines and objects, and specifying their behavior as static or dynamic. And third, all videos followed a similar structure using human voiceover where the video would first introduce the physics concept, then demonstrate a failed attempt followed by a successful solution. Note that the physics videos do not reveal the solution for a game level.

Fig. 2 shows an example of a game level targeting the physics concept of energy can transfer (ECT) using a springboard solution, and the connected physics video (i.e., ECT Springboard), and Fig. 3 depicts how the video was rendered in the game. Students could rewind a physics video as many times as they wanted while the video window was open. Importantly, students did not have access to the videos during gameplay. We intentionally limited the physics videos to either before or after levels to isolate the effect of the delivery time (i.e., *Before* or *After* conditions) compared to no delivery (i.e., *Control* condition). Appendix A shows the seven physics videos used in this study along with their links.

Students assigned to the *Before* condition received a message encouraging them to watch a learning support video (discussed below) to help them prepare for and solve the level they were about to enter—i.e., a learning support as an advance organizer. Students in the *After* condition received a message when they solved or quit a level that encouraged them to watch a video that could help them solve similar levels in the game—i.e., a learning support as a reflection (see Fig. 4 for both messages).

We made various design decisions when developing the version of *PP* that we used for this study to make sure the *Before* and *After* conditions were distinct. First, we ordered all the levels from easy to difficult ($r = .62$ between level number and difficulty) to ease students' entry into the game. Second, we ordered the levels in a way that students alternated between concepts (ECT or POT). This was done because if students in the *After* condition played two POT levels back-to-back, the support that students viewed after the first level could be considered as a before support for the second level, so this was avoided. Third, we counterbalanced the ordering of levels with associated supports (see Table 2) in that half the students in each treatment condition saw the supports in one set of 14 levels (counterbalancing A in Table 2) scattered throughout the 28 levels; the other half saw the supports in the other set of 14 levels (counterbalancing B in Table 2). For example, students in counterbalancing order A would first receive the ECT-Lever support for the Sunny Day level, while those in counterbalancing order B would receive that same support for the Spider Web level.

Fourth, we included each physics video (shown in Appendix A) in two of the relevant game levels (shown in Table 2) to potentially increase exposure to the support and facilitate generalization (i.e., the same support can be useful for multiple levels). For example, the ECT-Ramp physics video appears for both the Big Watermill and Ultimate Pinball levels in counterbalancing order A. Fifth, students were required to watch the entire support video the first time it was presented to them during gameplay. Specifically, the game was programmed to: (a) check if the student was viewing the physics video for the first time, and (b) if *a* was true then the game would wait until the physics video was completed before providing a button to allow the students close the video window. In all subsequent views of the physics video, students could close the video at will.

6.3. Other supports and features

Students always had access to *Hints* during gameplay, which point students to the correct solution per level (e.g., "Try drawing a ramp."). Students could also access *Game Tips* to review game mechanics (e.g., how to delete an object and nudge the ball) and receive quick reminders about how to create the simple machines (i.e., ramp, springboard, lever, and pendulum). These supports were accessed via buttons at the bottom of the screen. In the main menu of the game, students could select the next level to play and see their progress (i.e., the number of levels they completed, and coins collected).

6.4. Measures

6.4.1. In-game measures

PP records player actions and gameplay events into log files, such as: the number of levels solved, number of gold coins, number of silver coins, engagement with the physics videos (i.e., duration of views), and duration of gameplay. We parsed and extracted these data from log files for analysis. Specifically, we focused on the following six variables: (1) number of levels visited (including revisits to a level); (2) mean level duration in seconds (3) number of physics videos (i.e., the main learning support); viewed; (4) number of other supports (*Hints* and *Game Tips*) viewed; (5) proportion of levels in which the appropriate simple machine (i.e., pendulum, ramp, springboard, lever) was used in the participant's solution (i.e., Focal Machine Drawn Rate); and (6) success rate (i.e., Solution Rate) computed as the number of coins earned (gold + silver) divided by the number of levels visited (from #1). There were a few implausible values for level durations (presumably due to participants failing to close out the game in their browsers), so values greater than 30 min per level were clipped at 30 min (e.g., 140-min would be recoded as 30-min) prior to computing the mean level duration per participant.

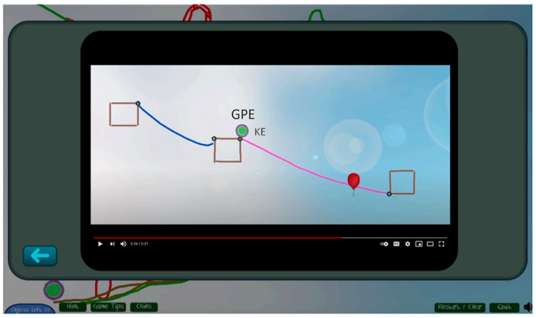

Fig. 3. Video player window showing an ECT-Ramp physics video within a level.

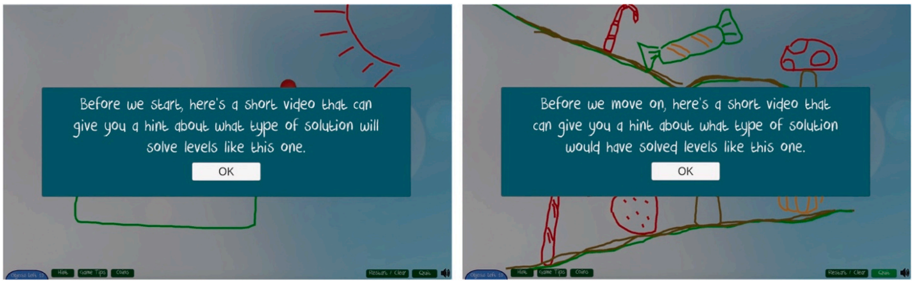

Fig. 4. The message for the physics video in the *Before* (left) and *After* conditions (right).

Table 2
The order of the focal levels and counterbalancing of showing the physics videos.

#	Level Name	Difficulty (Max = 10)	Physics Concept	Physics Video (Learning Support)	Counterbalance for Learning Support	
					A*	B*
1	Downhill (Warmup)	2	N1stL	None		
2	Lead the ball (Warmup)	2	N1stL	None		
3	Chocolate Factory	4	ECT	ECT-Ramp		x
4	Sunny Day	6	ECT	ECT-Lever	x	
5	Shark	5	POT	POT-lever-distance		x
6	On the Upswing	4	ECT	ECT-Pendulum	x	
7	Scale	4	POT	POT-lever-mass		x
8	Trunk Slide	5	ECT	ECT-Springboard		x
9	Diving Board World	7	POT	POT-Springboard	x	
10	Big Watermill	5	ECT	ECT-Ramp	x	
11	Spider Web	6	ECT	ECT-Lever		x
12	One at a Time	6	POT	POT-lever-distance	x	
13	Cloudy Day	6	ECT	ECT-Pendulum		x
14	Timing is everything	5	POT	POT-lever-mass	x	
15	Little Mermaid	4	ECT	ECT-Springboard	x	
16	Stiff Curtains	5	POT	POT-Springboard		x
17	Ultimate Pinball	5	ECT	ECT-Ramp	x	
18	Leprechaun	7	ECT	ECT-Lever		x
19	Need Fulcrum	5	POT	POT-lever-distance	x	
20	Uphill Battle	6	ECT	ECT-Pendulum		x
21	Crazy Seesaw	7	POT	POT-lever-mass	x	
22	Yippie!	6	ECT	ECT-Springboard	x	
23	Roller coaster	7	POT	POT-Springboard		x
24	Around the Tree	4	ECT	ECT-Ramp		x
25	Wavy	7	ECT	ECT-Lever	x	
26	Up in the Air	7	POT	POT-lever-distance		x
27	Maze	7	ECT	ECT-Pendulum	x	
28	Can Opener	6	POT	POT-lever-mass		x
29	Diving Board	6	ECT	ECT-Springboard		x
30	Perfect Toss	6	POT	POT-Springboard	x	

Note. *: The "x" indicates that the students assigned to counterbalance order A or B would receive the physics videos (supports) on the corresponding levels. ECT = Energy Can Transfer; POT = Properties of Torque; N1stL = Newton's First Law.

6.5. Physics understanding (learning assessment)

We created two isomorphic forms each including 28 illustrative multiple-choice items (Form A = 28 items, Cronbach's $\alpha = 0.80$; Form B = 28 items, Cronbach's $\alpha = 0.83$) covering the two physics competencies in this study (i.e., ECT and POT). To control for possible differences between the forms, we counterbalanced the pretest and the posttest. That is, half of the students in each condition received Form A for the pretest and Form B for the posttest, while the other half of the students in each condition received Form B for the pretest and Form A for the posttest.

Each form included 14 near-transfer items (designed in the context of *PP*; see Fig. 5, on the top), and 14 far-transfer items, similar to the Force Concept Inventory (Hestenes et al., 1992); see Fig. 5, on the bottom. These items were developed with the help of two physics experts and subjected to several pilot tests before administration in the current study.

We computed proportions for the knowledge tests as the number of items correctly answered divided by the number of items in the test, both overall ($n = 28$) and separately for the near-transfer ($n = 14$) and far-transfer ($n = 14$) items. An independent-sample *t*-test indicated that there were no significant differences in total scores among the two versions of the pretest ($p = .89$), but students scored slightly higher on Form B ($M = 18.1$, $SD = 4.79$) of the posttest compared to Form A ($M = 16.4$, $SD = 5.88$; $p = .049$). We, therefore, included a two-level (Form A or B) categorical variable (called *Posttest Form*) for what form the students received on the posttest as a covariate in the analyses.

6.5.1. Intrinsic Motivation Inventory (IMI)

To gauge enjoyment of the game and other measures of subjective experience, we modified and used the Intrinsic Motivation Inventory (IMI; Ryan, 1982) for this study. Our scale had 18 items presented as statements on a 7-point Likert scale ranging from "Not at all true" to "Very true" (see Appendix A). There were six subscales with two to four items per subscale, including *enjoyment* (e.g., "The game was fun to play"; $n = 5$; $\alpha = 0.91$), *perceived competence* (e.g., "I think I am pretty good at the game"; $n = 4$; $\alpha = 0.79$), *effort* (e.g., "I put a lot of effort into the game"; $n = 4$; $\alpha = 0.85$), *frustration* (e.g., "I felt very frustrated while playing the game"; $n = 3$; $\alpha = 0.75$), and *value* (e.g., "I believe playing the game could be beneficial to me"; $n = 2$; $\alpha = 0.75$). The IMI is known to be reliable and valid (Tsigilis & Theodosiou, 2003), with an overall Cronbach's α of 0.90 in the present study.

6.6. Procedure

The COVID-19 situation resulted in many schools temporarily pivoting to remote learning, which made recruiting students difficult. Thus, we developed a remote (via Zoom) after-school, summer camp, or in-class[1] program involving contests, certificates, individual and group play, and other fun team-building activities (e.g., a mini-game where students got to guess what happens next in a fun video). This way, parents, teachers, and students were more willing to register, participate, and stay for the entire program which would allow us to collect more data.

Once they were signed up to participate, we asked students and a parent or guardian to electronically sign the student assent and parental consent forms. The ECT and POT focal levels with three experimental conditions and learning supports were incorporated into the individual gameplay part of the program. The group gameplay involved levels *without* any supports targeting *levels with non-focal* physics concepts (i.e., not ECT or POT) as a part of the program. Thus, care was taken to minimize the influence of group gameplay on individual gameplay (the focus of the study), and students in all conditions (*Before*, *After*, and *Control*) participated in both individual and group gameplay.

For the students who participated in the study as a summer camp or an in-class activity, the study was completed across three days (see Fig. 6) including both individual and group gameplay; the other cohorts completed the study over 5 day as detailed below. On Day 1, after a short icebreaker activity (e.g., each student answered the question, "What is your favorite game?"), students completed a questionnaire including demographic questions and the physics understanding pretest (25 min). Afterward, students engaged in 50 min of individual gameplay (with five tutorial levels, and two warmup levels followed by 28 focal ECT and POT levels; see Table 2) followed by 45 min of group gameplay in groups of two to four students. On Day 2, after another short icebreaker activity (e.g., each student answered the question, "Who is your favorite scientist?"), several awards were distributed among students who performed well on Day 1's individual gameplay (i.e., *Level Domination* for the student who played most of the levels, *Gold Coin* for the student who collected the most gold coins, *Supersonic* for the student who solved the levels the quickest on average, and *Most Tenacious* for the student who persisted the most when solving game levels). Then, students continued playing the focal levels individually (shown in Table 2) for about 1 h. Students were then divided into the same groups as on Day 1 and were tasked to create their own game levels using *PP*'s level editor for about 1 h. Again, this part of the program was intended to provide fun activities for students and help them feel like they were participating in an after-school or summer camp program, and not a research study.

On Day 3, after a short introduction, students continued playing individually (the focal levels shown in Table 2) for about 45 min. Then they completed the posttest followed by the IMI questionnaire for about 25 min. Afterward, students were divided into their groups and played the game levels they created on the previous day for about 45 min. Finally, other awards were distributed to high-performing students with the addition of two new awards: the *Isaac Newton* award given to the student(s) who had the highest gain score on the physics test (receiving a $10 e-Gift card), and the *Creative Designer* award to each student in the group that designed the

[1] Students in Cohort 3 had the option to come to school or participate remotely. However, in both cases, while students were physically in class, they had to login via Zoom to participate.

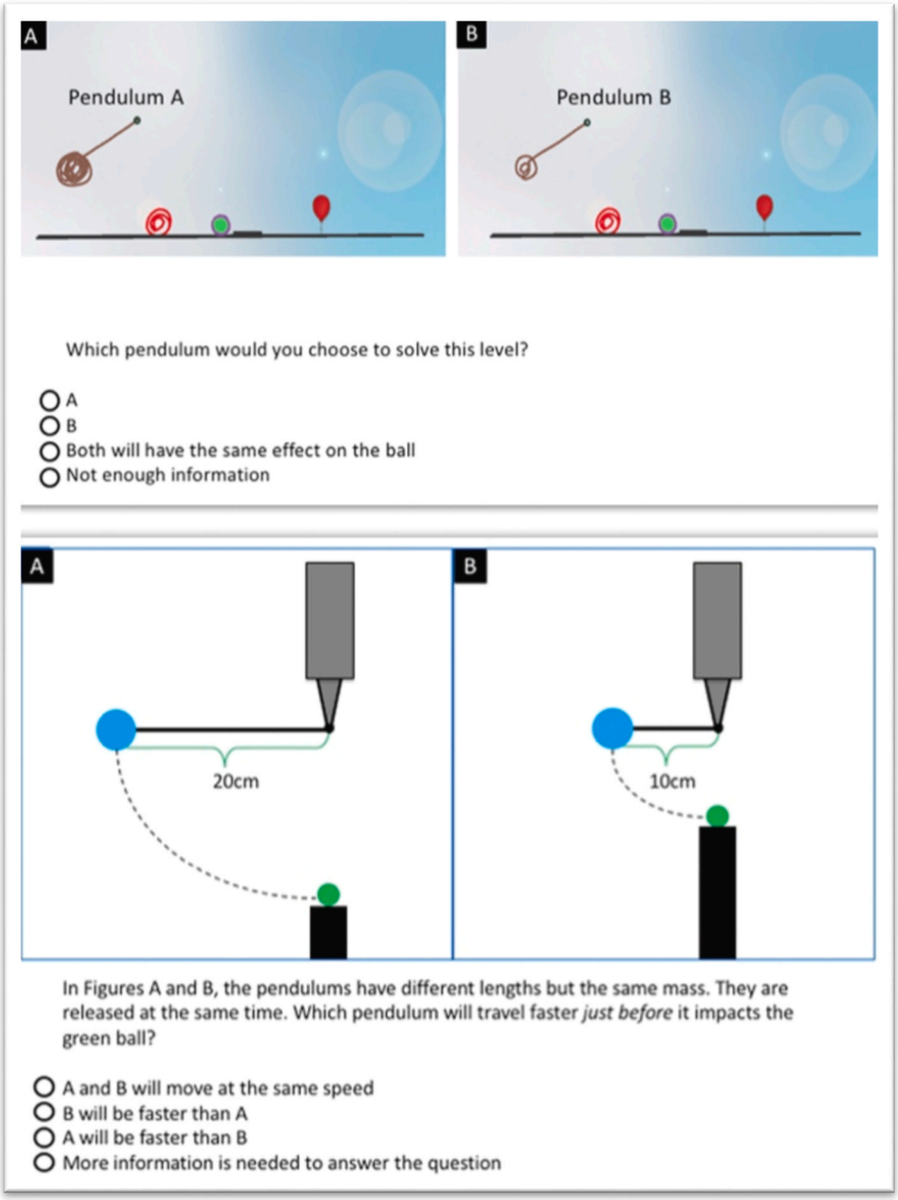

Fig. 5. Examples of near- (top) and far-transfer (bottom) test items for physics understanding.

most creative game level (selected by the students through an anonymous poll in Zoom); the winners also received $10 e-Gift cards.

The procedure for the students attending the after-school programs was very similar to the above summer camp and in-class students. The major difference was that we conducted the afterschool program across five days with shorter sessions per day (about 1 h per day). However, we designed the five sessions to align with the gameplay time and the order of the events from the three-session study. That is, collectively, students played individually for about 150 min (i.e., the tutorial, warmup, and focal levels), and also, they played and designed levels in groups for about 150 min. They also spent the same amount of time completing the pretest and posttest questionnaires (25 min per test). The main reason for the change in sessions was that we did not want to have students in front of computers for an additional 3 h after possibly being in front of a computer during the school day due to digital learning models.

While the research literature on educational games shows effective computer-based, educational games can positively influence learning, concern exists related to possible negative effects of the amount of time children spend playing video games. For example, Melo et al. (2020) noted the negative consequences (e.g., addiction) of extended screen time when playing video games. To avoid such negative consequences, we intentionally limited the amount of time spent in front of the screen during our study (e.g., spreading out game-based learning activities across several days) to prevent possible adverse effects. We also interspersed other activities to reduce any ill effects of just playing the game (e.g., ice-breaker activities). Finally, well-designed educational games that include

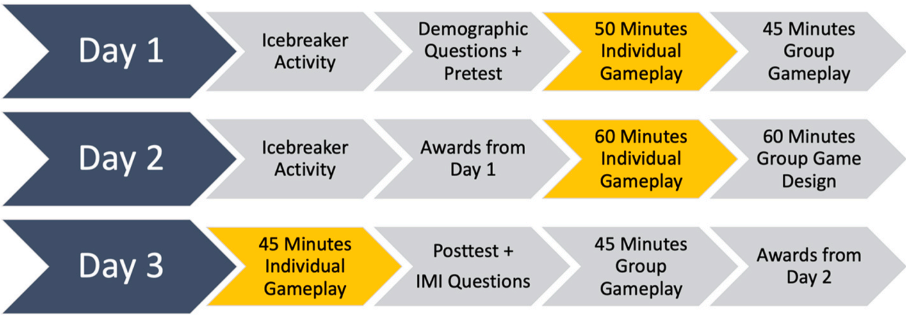

Fig. 6. The procedure during the three days. Note: IMI = Intrinsic Motivation Inventory; the highlighted steps indicate individual gameplay including the focal levels in Table 2.

well-designed learning supports directing students to the knowledge they need to gain, should have fewer negative consequences, if any, compared to commercial violent video games that could have negative effects (Ritterfeld et al., 2005). *Physics Playground* and similar educational games promote a fun approach to learning (e.g., Rahimi, 2020; Shute et al., 2020a; Yang et al., 2021), controlled by adult supervision (e.g., teachers), that can help improve students' self-efficacy and prepare them for formal learning of difficult concepts (Gee, 2003).

6.7. Data Exclusion

Of the sample of 204 participants, 1 was discarded due to a data entry error where the experimental condition could not be determined, 4 were missing a pretest, 40 a posttest, and 10 were missing both. There were multiple reasons for the large number of missing posttests including running out of time, noncompliance, and lack of motivation due to the pandemic. We elected to proceed with the remaining 149 participants rather than impute the missing data due to the large volume of missing data (about 27%) and likely non-random missingness. Importantly, there was not a statistically significant association between pre- or post-missing data and assignment to conditions (χ^2 *ps* > .45), gender (*ps* > .18), ethnicity (*ps* > .45), and age (*rs* < 0.06, *ps* > .48). In addition, three participants were missing gameplay data or survey data, so the number of observations varies slightly for analyses involving these measures.

7. Results

We used R (R Core Team, 2021) to run our analyses. Specifically, we used linear mixed-effects regression models for the analyses with cohort as a random intercept, using the lme4 package (Bates et al., 2015) with Satterthwaite's degrees of freedom method from lmerTest (Kuznetsova et al., 2017). Condition (*Before*, *After*, *Control*) was included as a three-level categorical fixed effect with the

Table 3
Descriptive Statistics of the Key Variables Analyzed (n = 149).

	After (*n* = 46)	Before (*n* = 50)	Control (*n* = 53)
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Gameplay			
Level Duration (seconds) [92 to 939]	355.20 (157.47)	377.49 (164.74)	257.84 (96.27)
No. Levels Visited [0 to 77]	24.27 (13.95)	23.33 (13.16)	28.81 (17.03)
No. Physics Videos Viewed [0 to 31]	11.33 (7.57)	11.83 (6.68)	0.00 (0.00)
No. Other Supports Viewed [0 to 86]	11.62 (14.37)	9.25 (8.21)	13.92 (15.51)
Focal Machine Drawn Rate [0 to 1]	0.51 (0.14)	0.52 (0.14)	0.50 (0.14)
Solution Rate [0, 1]	0.29 (0.11)	0.47 (0.21)	0.49 (0.18)
Pretest and Posttest [0 to 14]			
Pretest Near	9.15 (2.32)	7.78 (3.02)	7.91 (3.01)
Pretest Far	7.48 (2.40)	7.18 (2.57)	6.51 (2.49)
Posttest Near	10.17 (2.51)	9.74 (3.47)	9.11 (2.89)
Posttest Far	7.96 (2.54)	7.74 (3.15)	7.11 (2.97)
IMI Sub-scales [1 to 7]			
Enjoyment	5.53 (1.08)	5.56 (1.16)	5.00 (1.50)
Competence	4.48 (1.27)	4.38 (1.29)	4.46 (1.24)
Effort	5.37 (1.11)	5.47 (1.30)	5.46 (1.03)
Frustration	4.21 (1.66)	4.08 (1.35)	4.80 (1.58)
Value	4.78 (1.54)	4.91 (1.56)	4.38 (1.69)

Note. The observed range of variables is included in square brackets.

Control condition as the reference group. Pairwise comparisons were conducted using the emmeans package (Lenth et al., 2019), also with Satterthwaite's degrees of freedom method. We used two-tailed tests with a $p < .05$ cutoff for significance. We report the standardized beta coefficients in the main text and unstandardized coefficients in regression tables. Tables 3 and 4 provide descriptive statistics and partial Pearson correlations for the key variables analyzed.

8. Preliminaries

We first tested for adequacy of random assignment by regressing total pretest score on condition, and found no significant differences (*ps* > .16). Similarly, condition assignment was independent of gender, ethnicity, and counterbalancing of the pretest and posttest (*ps* > .22*)*. Next, we checked whether the total time spent on gameplay was equivalent across conditions, and found this to be the case (*ps* > .63). Turning to the experimental conditions, on average, students viewed the physics videos about 12 times (median = 10 with a range of 1–31) with no significant differences ($p = .87$) between the *Before* and *After* conditions. However, only 47% of students viewed all the seven supports at least once, again with no significant differences among conditions ($p = .98$). Thus, random assignment was successful and exposure to the learning supports was equivalent across the two experimental conditions.

9. Gameplay data

For RQ1 (*What are the effects of learning* supports*' delivery timing [before or after playing a game level vs. no-*support *control] on students' gameplay and in-game performance?)* we regressed our key gameplay measures (Table 5) on Condition after controlling for the total pretest score. The results indicated that compared to the students in the *Control* condition, those in the *Before* ($\beta = -0.51$, $p = .005$) and *After* ($\beta = -0.50$, $p = .007$) conditions visited fewer levels (about 7 levels) overall, but the students in the *Before* ($\beta = 0.73$, $p < .001$) and *After* ($\beta = 0.52$, $p = .006$) conditions stayed longer on each level compared to the students in the *Control* condition. We also found that participants in all three conditions were equally likely to use the appropriate simple machine (e.g., lever, pendulum) highlighted in the physics videos than those assigned to the *Control* condition (*ps* > .25), although those in the *Control* condition used about five more of the other supports (i.e., *Hints* and *Game Tips*) than those in the *Before* ($p = .002$) condition and eight more of the other supports than those in the *After* ($p = .06$) condition. There were no significant differences (*ps* > .24) between the *Before* and *After* conditions for these gameplay variables. However, students in the *After* condition had a significantly lower solution rate (i.e., earned about 25% fewer coins per level attempt) than the students in the *Before* ($\beta = -1.12$, $p < .001$) and *Control* ($\beta = -1.28$, $p < .001$) conditions, which were comparable ($p = .31$).

9.1. Learning measures

To address RQ2 (*What are the effects of learning* supports*' delivery timing [before or after playing a game level vs. no-*support *control] on students' learning [near and far transfer] controlling for incoming knowledge?)*, we investigated Condition effects on near- and far-transfer posttest scores after controlling for the corresponding near- and far-pretest scores and posttest form (A or B; see Table 6). There were no significant differences among conditions for either near- or far-transfer posttest scores without any additional covariates (*ps* > .42; Models 1 and 2 in Table 6).

Because the previous analyses (Table 5) indicated that the learning supports resulted in different patterns of gameplay and outcomes, we examined the effects of Condition on students' near- and far-transfer posttest scores after controlling for the five gameplay variables (i.e., mean level duration, total number of levels attempted, use of other supports [i.e., *Hints* and *Game Tips*], use of the target simple machine, and level success [i.e., Solution Rate]). Indeed, after adjusting for these gameplay measures, students in the *After*

Table 4
Partial pearson correlations among the key variables analyzed controlling for pretest Total.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Level Duration												
2. No. Levels Visited	−.16											
3. No. Physics Videos Viewed [a]	−.08	**.96**										
4. No. Other Supports Viewed	.03	**.48**	**.34**									
5. Focal Machine Drawn Rate	.11	−.14	−.11	.00								
6. Solution Rate	−.11	.09	**.27**	−.04	**.42**							
7. Posttest (Near)	.04	**.24**	**.22**	**.18**	.01	**.16**						
8. Posttest (Far)	.03	**.20**	**.22**	**.18**	.14	**.26**	**.51**					
9. Enjoyment	**.16**	.05	**.16**	.09	.00	.02	**.23**	.14				
10. Competence	−.02	**.20**	**.31**	−.03	.02	**.27**	.10	.14	**.44**			
11. Effort	.04	.12	.15	**.16**	−.04	.02	**.18**	.09	**.54**	**.28**		
12. Frustration	−.03	−.01	−.1	.06	.05	.00	−.06	**−.19**	**−.32**	**−.33**	.00	
13. Value	**.19**	.05	.11	.12	.01	−.02	**.17**	.15	**.76**	**.40**	**.52**	**−.33**

Note. Computed correlations used Pearson-method with pairwise-deletion. Bolded correlations are above 0.15 a threshold for small to large correlations.

[a] Only the treatment conditions' data was used to compute the correlations for Physics Videos Viewed as it was not possible to view the Physics Videos in the *Control* condition; for other variables, we used the data from all the conditions.

Table 5
Regression models with gameplay variables as dependent variables and condition and pretest as predictors.

Predictors	Model 1		Model 2		Model 3		Model 4		Model 5	
	Level Duration		Levels Visited		Other Supports Viewed		Focal Machine Drawn Rate		Solution Rate	
	B	p	B	p	B	p	B	p	B	p
(Intercept)	308.77	**<0.001**	13.80	**0.006**	14.52	**0.003**	0.51	**<0.001**	0.32	**<0.001**
Condition [After]	78.22	**0.006**	−7.48	**0.007**	−4.91	0.054	−0.02	0.407	−0.25	**<0.001**
Condition [Before]	108.97	**<0.001**	−7.71	**0.005**	−7.98	**0.001**	0.01	0.738	−0.03	0.309
Pretest	−1.77	0.494	1.06	**<0.001**	0.20	0.397	0.00	0.634	0.01	**<0.001**
Random Effects										
σ^2	18,269.27		172.21		145.69		0.02		0.02	
τ_{00}	3291.54 $_{\text{Cohort}}$		25.95 $_{\text{Cohort}}$		29.97 $_{\text{Cohort}}$		0.005 $_{\text{Cohort}}$		0.005 $_{\text{Cohort}}$	
ICC	0.15		0.13		0.17		0.22		0.18	
N	5 $_{\text{Cohort}}$		5 $_{\text{Cohort}}$		5 $_{\text{Cohort}}$		5 $_{\text{Cohort}}$		5 $_{\text{Cohort}}$	
Observations	146		146		146		146		146	
Marginal R^2/Conditional R^2	0.093/0.231		0.143/0.255		0.062/0.222		0.008/0.229		0.318/0.439	

Note. Significant ($p < .05$) effects are bolded. The reference group is the *Control* condition. B = unstandardized coefficient.

Table 6
Regression models predicting learning outcomes—near and far transfer.

Predictors	Model 1		Model 2		Model 3		Model 4	
	Posttest (Near)		Posttest (Far)		Posttest (Near)		Posttest (Far)	
	B	p	B	p	B	p	B	p
(Intercept)	4.41	**<0.001**	4.58	**<0.001**	2.23	0.062	1.96	**0.040**
Condition [After]	0.11	0.822	0.04	0.933	1.49	**0.011**	1.48	**0.006**
Condition [Before]	0.39	0.419	0.37	0.424	0.96	0.054	0.76	0.105
Pretest (Near)	0.61	**<0.001**			0.46	**<0.001**		
Posttest Form [B]	0.60	0.123	−2.56	**<0.001**	0.65	0.086	−2.06	**<0.001**
Pretest (Far)			0.60	**<0.001**			0.45	**<0.001**
Solution Rate					4.38	**0.002**	4.81	**<0.001**
Levels Visited					0.04	**0.020**	0.03	0.080
Level Duration					0.00	0.416	0.00	0.919
Other Supports Viewed					0.02	0.185	0.02	0.134
Focal Machine Drawn					−1.28	0.434	−0.32	0.838
Random Effects								
σ^2	5.50		5.07		4.90		4.59	
τ_{00}	0.47 $_{\text{Source}}$		0.38 $_{\text{Source}}$		0.39 $_{\text{Source}}$		0.00 $_{\text{Source}}$	
ICC	0.08		0.07		0.07			
N	5 $_{\text{Source}}$		5 $_{\text{Source}}$		5 $_{\text{Source}}$		5 $_{\text{Source}}$	
Observations	149		149		146		146	
Marginal R^2/Conditional R^2	0.337/0.389		0.359/0.403		0.444/0.485		0.483/NA	

Note. Significant ($p < .05$) effects are bolded. The reference group is the *Control* condition. B = unstandardized coefficient. Model 4 yielded a singular fit due to zero variance in the random intercept for cohort.

condition showed significantly higher near-transfer ($\beta = 0.50$, $p = .011$) and far-transfer ($\beta = 0.50$, $p = .006$) scores compared to the *Control* condition (Models 3 and 4 in Table 6). We observed a similar pattern for the *Before* vs. *Control* comparison ($\beta = 0.32$; $p = .05$ and $\beta = 0.26$; $p = .11$), but the differences were not significant. There were also no differences ($ps > .16$) between the two experimental conditions. Of the five gameplay variables, the solution rate ($\beta's$ of 0.28 and 0.32; $ps < .01$) was the most robust predictor of students' posttest scores (near and far, respectively) followed by the number of levels visited ($\beta's$ of 0.19 and 0.13 [$ps < .08$]), whereas the other three variables were nonsignificant predictors ($ps > 0.13$). Rerunning the models without these three nonsignificant predictors (see Appendix C) yielded similar results for the *After* vs. *Control* condition (βs of 0.46 and 0.48; $ps < .02$). Additionally, the *Before* vs. *Control* condition difference was similar for far-transfer ($\beta = 0.24$; $p = .11$) but became significant (β of 0.32, $p = .04$) for near-transfer.

We also examined the effect of viewing the physics videos on learning outcomes. As discussed earlier, we found the physics videos to be significantly correlated with learning outcomes in our previous studies (Bainbridge et al., 2022; Shute et al., 2020a). The partial correlation analysis showed a significant correlation between the number of times students viewed physics videos and the near- ($r = 0.24$, $p = .02$) and far-transfer posttest scores ($r = 0.22$, $p < .01$) controlling for students' pretest scores.

10. Subjective perceptions

Finally, to address RQ3 (*What are the effects of learning* supports' *delivery timing [before or after playing a game level vs. no*-support *control] on students' subjective perceptions of gameplay [i.e., enjoyment, competence, frustration, and value]?)*, we regressed each of the five IMI measures on Condition after controlling for total pretest scores (see Table 7). Starting with comparisons between students in the experimental and *Control* conditions, students in the *Before* condition self-reported lower frustration ($\beta = -0.47$; $p = .02$) and more enjoyment ($\beta = 0.41$; $p = .03$) than those in the *Control* condition. A similar effect was found for students in the *After* condition, with lower frustration ($\beta = -0.42$; $p = .04$) but not for enjoyment (though in the same direction; $\beta = 0.33$; $p = .10$) as compared to those in the *Control* condition. There were no significant differences for the experimental vs. *Control* contrasts for self-reports of competence, effort, and value (*ps* > .19) and none of the five *After* vs. *Before* effects were significant (*ps* > .58).

To separate the effects of the learning supports from other aspects of gameplay, we reran the above five models after controlling for gameplay success (solution rate) and exposure (number of levels attempted) as these were the same variables that predicted posttest scores (see Table 6. In addition to retaining the above significant effects pertaining to enjoyment and frustration, there was a significant *After* vs. *Control* effect for self-reported competence ($\beta = 0.56$; $p = .02$) after adjusting for gameplay (see Appendix D).

11. General discussion

We examined the effects of in-game learning supports' delivery timing on students' game performance, learning, and subjective perceptions in *Physics Playground*. In a between-subjects design ($n = 149$), the learning-support videos were delivered before or after students engaged in playing a game level; there was also a no-support control condition. In what follows, we review our main findings alongside our research questions followed by a discussion of limitations and future work.

12. Main findings

Our main findings are that: (1) including learning supports in educational games may lead to some tradeoffs compared to providing no supports (i.e., those playing with supports may have less exposure to game levels, but more time playing each game level along with less frustration); (2) providing learning supports after playing a game level resulted in lower in-game success compared to the *Before* and *Control* conditions; (3) compared to no support, students learned more and reported greater competence when the learning supports were provided after rather than before engaging with a game level, but only when controlling for in-game behaviors and performance; and (4) the significant differences we observed pertained to contrasts involving each support condition and the no-support control; when we compared the two treatment conditions there were no major differences between them. In other words, the advantage of the *After* condition was relative to the difference of this condition and the *Control* condition, rather than a direct comparison between *After* and *Before*.

Starting with the tradeoffs, including learning supports in the treatment conditions resulted in students spending more time per level and viewing the supports whereas students in the *Control* condition used that time to visit more game levels. Thus, students in the treatment conditions had less exposure to a diverse set of game levels, but were more focused on each level. This tradeoff might be resolved when the overall gameplay time is not fixed and students could finish the game at their own pace leading to an equal game exposure. However, this tradeoff might not be due just to the time limitation in the treatment conditions. Alternatively, students in the *Control* condition might just have focused on having fun, playing, and solving levels with any solution possible rather than focusing on the underlying physics (i.e., the right solution). To this point, an unsupported hypothesis was that the students in the *Before* condition would be more likely to use an appropriate simple machine (e.g., lever, springboard) in their solutions than the students in the other conditions. One reason for this finding could be that students in the *After* and *Control* conditions could correctly identify the appropriate simple machine of the game levels' solution from the appearance of the game levels. This would also explain why students in the *Before* condition did not have better solution rates (i.e., the number of coins earned divided by the number of levels visited) than those in the *Control* condition.

Table 7
Regression models predicting subjective perception outcome variables.

Predictors	Model 1		Model 2		Model 3		Model 4		Model 5	
	Enjoy		Competence		Effort		Frustration		Value	
	B	p	B	p	B	p	B	p	B	p
(Intercept)	4.12	**<0.001**	4.07	**<0.001**	5.11	**<0.001**	4.39	**<0.001**	3.79	**<0.001**
Condition [After]	0.42	0.100	−0.03	0.909	−0.20	0.394	−0.65	**0.040**	0.23	0.479
Condition [Before]	0.53	**0.034**	−0.09	0.722	−0.11	0.637	−0.74	**0.016**	0.41	0.194
Pretest	0.06	**0.009**	0.03	0.243	0.03	0.129	0.03	0.304	0.05	0.086
Random Effects										
σ^2	1.51		1.60		1.25		2.34		2.45	
τ_{00}	0.05 Source		0.00 Source		0.08 Source		0.00 Source		0.10 Source	
ICC	0.03				0.06				0.04	
N	5 Source		5 Source		5 Source		5 Source		5 Source	
Observations	146		145		146		146		147	
Marginal R^2/Conditional R^2	0.087/0.119		0.011/NA		0.019/0.079		0.049/NA		0.034/0.071	

Note. Significant ($p < .05$) effects are bolded. The reference group is the *Control* condition. B = unstandardized coefficient. Models 2 and 4 yielded singular fits due to zero variance in the random intercept for cohort.

Moreover, the finding that students' solution rate in the *After* condition was significantly *lower* compared to the students in the *Before* and *Control* conditions provides some evidence for a slight advantage of the *Before* condition in improving students' in-game performance. The lower solution rate for the *After* condition could be related to (a) lower exposure to game levels, and (b) delayed delivery of supports (which could affect their success rate negatively).

Turning to learning outcomes, we found a significant correlation between posttest scores (near- and far-transfer) and the number of physics videos viewed in the treatment conditions controlling for students' pretest scores. This finding aligns with our previous research studies (Bainbridge et al., 2022; Shute et al., 2020a) and the broader literature on the positive effects of well-designed learning supports in educational games on students' learning (e.g., Cai et al., 2022; Vrugte et al., 2015; Wouters & van Oostendorp, 2013). Further analyses, focusing on the effects of the supports' delivery timing on learning outcomes revealed an advantage for the *After* compared to the *Control* condition, but only when controlling for gameplay variables, particularly solution rate. Thus, providing supports after attempting game levels appears to reduce success in the short term (i.e., in-game performance), but can lead to better learning outcomes. This finding aligns with Westera (2022) who has reinforced the distinction between students' learning and in-game performance when playing educational games.

This advantage of providing supports after playing game levels for students' learning is supported by the literature on formative feedback (Shute, 2008). Students in the *After* condition may have benefited from the corrective function of the learning supports as valuable feedback which led them to learn better compared to the students in the *Control* condition. Students in the *Before* condition may not have paused after gameplay and figured out how to correct their mistakes—both related to game mechanics (e.g., choosing the wrong simple machine) and physics concepts (e.g., how the length of one of the arms of torque can affect how torque functions). Therefore, students' mistakes in the *Before* condition might have been repeated during gameplay and negatively affected their learning. Moreover, our findings align with what Cloude et al. (2021) found—i.e., a positive effect of students' reflections on the topics they learned after solving game levels in a science game.

Another reason that the *After* supports were more effective than the *Before* supports on students' learning when compared to the *Control* could be explained by the productive failure literature (Kapur, 2016; Kapur & Bielaczyc, 2012; Schwartz et al., 2011). Kapur and colleagues (2012) found that providing supports after students' attempted solutions did not facilitate their problem-solving performance during learning (similar to what we found regarding students' game performance in the *After* condition). However, students' conceptual understanding and knowledge transfer on the posttest were significantly higher compared to the students who received learning supports before solving problems (similar to what we found for the effects of the *After* condition after controlling for in-game measures). Moreover, the productive failure literature indicates that using delayed supports allows students to explore, discover, struggle, and be ready to learn what they are supposed to learn after a learning activity (Kapur, 2012) which is what, in our view, the *After*-condition students experienced in our study.

Turning to the subjective outcomes, providing supports resulted in lower frustration compared to providing no supports. There are several reasons why physics videos were effective in reducing students' frustration levels. First, the physics videos used the same look and feel of the game and illustrated the game mechanics while communicating the underlying physics concepts. Moreover, they included examples of a failed attempt followed by a successful attempt. Therefore, the supports could cue students into how to approach solving the level (for both *Before* supports and when solving a similar level after receiving the first *After* support) as well as additional guidance about the game mechanics than those in the *Control* condition, who may have had to resort to more trial and error and thereby experienced more frustration.

Apart from the effectiveness of providing learning supports vs. no supports in terms of lowering frustration, providing supports before or after gameplay did not show any difference in students' frustration level. This finding contrasts with the productive failure literature, which would expect that students receiving supports after a learning activity might experience more difficulty or frustration (Kapur, 2012). One reason could be that students in the *After* condition might have had mixed successes and failures—not just failures; whereas in productive failure, instructional activities are designed for students to experience failure before receiving learning supports. Further investigations are needed to identify the effects of delayed learning supports in educational games when the likelihood of failing is high.

13. Implications

Our findings shed light on several design considerations for including supports in educational games which led us to provide the following suggestions. First, include well-designed learning supports that are developed using learning theories and are well-integrated well into the game environment (i.e., have the look and feel of the game). Such learning supports can maximize students' game performance and learning, compared to external learning supports that take learners away from the fun of the gameplay. Second, depending on the aim of the supports (i.e., game or content-related supports; Yang et al., 2021), one can decide to show them before or after playing game levels. For instance, game-related supports (i.e., related to game mechanics) might be more impactful when shown before rather than after engaging in playing a level while content-related supports might be more helpful when shown after rather than before playing a game level. Note that our findings did not show any differences between supports delivered before or after playing game levels. Also, note that our supports (the physics videos) combined game- and content-related supports. Thus, an alternative approach, and our third suggestion, would be to develop two separate learning supports: game and content-related, then show game-related supports before engaging in gameplay and content-related supports after engaging in gameplay. However, this suggestion needs to be investigated in future research.

14. Limitations and future work

Our study was conducted during the COVID-19 pandemic, with some students participating toward the beginning and others in the later stages of the pandemic. Additionally, most of the data collection was done remotely and across five different cohorts, all factors which may have increased variability in the data. Moreover, we ended up not using the data from 53 students because their data was partially completed (e.g., pretest or posttest or gameplay data were missing). As a result, the study might have been underpowered to detect some of the effects, especially for the *Before* vs. *Control* contrast for learning measures (Models 3 and 4 in Table 6). Thus, replicating with a larger sample and in post-pandemic educational settings is warranted. We also only used *Physics Playground*; and other educational games with supports before or after gameplay might yield different results than what we found. Thus, further research is needed to evaluate the generalizability of our findings and to see if what we found holds for other demographics, educational games, and subject matter.

In the future, the effects of providing learning supports *during* gameplay (when students need them) should also be investigated. Additionally, future studies could investigate the effects of students' control on accessing learning supports vs. forced learning supports. One of our studies (Rahimi et al., 2021) investigated the effects of an incentive system in bringing students' attention to learning supports by incentivizing the first-time use of the supports. We found that students used the learning supports even when viewing the supports was not incentivized. In the current study, we provided forced supports to the students without their control and with no incentives. A future study could compare these two support delivery systems on students' game performance, learning, and other self-report measures (e.g., IMI data).

Another future direction for research in this area could be to investigate the relationship between the timing of the learning supports relative to students' prior knowledge. It is likely that the delivery timing of learning supports will have different effects for players with different levels of prior knowledge. A pervasive phenomenon in multimedia instructional design, dubbed "the individual differences principle" by Mayer (2020) and the "expertise reversal effect" by Kalyuga (2007), is that learning supports that aid students with lower prior knowledge often convey no benefit to students with higher prior knowledge. Students with low prior knowledge might benefit more from learning supports delivered *before* a learning activity because many of these supports aim to reduce cognitive load through pretraining which exposes students to key concepts beforehand (Mayer, 2020). Consequently, students with low prior knowledge could spend more mental energy on building accurate mental models during the learning activity (Mayer, 2020). On the other hand, students with high prior knowledge already have this information and could devote less cognitive load to essential processing during gameplay. For these learners, supports run the risk of reducing enjoyment while conveying no benefit (Mayer et al., 2002). We did not investigate moderation by individual differences in prior knowledge here due to a lack of statistical power, which is a viable future direction.

In conclusion, the timing of learning supports' intended to bolster student learning hinges on the *design* of the supports—using established theories of learning and motivation. Such well-designed supports should also be tightly integrated into the game environment (i.e., using the look and feel of the game), and combine game mechanics and content knowledge. Following these guidelines with suitable timing can enhance educational games' impact on students' game performance and learning.

Credit author statement

Seyedahmad Rahimi: Writing – original draft, Formal analysis, Conceptualization, Methodology, Data collection, Data curation, Software, Visualization, Project administration, Writing – review & editing. **Valerie J. Shute**: Funding acquisition, Conceptualization, Methodology, Data collection, Formal analysis, Project administration, Writing – review & editing. **Curt Fulwider**: Data collection, Data curation, Writing – review & editing. **Katie Bainbridge**: Conceptualization, Methodology, Data collection, Writing – review & editing, Project administration. **Renata Kuba**: Data collection, Writing – review & editing, Project administration. **Xiaotong Yang**: Project administration, Data collection, Writing – review & editing, Project administration. **Ginny Smith**: Project administration, Data collection, Writing – review & editing. **Ryan S. Baker**: Funding acquisition, Conceptualization, Writing – review & editing. **Sidney K. D'Mello**: Funding acquisition, Conceptualization, Methodology, Formal analysis, Data curation, Writing – review & editing.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this manuscript.

Acknowledgments

This work was supported by the US National Science Foundation [award number #037988] and the US Department of Education [award number R305A170376].

Appendices.

Appendix A

Physics Videos Included in this Study.

Focal Concept	Solutions	Physics Videos	Links	Duration (sec.)
ECT	Ramp	ECT Ramp	https://youtu.be/QKGNMfQPvN4	37
	Pendulum	ECT Pendulum	https://youtu.be/YT9d-o4VV-k	52
	Springboard	ECT Springboard	https://youtu.be/zNjPjo-2k88	34
	Lever	ECT Lever	https://youtu.be/TIyqwOfhxgE	41
POT	Springboard	POT Springboard	https://youtu.be/XPgDJg7lGN0	27
	Lever	POT Distance Lever	https://youtu.be/S9mi1GCtUZg	55
		POT Mass Lever	https://youtu.be/Jswujpf5vTc	55

Note: ECT = Energy can Transfer, POT = Properties of Torque.

Appendix B

Internal Motivation Inventory (IMI).

Enjoyment
I enjoyed playing the game very much
This activity was fun to do.
I thought this was a boring activity. [R]
The game did not hold my attention at all. [R]
I would describe the game as very interesting.
Perceived Competence
I think I am pretty good at this activity.
I think I did pretty well at this activity, compared to other students.
I am satisfied with my performance in the game.
This was an activity that I couldn't do very well. [R]
Effort
I put a lot of effort into this.
I didn't try very hard to do well at this activity. [R]
I tried very hard on this activity.
I didn't put much energy into this. [R]
Frustration
I did not feel frustrated at all while playing the game. [R]
I felt very frustrated while playing the game.
I was very relaxed while playing the game. [R]
Value
I would be willing to play the game again because it has some value to me.
I believe playing the game could be beneficial to me.

Note: The items were 1–7 Likert scale: From 1 = "Not at all true" to 7 = "Very true". R = reversed coded for analysis.

Appendix C

m, Regression Models Predicting Learning Outcomes with Significant In-game Predictors Only.

Predictors	Posttest (Near)			Posttest (Far)		
	B	β	p	B	β	p
(Intercept)	2.36	−0.29	**0.009**	2.00	0.15	**0.006**
Condition [After]	1.36	0.46	**0.017**	1.41	0.48	**0.006**
Condition [Before]	0.95	0.32	**0.043**	0.71	0.24	0.108
Solution Rate	3.56	0.23	**0.004**	4.48	0.30	**<0.001**
Levels Visited	0.05	0.23	**0.001**	0.04	0.19	**0.004**
Pretest (Near)	0.46	0.44	**<0.001**			
Posttest Form [B]	0.65	0.22	0.087	−2.08	−0.71	**<0.001**
Pretest (Far)				0.45	0.39	**<0.001**
Random Effects						
σ^2	4.90			4.56		
τ_{00}	0.40 $_{\text{Source}}$			0.00 $_{\text{Source}}$		
ICC	0.08					
N	5 $_{\text{Source}}$			5 $_{\text{Source}}$		
Observations	146			146		
Marginal R^2/Conditional R^2	0.430/0.473			0.480/NA		

Note. Bold faced = $p < .05$ or $p < .001$. The reference group is the Control condition. B = unstandardized coefficient. Model 4 yielded a singular fit due to zero variance in the random intercept for cohort.

Appendix D

Regression Models Predicting Subjective Perception Outcome Variables Controlling for Gameplay Success.

Predictors	Enjoy		Competence		Effort		Frustration		Value	
	B	p	B	p	B	p	B	p	B	p
(Intercept)	3.82	**<0.001**	3.00	**<0.001**	4.97	**<0.001**	4.71	**<0.001**	3.69	**<0.001**
Condition [After]*	0.61	0.053	0.71	**0.017**	−0.13	0.654	−0.79	**0.039**	0.31	0.444
Condition [Before]*	0.58	**0.027**	0.07	0.787	−0.02	0.918	−0.73	**0.020**	0.43	0.199
Solution Rate	0.64	0.349	2.57	**<0.001**	−0.11	0.857	−0.66	0.427	−0.09	0.921
Levels Visited	0.01	0.398	0.02	**0.016**	0.01	0.343	−0.01	0.579	0.01	0.466
Pretest	0.05	0.083	−0.02	0.339	0.03	0.213	0.04	0.233	0.05	0.178
Random Effects										
σ^2	1.53		1.37		1.20		2.34		2.46	
τ_{00}	0.08 $_{Source}$		0.04 $_{Source}$		0.06 $_{Source}$		0.02 $_{Source}$		0.10 $_{Source}$	
ICC	0.05		0.03		0.05		0.01		0.04	
N	5 $_{Source}$		5 $_{Source}$		5 $_{Source}$		5 $_{Source}$		5 $_{Source}$	
Observations	143		142		143		143		144	
Marginal R^2/Conditional R^2	0.107/0.152		0.165/0.189		0.029/0.075		0.045/0.052		0.039/0.076	

Note. Bold faced = $p < .05$ or $p < .001$.

* The reference group is Control condition. B = unstandardized coefficient.

References

Adams, D. M., & Clark, D. B. (2014). Integrating self-explanation functionality into a complex game environment: Keeping gaming in motion. *Computers & Education, 73*, 149–159.

Bainbridge, K., Shute, V., Rahimi, S., Liu, Z., Slater, S., Baker, R. S., & D'Mello, S. (2022). Does embedding learning supports enhance transfer during game-based learning? A case study with *Physics Playground*. *Learning and Instruction, 77*, 1–11. https://doi.org/10.1016/j.learninstruc.2021.101547

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65–79. https://doi.org/10.1016/j.compedu.2013.08.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*(6), 740–756.

Boldbaatar, N., & Şendurur, E. (2019). Developing educational 3D games with StarLogo: The role of backwards fading in the transfer of programming experience. *Journal of Educational Computing Research, 57*(6), 1468–1494. https://doi.org/10.1177/0735633118806747

Bruner, J. S. (1971). The process of education" revisited. *Phi Delta Kappan, 53*(1), 18–21.

Byun, J., & Joung, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science & Mathematics, 118*(3–4), 113–126. https://doi.org/10.1111/ssm.12271

Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of scaffolding in digital game-based learning on student's achievement: A three-level meta-analysis. *Educational Psychology Review, 34*(2), 537–574. https://doi.org/10.1007/s10648-021-09655-0

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. https://doi.org/10.3102/0034654315582065

Cloude, E. B., Carpenter, D., Dever, D. A., Azevedo, R., & Lester, J. (2021). Game-based learning analytics for supporting adolescents' reflection. *Journal of Learning Analytics, 8*(2), 51–72. https://doi.org/10.18608/jla.2021.7371

Conati, C., Jaques, N., & Muir, M. (2013). Understanding attention to adaptive hints in educational games: An eye-tracking study. *International Journal of Artificial Intelligence in Education, 23*(1), 136–161. https://doi.org/10.1007/s40593-013-0002-8

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

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.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (Vol. 1). Palgrave/Macmillan.

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media, 2*(1), 5–16.

Gresalfi, M. S., & Barnes, J. (2016). Designing feedback in an immersive videogame: Supporting student mathematical engagement. *Educational Technology Research & Development, 64*(1), 65–86. https://doi.org/10.1007/s11423-015-9411-8

Habgood, M., & Ainsworth, S. (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. https://doi.org/10.1080/10508406.2010.508029

Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming, 36*(4), 483–498. https://doi.org/10.1177/1046878105282276

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141–158. https://doi.org/10.1119/1.2343497

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

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

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13–24. https://doi.org/10.1016/j.iheduc.2004.12.001

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

Kuba, R., Rahimi, S., Smith, G., Shute, V., & Dai, C.-P. (2021a). Using the first principles of instruction and multimedia learning principles to design and develop in-game learning support videos. *Educational Technology Research and Development, 69*(2), 1201–1220. https://doi.org/10.1007/s11423-021-09994-3

Kuba, R., Shute, V., & Rahimi, S. (2021b). Students' perceived competence and extrinsic and intrinsic motivation in a physics educational game. *American Educational Research Association International Conference, Online*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lei, H., Chiu, M. M., Wang, D., Wang, C., & Xie, T. (2022). Effects of game-based learning on students' achievement in science: A meta-analysis. *Journal of Educational Computing Research.* , Article 07356331211064543. https://doi.org/10.1177/07356331211064543

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *Package 'emmeans*. https://github.com/rvlenth/emmeans.

Liao, C.-W., Chen, C.-H., & Shih, S.-J. (2019). The interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behavior patterns in a digital game-based learning environment. *Computers & Education, 133*, 43–55. https://doi.org/10.1016/j.compedu.2019.01.013

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. (2020). *Multimedia learning*. Cambridge University Press.

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*(3), 147–154. https://doi.org/10.1037/1076-898X.8.3.147

Melo, C., Madariaga, L., Nussbaum, M., Heller, R., Bennett, S., Tsai, C.-C., & van Braak, J. (2020). Editorial: Educational technology and addictions. *Computers & Education, 145*(C). https://doi.org/10.1016/j.compedu.2019.103730

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research & Development, 50*(3), 43–59.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94*(3), 598–610. https://doi.org/10.1037/0022-0663.94.3.598

O'Neil, H. F., Chung, G. K. W. K., Kerr, D., Vendlinski, T. P., Buschang, R. E., & Mayer, R. E. (2014). Adding self-explanation prompts to an educational computer game. *Computers in Human Behavior, 30*, 23–28. https://doi.org/10.1016/j.chb.2013.07.025

Pando Cerra, P., Fernández Álvarez, H., Busto Parra, B., & Iglesias Cordera, P. (2022). Effects of using game-based learning to improve the academic performance and motivation in engineering studies. *Journal of Educational Computing Research.* , Article 07356331221074022. https://doi.org/10.1177/07356331221074022

R Core Team. (2021). R: The R project for statistical computing. https://www.r-project.org/.

Rahimi, S. (2020). Inspire, Instruct, or Both? Game-based and support of creativity. *[Doctoral Dissertation]*. Florida State University.

Rahimi, S., Shute, V., Kuba, R., Dai, C.-P., Yang, X., Smith, G., & Alonso-Fernández, C. (2021). The use and effects of incentive systems on learning and performance in educational games. *Computers & Education, 165*, 1–17. https://doi.org/10.1016/j.compedu.2021.104135

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research, 49*(2), 181–196. https://doi.org/10.3102/00346543049002181

Rickards, J. P., & Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*(3), 354. https://doi.org/10.1037/h0036349

Ristanto, R. H., Kristiani, E., & Lisanti, E. (2022). Flipped classroom–digital game based learning (FC-DGBL): Enhancing genetics conceptual understanding of students in bilingual programme. *Journal of Turkish Science Education, 19*(1), 332–352.

Ritterfeld, U., Weber, R., Vorderer, P., & Bryant, J. (2005). Video games for entertainment and education. In *Playing video games: Motives, responses, and consequences* (pp. 399–413). Lawrence Erlbaum Associates Publishers.

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

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450–461. https://doi.org/10.1037/0022-3514.43.3.450

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*(4), 759–775. https://doi.org/10.1037/a0025140

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

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

Shute, V., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 43–58). New York, NY: Springer.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C.-P., … Sun, C. (2020a). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning, 37*(1), 127–141. https://doi.org/10.1111/jcal.12473

Shute, V., Smith, G., Kuba, R., Dai, C.-P., Rahimi, S., Liu, Z., & Almond, R. (2020b). The design, development, and testing of learning supports for the Physics Playground game. *International Journal of Artificial Intelligence in Education, 31*, 357–379. https://doi.org/10.1007/s40593-020-00196-1

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8.

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

Tsai, Y.-L., & Tsai, C.-C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education, 125*, 345–357. https://doi.org/10.1016/j.compedu.2018.06.020

Tsai, Y.-L., & Tsai, C.-C. (2020). A meta-analysis of research on digital game-based science learning. *Journal of Computer Assisted Learning, 36*(3), 280–294. https://doi.org/10.1111/jcal.12430

Tsigilis, N., & Theodosiou, A. (2003). Temporal stability of the intrinsic motivation inventory. *Perceptual & Motor Skills, 97*(1), 271–280. https://doi.org/10.2466/pms.2003.97.1.271

Vrugte, J. ter, Jong, T. de, Wouters, P., Vandercruysse, S., Elen, J., & Oostendorp, H. van (2015). When a game supports prevocational math education but integrated reflection does not. *Journal of Computer Assisted Learning, 31*(5), 462–480. https://doi.org/10.1111/jcal.12104

Westera, W. (2022). Examining the productiveness of failure in serious gaming. *International Journal of Technology Enhanced Learning, 14*(1), 66–87. https://doi.org/10.1504/IJTEL.2022.120561

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, X., Rahimi, S., Shute, V., Kuba, R., Smith, G., & Alonso-Fernández, C. (2021). The relationship among prior knowledge, accessing learning supports, learning outcomes, and game performance in educational games. *Educational Technology Research and Development, 69*(2), 1055–1075. https://doi.org/10.1007/s11423-021-09974-7