International Journal of Artificial Intelligence in Education
https://doi.org/10.1007/s40593-021-00250-6

ARTICLE

Assessing the Effects of Open Models of Learning and Enjoyment in a Digital Learning Game

Xinying Hou, et al. ***[full author details at the end of the article]***

Accepted: 22 March 2021/Published online: 13 April 2021

Abstract

Digital learning games are designed to foster both student learning and enjoyment. Given this goal, an interesting research topic is whether game mechanics that promote learning and those that promote enjoyment have different effects on students' experience and learning performance. We explored these questions in *Decimal Point*, a digital learning game that teaches decimal numbers and operations to 5th and 6th graders, through a classroom study with 159 students and two versions of the game. One version encouraged playing and learning through an *open learner model* (OLM, $N=$ 55), while one encouraged playing for enjoyment through an analogous *open enjoyment model* (OEM, $N=54$). We compared these versions to a control version that is neutral with respect to learning and enjoyment ($N=50$). While students learned in all three conditions, our results indicated no significant condition differences in learning outcomes, enjoyment, or engagement. However, the learning-oriented group engaged more in re-practicing, while the enjoyment-oriented group demonstrated more exploration of different mini-games. Further analyses of students' interactions with the open learner and enjoyment models revealed that students who followed the learner model demonstrated better in-game learning and test performance, while following the enjoyment model did not impact learning outcomes. These findings indicate that emphasizing learning or enjoyment can lead to distinctive game play behaviors, and that open learner models can be helpful in a learning game context. In turn, our analyses have led to preliminary ideas about how to use AI to provide recommendations that are more aligned with students' dynamic learning and enjoyment states and preferences.

Keywords Digital learning game · Decimal numbers · Mediation analysis · Student modeling · Open learner model · Enjoyment

An earlier version of this study was presented at the 21st International Conference on Artificial Intelligence in Education (AIED 2020) and published in Hou et al. (2020). In this work, we conduct additional analyses to investigate how students referred to and made gameplay choices based on the provided dashboards, as well as the impact of these choices.

Introduction

Recent advances in artificial intelligence (AI) in education have enabled many kinds of adaptive learning support in instructional platforms, ranging from individualized problem sequencing (Corbett & Anderson, 1994) to conversational agents (Lin et al., 2020). At the same time, there is active research on opening up the underlying models that guide these adaptivity features to the learners, in an effort to connect with them and promote system transparency. Such efforts have resulted in the development of open learner models (OLM - Bull, 2020; Bull & Kay, 2010), which are often accompanied by recommendation features that suggest the optimal learning pathways based on the learners' parameters (Dascalu et al., 2016). The use of these models has led to improved student learning in various domains and systems (Bodily & Verbert, 2017).

On the other hand, many of the current learner models and recommender systems are based on metrics of students' learning performance, such as their skill masteries and self-regulation (Hummel et al., 2007; Papamitsiou et al., 2018; Xie et al., 2019). It is less clear if grounding the adaptivity and recommendations in a different construct, such as affect or engagement, would result in better learning. This question is especially relevant in the area of digital learning games, which typically aim to promote both learning and enjoyment. While some studies have shown that game enjoyment is positively correlated with learning outcomes (Anderman & Dawson, 2011; Fu et al., 2009; Liu et al., 2011), others instead reported a trade-off, where games led to more enjoyment but less learning than traditional approaches (Greipl et al., 2018; Pittman, 2013; Plass et al., 2013).

Our research aims to elucidate this relationship between learning and enjoyment, by exploring and comparing the effects of recommender systems that optimize for each construct. In particular, we made use of adaptive dashboard technologies that capture data about students' learning or enjoyment and present this information back to the students, along with recommendations on how to maximize either factor. While data-driven adaptivity has been implemented in learning games in many forms, such as content generation (Hooshyar et al., 2018) and dynamic difficulty adjustment (Sampayo-Vargas et al., 2013), we chose this dashboard approach for several reasons. First, we would like to leverage the advantages of a transparent student assessment model, which has been shown to improve student engagement (Sarkar & Cooper, 2018). Second, the benefits of a learning-oriented dashboard have been validated in numerous studies on *open learner models* (see a recent review by Bodily et al., 2018), while the use of an enjoyment-oriented dashboard is a novel idea that we would like to explore. Finally, the use of suggestive, but not prescriptive, dashboard recommendations would allow us to examine when and how students made use of these recommendations, in order to better understand the effects of learning- and enjoyment-oriented dashboard design.

We created our study in the context of *Decimal Point*, a game that supports middle school students in learning about decimal numbers and their operations (McLaren et al., 2017). In *Decimal Point*, students can select from twenty-four mini-games to practice their decimal skills. Each mini-game is essentially one of five "game types," each targeting a different decimal operation. Our study of the game compared the learning- and enjoyment-oriented features through three conditions. The Learning-oriented Condition (LC) displays the student's current skill level across different decimal skills, through an open learner model, and recommends more playing of the mini-games they are weakest at. The Enjoyment-oriented Condition (EC) displays the student's current enjoyment levels, through an analogous open enjoyment model, and recommends more playing of the mini-games they enjoy the most. Finally, the Control Condition (CC) does not show any learning- or enjoyment-related information. In this setting, our research questions are as follows:

> RQ1: *Is there a difference in learning outcomes, self-reported enjoyment, or game play behavior between students in the three conditions?*
> RQ2: *In the Learning-oriented Condition, how is following the recommendation of the open learner model associated with learning outcomes and enjoyment?*
> RQ3: *In the Enjoyment-oriented Condition, how is following the open enjoyment model associated with learning outcomes and enjoyment?*

Through investigating these research questions, our work makes three contributions to the digital learning game and AI in education literature. First, we show that the learning- and enjoyment-oriented condition designs can lead to distinct game play patterns that have different implications for learning. Second, we present an in-depth analysis of the integration of an open learner model, as well as its analogous open enjoyment model, in a digital learning game context. Third, we derive general insights into the role of human-AI interaction in adaptive learning game features, by examining how students make use of and react to the learning- and enjoyment-oriented recommendations.

Background

AI-Based Adaptive Learner Support

The ability to provide adaptive support that caters to individual learner differences is one of the key characteristics of intelligent learning technologies (Koedinger et al., 2013). Based on a review by Aleven et al. (2016), modern learning systems can adapt to many psychological dimensions of the learner (e.g., knowledge, motivation, and self-regulation) at different time scales (during a problem-solving task, between tasks, or between iterations of the system). This wide range of adaptivity is enabled by a rich body of AI methodologies, ranging from traditional cognitive models (e.g., model tracing, example tracing, constraint-based tutoring) to data-driven machine learning techniques (e.g., Bayesian network, reinforcement learning, deep learning; see reviews by Aleven et al., 2016; Brusilovsky, 2001; Vandewaetere & Clarebout, 2014; VanLehn, 2016).

Adaptivity is also a popular feature of digital games, and often comes in the form of dynamic difficulty adjustment (Ang & Mitchell, 2019; Baldwin et al., 2016; Frommel et al., 2018; Zohaib, 2018), where the game attempts to match its level of difficulty with the player skill. The primary motivation of this feature is that players would be in their most engaged state, often referred to as *flow*, when the game challenges are closely aligned with their skills (Csikszentmihalyi, 1990); otherwise, players would feel bored if the game was too easy, or frustrated if the game was too difficult. While this focus on player engagement is different from the focus of learning systems, which aims to promote student knowledge and learning, there are many common approaches to implementing adaptivity in these two platforms, for instance with probabilistic methods (Bunian et al., 2018), reinforcement learning (Hagelback & Johansson, 2009), and neural networks (Li et al., 2010).

At the intersection of learning systems and digital games, digital learning games can benefit from the many advances in learner/player adaptive support in both areas. For example, in a physical education game, Xu et al. (2019) built algorithms to help the system select suitable training material according to students' real-time performance. Similarly, in *Trento Play&Go*, a gamified urban mobility system, Khoshkangini et al. (2017) constructed a system to automatically generate and recommend personalized challenges tailored to the students' preferences and history. As another example, in a mathematics game for 3rd graders, *ST Math*, Peddycord-Liu et al. (2017) used linear regression to uncover the predictive relationships between different math objectives, which led to a method for recommending the next math objective for students to play based on their current progress.

However, much of this modeling and adaptation process is performed behind the scenes, inaccessible to the students themselves. At the same time, research across different types of systems has uncovered the benefits of opening up the user assessment model, resulting in increased system transparency and user engagement (Bodily & Verbert, 2017; Bull & Kay, 2010; Malacria et al., 2013; Sarkar & Cooper, 2018). Therefore, in our work, we would like to experiment with a more transparent form of adaptivity, through a personalized dashboard that displays relevant assessment information to the students, in addition to providing game play recommendations.

Open Learner Models

The type of transparent dashboard we plan to incorporate has also been used in many educational systems, where it is often referred to as an *open learner model* (OLM), to display the internal learner model to students, with the goal of promoting their metacognitive abilities (Bodily et al., 2018; Bull, 2020). Here the key assumption is that students would use the open learner model to reflect on their learning progress and make decisions accordingly, leading to better self-regulation and learning outcomes. This proposition has been supported by a large number of empirical studies (e.g., Bodily et al., 2018; Bull et al., 2016; Jivet et al., 2018; Long & Aleven, 2013).

While OLMs are popular in intelligent tutoring systems and adaptive learning platforms, they have not seen wide adoption in learning games (e.g., Jasin et al., 2017). One of the earliest research in this area was conducted by Chen et al. (2007), who built a learning game, called *My-Pet-Our-Pet*, for Chinese idioms. The game's OLM is represented by a pet companion whose attributes are based on several learning dimensions. Results from this study indicated that the game materials alone did not lead to significant learning improvements, while the game materials in combination with OLM did. Follow-up research by Chen et al. (2011) further refined the in-game OLM, changing its representation from a pet companion to the student avatar, in order to help students understand their learning progress and enhance their feelings of self-awareness. As another example, Leonardou et al. (2019) combined an educational game for learning multiplication tables with OLM elements, and found that primary school students had positive reactions towards the OLM approach in games. However, these prior studies did not closely examine students' interactions with the OLM to see how frequently they consulted the model, and how this behavior impacted their learning. Addressing this area is one of the primary goals of our work.

A notable example of using OLMs to support both learning and self-regulation is *Lynnette* (Long & Aleven, 2016), an algebra tutor with gameful design elements. The tutor presents a dashboard that shows how well the student did on each problem type, while also offering immediate feedback on problem selections; students would receive positive feedback when selecting a problem type they needed to improve on, and negative feedback for selecting a type they have already mastered. Results from a randomized experiment by Long and Aleven (2016) showed that students learned better when they were allowed to select which problem types to do and received feedback on their selections, compared to when the system had full control over problem selection. As we will later discuss, the experimental design of *Lynnette* has key similarities with the *Decimal Point* study on which we base our analysis.

Learning and Enjoyment in Digital Learning Games

Digital learning games are instructional tools that aim to promote student learning through engaging game environments (Dondlinger, 2007; Gee, 2003; Young et al., 2012). A special characteristic of this type of environment is the focus on both learning and enjoyment. Learning is the primary objective and is fostered through evidence-based intervention techniques such as immediate feedback (Burgers et al., 2015) and self-reflection prompts (Moreno & Mayer, 2004). In addition, games improve students' motivation and engagement primarily by fostering enjoyment (Annetta et al., 2009; Moreno & Mayer, 2007; Tobias & Fletcher, 2007), and several studies have shown a positive correlation between enjoyment and learning outcomes (Anderman & Dawson, 2011; Fu et al., 2009; Liu et al., 2011). Engagement is often hypothesized to be one of the mechanisms through which digital learning games support learning (e.g., McLaren et al., 2017), and while this pathway has rarely been experimentally tested, one recent study found that students' disengaged behaviors mediated learning benefits (Richey et al., under review). This study was with our learning game, *Decimal Point,* compared to a non-game control (i.e., a computer tutor with the same math content, but presented in a standard, non-game manner).

On the other hand, there is a well-known challenge in maintaining the balance between the learning and enjoyment aspects of digital learning games (Kickmeier-Rust & Albert, 2010; Shute et al., 2019; Van Eck, 2006). According to Charsky and Ressler (2011), students who played a learning game enhanced with expert-generated concept maps reported lower motivation than their baseline motivation for regular classroom studies, because the concept maps focused students on the difficulty of learning and extrinsic reward in game playing, thereby negating the game's fun factor. In another study, Greipl et al. (2018) reported that, when learning basic math skills from a game, students reported more fun but made more estimation errors than when learning with paper and pencil; the authors also framed the connection between learning and enjoyment as a trade-off, where efforts to improve one factor may be detrimental to the other, rather than being synergistic and supportive.

While prior research has reported conflicting results regarding the relationship between learning and enjoyment, these studies also differ widely in their game design and instructional domain. To further elucidate this issue, we believe it would be enlightening to compare two different versions of the *same* learning game, where one version focuses on the learning aspect of the game and the other on the enjoyment aspect. To our knowledge, only a handful of prior studies in this direction have been conducted. For example, a study by Habgood and Ainsworth (2011) compared three variations of a math game: an intrinsic version with math content integrated into game elements, an extrinsic version which inserted math problems with symbolic representations into the game environment, and a control version where the learning materials have no connection with the game mechanics. Their results indicated that students learned more from the intrinsic version of the game under fixed time limits. As another example, Erhel and Jamet (2013) manipulated how undergraduate students perceived the same multimedia environment as either a learning module or a game. Based on their findings, the learning module group achieved deeper learning while reporting the same level of motivation as the game group, but performed worse than the game group when instructional feedback was added to both conditions, suggesting that a game environment can be helpful if it provides sufficient instructional support. Another study done by Wechselberger (2013) adopted a similar strategy with high school students and found that enjoyment is not affected by playful or serious framing.

Building on prior research, our goal was to experiment with a contrast between learning and enjoyment, one that relies on guiding students towards learning- or enjoyment-oriented goals. This work extends prior research on open learner models in intelligent tutoring systems by seeking to replicate the few prior studies that have implemented open learner models in digital learning games. Additionally, we have experimentally tested the effects of promoting enjoyment, a key hypothesized mechanism for supporting learning in digital learning games that has, in practice, produced mixed results when it has been manipulated through a game's content representation of students' prior perspectives. There is reason to expect both the open learner model and the open enjoyment model would support learning, but through different mechanisms (i.e., the open learner model through targeted practice and the open enjoyment model through increased enjoyment and engagement).

Decimal Point

Decimal Point is a 2D single-player web game that can be played in a browser on a PC or tablet. The game features a fantasy world where the players, who are middle school students, travel through the Decimal Point Amusement Park to help alien friends learn about decimal numbers and their operations. As shown in Fig. 1, the game map consists of eight theme areas (e.g., *Haunted House*, *Wild West*, *Space Adventure*), each with its own leitmotif and mini-games (e.g., *Enter If You Dare* in the *Haunted House* and *Lasso the Bronco* in the *Wild West*). While the game map is designed to facilitate an immersive experience, game play occurs inside the mini-games, where students complete a variety of playful activities (e.g., entering a haunted house, shooting western targets, launching a spaceship) that each connect with a type of decimal exercise. Based on the nature of its activities, each mini-game is characterized by one of the five game types in Table 1. Each game type was designed to target one of four established

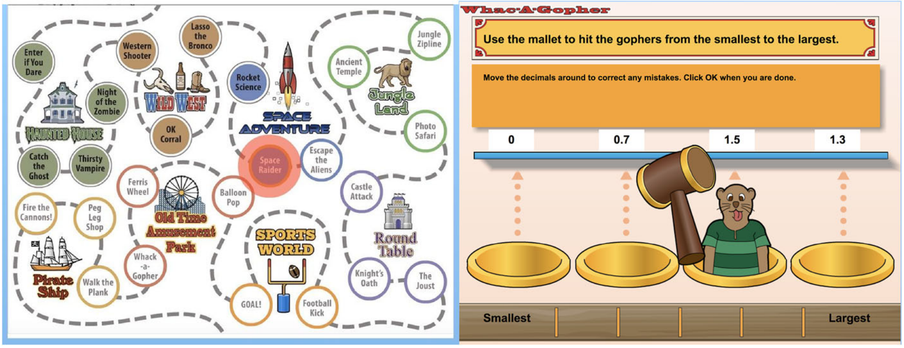


Fig. 1 The game map where students can select among 24 mini-games to play (left), and an example mini-game, *Whack-A-Gopher*, in the *Sorting* type and *Old Time Amusement Park* theme (right)

decimal misconceptions: (1) longer decimals are larger, (2) shorter decimals are larger, (3) the integer and decimal portions are independent, and (4) decimals smaller than 1.0 are negative (McLaren et al., 2017).

As an example, in the mini-game *Whack-A-Gopher* (Fig. 1), students have to hit the four gophers in the correct order based on their associated number labels. These number labels were set up to target the misconception that decimals smaller than 1.0 are negative (Isotani et al., 2010; Stacey et al., 2001). They also need to be quick in thinking and acting, as the gophers pop up and retreat at random times. Once the four gophers have been hit, students receive immediate feedback about the correctness of their ordering, and can rearrange the number labels if they are incorrect. After successfully finishing this activity, students are prompted to self-explain their answer by selecting from a multiple-choice list of possible explanations. This stage is based on prior research that has demonstrated the learning benefits of self-explanations (Chi et al., 1989, 1994), including in digital learning games (Johnson & Mayer, 2010). Every mini-game has a similar outline of problem-solving embedded in game activities, followed by self-explanation. Students don't face any penalty on incorrect responses and can resubmit answers as many times as needed; however, they are not allowed to move forward without correctly solving all the problems in the mini-game they chose. More details about the instructional content of the mini-game problems can be found in McLaren et al. (2017).

Table 1 The list of game types and their game activities in *Decimal Point*

Game type	Activity
Number Line	Locate the position of a decimal number on the number line
Addition	Add two decimal numbers by entering the carry digits and the sum
Sequence	Fill in the next two numbers of a sequence of decimal numbers
Bucket	Compare given decimal numbers to a threshold number and place each number in a "less than" or "greater than" bucket.
Sorting	Sort a list of decimal numbers in ascending or descending order

An initial study of *Decimal Point*, where students had to play all mini-games in a canonical order, showed that the game yielded more learning and enjoyment than a conventional tutor with the same instructional content (McLaren et al., 2017). Subsequent studies have experimented with providing agency, i.e., the freedom to select which mini-games to play and when to stop (Harpstead et al., 2019; Nguyen et al., 2018), to students. These studies revealed no differences in test scores or enjoyment between students who were and were not offered agency, but Harpstead et al. (2019) found that students in the former group had the same learning gains while completing fewer mini-games than the latter, suggesting that *Decimal Point*, in its canonical form, may contain more learning content than it requires and that students are able to self-regulate successfully in determining when to quit playing. Our study builds on these prior studies by retaining the agency feature while also providing students with learning- or enjoyment-oriented recommendations to aid them in decision making throughout their game play.

Context and Methods

In order to compare and evaluate the effectiveness of the learning- and enjoyment-oriented game features in *Decimal Point*, we conducted a study with 5th and 6th grade students in two public schools in a mid-sized U.S. city. 196 students originally participated in the study, which was conducted during students' regular class times and lasted six days. The materials included a pretest, game play, evaluation questionnaire and posttest during the first five days, followed by a delayed posttest one week later. Participants completed the pretest and demographic questionnaire on the first day, played the game in three class days, then completed an evaluation survey and posttest right after finishing the game, as well as a delayed posttest one week later. They had one hour of class time per day to go through the above activities, but typically took fewer than two hours to finish the game (M = 1.61 h, SD = 0.57). After the study, 35 students were removed from our analyses due to not finishing all of the materials. Using the outlier criteria from a prior study in *Decimal Point* (Nguyen et al., 2018), we excluded two students whose gain scores from pretest to posttest were 2.5 standard deviations away from the mean (M = 5.27, SD = 6.00). Thus, our final sample included 159 students (82 males, 77 females), whose age range was from 10 to 12 years old (M = 10.94, SD = 0.64). The full log data from the study is archived in the DataShop repository (Koedinger et al., 2010), in dataset number 3086.[1]

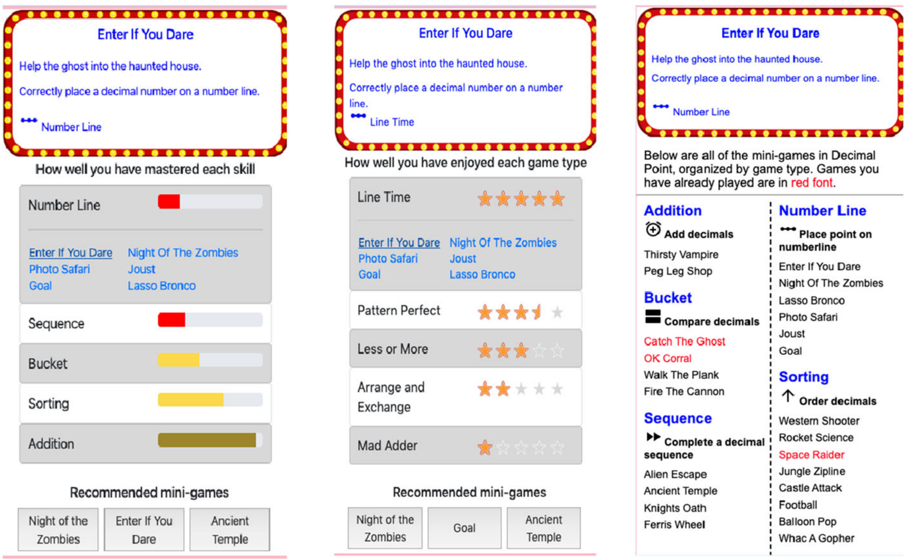
Intervention Design

Each student was randomly assigned to one of three conditions: Learning-oriented Condition (LC; N = 55), Enjoyment-oriented Condition (EC; N = 54), or Control Condition (CC; N = 50). Students could select the mini-games to play in any order, where a mini-game round is defined as a complete play through the decimal problems and self-explanation question in that mini-game. In both the LC and EC settings, students could choose to stop playing any time after completing at least 24 rounds. In the CC setting,

[1] https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=3086

which was equivalent to the High Agency condition in Nguyen et al. (2018) and Harpstead et al. (2019), students had the option to play another round of each after finishing the first two rounds of all 24 mini-games. Additionally, each condition featured a different dashboard (see Fig. 2) attached to the main game map in Fig. 1. After finishing a mini-game round, students would be taken back to the game map, where they could make their next mini-game selection from the updated dashboard.

The Learning-oriented Condition was designed to encourage students to play the game types that they needed the most improvement on. After the student finished a mini-game round and returned to the game map, the open learner model dashboard (Fig. 2a) would display their updated mastery of each game type, based on Bayesian Knowledge Tracing (BKT - Corbett & Anderson, 1994). BKT represents student learning of a targeted skill by a Hidden Markov Model with two states: Mastered and Unmastered. A standard BKT model has four parameters: (1) p_{init}: the probability of starting at the Mastered state (i.e., knowing the skill before interacting with the system), (2) p_{transit}: the probability of transitioning from Unmastered to Mastered, (3) p_{slip}: the probability of making an incorrect answer while being in the Mastered state (i.e., slipping), and (4) p_{guess}: the probability of making a correct answer while being in the Unmastered state (i.e., making a lucky guess). The goal of BKT is to infer the value of p_{learn}, which denotes the probability of the student being in the Mastered state, from a sequence of observations about their answer correctness (Yudelson et al., 2013). Based on prior work, the BKT parameters in our study were initially set to $p_{\mathrm{init}} = 0.4$, $p_{\mathrm{transit}} = 0.05$, $p_{\mathrm{slip}} = p_{\mathrm{guess}} = 0.299$ (Baker, R. - personal correspondence).



(a) (b) (c)

Fig. 2 The dashboards shown along with the game map in the (**a**) Learning-oriented, (**b**) Enjoyment-oriented, and (**c**) Control condition. Clicking on a game type revealed the mini-games in that game type, which could be selected for playing. The game types in the Enjoyment-oriented Condition (b) were renamed in keeping with maximizing enjoyment and to be more playful, e.g., *Addition* became *Mad Adder*. The equivalent game types are next to one another in Figs. 2(a) and 2(b)

In the Learning-oriented Condition, if a game type had not been played so far, it would start at the initial mastery value of $p_{\mathrm{init}} = 0.4$. If the mastery value p_{learn} of any game type changed after the most recent mini-game round, there would be a short animation of the skill bar increasing or decreasing to draw the student's attention to this change. The game types were also ordered so that those at the top had the *lowest* mastery value. In addition, the dashboard selected three mini-games from the two least-mastered game types to recommend to the student. For example, in Fig. 2a, the first two recommended mini-games (*Night of the Zombies* and *Enter If You Dare*) came from the top game type (*Number Line*), and the third recommended mini-game (*Ancient Temple*) came from the second top game type (*Sequence*); the selection of specific mini-games within a game type was random. This recommendation was only displayed after a student had played three mini-game rounds, when initial data about their learning would be available. Students in this condition could either follow the recommendation or make their own selection. To encourage interaction with the dashboard, mini-game selection from the game map was not allowed.

The Enjoyment-oriented Condition was designed to encourage students to play the game types for which they previously indicated their preference. After finishing a mini-game round, each student was asked to rate their enjoyment of that mini-game, on a scale from 1 ("not fun") to 5 ("crazy fun"). The rating interface (Fig. 3) is based on the prior work of Read and MacFarlane (2006) in which they designed a "fun-o-meter" used to collect children's ratings on playful activities. The open enjoyment model dashboard then displayed the student's enjoyment score of each game type, which was the average of all their mini-game ratings in that type, rounded up to half a star (Fig. 2b). If the enjoyment score of any game type changed after the most recent mini-game round, there would be a short animation of the stars being filled or emptied to draw the student's attention to this change. If a game type had not been rated yet, it would start at the zero star position. The game types were also ordered so that those at the top had the *highest* enjoyment score. Similar to the Learning-oriented Condition, two mini-games in the top game type and one in the second top type were also recommended (chosen

Fig. 3 A screenshot of the fun-o-meter that asked students in the Enjoyment-oriented Condition to rate their enjoyment of the most recent mini-game round

randomly within type); students could follow the recommendation or make their own selection from the dashboard, but not from the game map (which was done to be sure students paid attention to the open enjoyment model dashboard).

The Control condition was equivalent to the High Agency condition in prior studies by Nguyen et al. (2018) and Harpstead et al. (2019), where students played two rounds of mini-game per selection (i.e., they played each selected mini-game twice, with different content but the same mechanics). The dashboard listed the mini-games and their corresponding skills, where the completed ones were highlighted in red (Fig. 2c). However, in this condition, no information about the student's learning or enjoyment state was provided. After finishing the first two rounds of all 24 mini-games, students had the option to play another round of each.

Assessment

In addition to gameplay, students were also asked to complete pre, post, and delayed tests of their decimal knowledge as well as a questionnaire about their enjoyment of the experience.

Pretest, Posttest, and Delayed Posttest. Each student received an online pretest before game play on the first day, a posttest immediately after game play, and a delayed posttest one week after the posttest. Each test consists of 43 questions. Most questions were worth one point each, while some multi-part questions were worth several points, for a total of 52 points per test. The questions were designed to probe for specific decimal misconceptions and involved either one of the five decimal activities targeted in Table 1 or conceptual questions (e.g., "Is a longer decimal number larger than a shorter decimal number?"). Three test forms (A, B and C) that were isomorphic and positionally counterbalanced across conditions were used. A series of one-way ANOVAs showed no difference in terms of performance among the three versions of the test at pretest, $F(2, 156) = 0.480$, $p = .620$, posttest, $F(2, 156) = 1.496$, $p = .227$, or delayed posttest, $F(2, 156) = 1.302$, $p = .275$.

Questionnaires and Survey Immediately after finishing the game on the last day of game play, students in all three conditions were asked to rate several statements about their enjoyment of the experience on a Likert scale from 1 ("strongly disagree") to 5 ("strongly agree"). These statements are based on existing measurement scales and pertain to three enjoyment constructs: multidimensional engagement, game engagement, and the enjoyment dimension of achievement emotions (Table 2). In the multidimensional engagement construct, we excluded the behavioral/cognitive engagement subscale from analysis, due to its low reliability, and only reported the results for affective engagement. We then averaged the ratings for individual items in each construct to produce a representative score for that construct. In addition, the ratings for negative statements such as "I felt frustrated or annoyed" were reverse coded so that, across all constructs, a higher score indicates more enjoyment. We refer to these scores as *post-game enjoyment scores*, to distinguish them from the mini-game ratings (Fig. 3), which were collected only in the Enjoyment-oriented Condition. After the game, students were also asked to reflect on their game play behavior, e.g. "How many mini-games did you play, and why did you play this number of mini-games?"

Table 2 Post-intervention survey items

Construct (item count)	Example statement	Cronbach's α
Affective engagement (3) (Ben-Eliyahu et al., 2018)	I felt frustrated or annoyed.	.78
Behavioral/cognitive engagement (3) (Ben-Eliyahu et al., 2018)	I tried out my ideas to see what would happen.	.54
Game engagement (5) (Brockmyer et al., 2009)	I lost track of time.	.74
Achievement emotion (6) (Pekrun, 2005)	Reflecting on my progress in the game made me happy.	.89

Results

We first checked whether students learned by playing *Decimal Point*. A repeated-measures ANOVA showed a significant difference for all students, in all conditions, between pretest and posttest scores, $F(1, 158) = 132.882$, $p < .001$, as well as between pretest and delayed posttest scores, $F(1, 158) = 239.414$, $p < .001$. In other words, in all three conditions students' performance improved after playing the game. Next, we investigated our main research questions. Given that the conditions (CC, LC and EC) were randomly assigned, we did not expect systematic differences on the pretest based on condition; we used analyses of covariance (ANCOVA) to assess condition effects on posttest and delayed posttest while controlling for individual variation in pretest.

Condition Effect on Learning and Enjoyment

> RQ1: *Is there a difference in learning outcomes, self-reported enjoyment, or game play behavior between students in the three conditions?*

Descriptive statistics about students' test scores and post-game enjoyment ratings in each condition are included in Table 3. From a one-way ANOVA, we observed no significant differences across conditions in pretest scores, $F(2, 156) = 1.915$, $p = .151$. With pretest score as a covariate, an ANCOVA showed no significant condition differences in posttest scores, $F(2, 155) = 0.201$, $p = .818$, or delayed posttest scores,

Table 3 Descriptive statistics of test performance and self-reported enjoyment scores by condition

Category	CC	EC	LC
Pretest scores M (*SD*)	26.68 (8.89)	24.76 (9.55)	23.09 (9.65)
Posttest scores M (*SD*)	32.12 (8.01)	29.76 (10.25)	28.42 (11.31)
Delayed posttest scores M (*SD*)	32.84 (8.90)	31.74 (10.12)	30.05 (10.06)
Achievement emotion M (*SD*)	3.46 (1.02)	3.49 (0.88)	3.55 (0.94)
Game engagement M (*SD*)	3.00 (0.90)	3.14 (0.98)	3.18 (0.80)
Affective engagement M (*SD*)	3.66 (0.94)	3.42 (1.04)	3.58 (0.85)

$F(2, 155) = 0.143$, $p = .867$. Similarly, based on a series of one-way ANOVA, there were no significant differences across conditions in achievement emotions, $F(2, 156) = 0.118$, $p = .889$, game engagement, $F(2, 156) = 0.597$, $p = .552$, or affective engagement, $F(2, 156) = 0.886$, $p = .414$. In other words, there was no condition effect on learning or self-reported enjoyment.

As students were able to make their own mini-game selections in all three conditions, our measure of game play behavior is based on three factors: the number of mini-game rounds that students played in each condition, the variety of mini-games played, and the frequency of mini-game type switching. Because CC students could not replay mini-games until after they had played 48 rounds, and were required to play two rounds of mini-game per selection, their game play behavior measures were necessarily different from those in LC and EC, so we focused our comparisons on the LC and EC groups. First, a Kruskal-Wallis test showed significant differences between the two conditions in the number of rounds where the LC students ($M = 33.20$, $SD = 9.86$) played significantly more rounds than the EC students ($M = 26.65$, $SD = 4.59$), $p = .002$. In short, the EC students tended to play fewer rounds of the mini-games compared with students in LC.

Second, we defined a new metric for each student called *replay rate*, which is the number of times a student reselected a mini-game beyond the first try divided by their total number of mini-game selections. A high replay rate (close to 1) indicates that the student played more rounds of the same mini-games, while a low rate (close to 0) indicates the student played fewer rounds of more mini-games (i.e., playing a wider variety of mini-games). We employed a Kruskal-Wallis test and observed significant differences in replay rates between the LC and EC students, $H = 42.41$, $p < .001$; LC students ($M = 0.44$, $SD = 0.20$) had a significantly higher replay rate than EC students ($M = 0.15$, $SD = 0.17$). In other words, LC students tended to replay more rounds of the mini-games they had already played than those in EC. Preliminary analysis of students' reflections on their game play behavior revealed a similar picture. In their responses to the post-game questionnaire item "How many mini-games did you play, and why did you play this number of mini-games?", many students in the EC group (25/54) mentioned trying out every available mini-game, e.g., "I really wanted to finish the whole map and see all the things filled in with color." On the other hand, fewer LC students (10/55) touched on this idea, while 17 of them instead mentioned the mastery scores as motivation for playing, e.g., "I was trying to get all the decimal category skill bars full."

Third, we defined a new metric for each student called *interleaving rate*, which is the number of times a student switched the mini-game type between two consecutive mini-game selections, divided by the total number of times they could have switched. A high interleaving rate (close to 1) indicates that the student made more mini-game type switches, while a low rate (close to 0) indicates the student played a particular mini-game type through multiple mini-game rounds. A one-way ANOVA showed significant differences in interleaving rate between the LC and EC group, $F(1, 107) = 28.20$, $p < 0.001$. LC students ($M = 0.46$, $SD = 0.20$) had a significantly lower interleaving rate than EC students ($M = 0.66$, $SD = 0.19$). In other words, students in EC tended to interleave the game types while playing more than those in LC.

Student Interaction with the Open Learner Model

To address our remaining two research questions, we first need to define the expected behavior from the open learner or enjoyment model. As previously described, both models encouraged more playing of specific game types by positioning these types at the top of the dashboard. In the open learner model, the top game types involved decimal operations for which the student was weakest; in the open enjoyment model, the top game types were those the student enjoyed the most, based on their own mini-game ratings. In addition, three mini-games were also recommended to the student at each turn after the first three rounds, with the specific mini-games chosen randomly from the top two game types. Therefore, we defined the behavior of following the learner or enjoyment model as selecting a mini-game from the top two game types on and after the fourth round. For each student, we then computed the *model following rate*, which is the number of times this behavior occurred divided by the maximum number of times it could have occured (i.e., the student's total number of mini-game rounds minus three). This metric reflects the frequency with which a student followed the learner or enjoyment model's recommendation in the course of their game play. The mean and standard deviation of model following rate were $M = 0.47$, $SD = 0.20$ in the Learning Condition ($N = 55$) and $M = 0.46$, $SD = 0.16$ in the Enjoyment Condition ($N = 54$). From a one-way ANOVA, we observed no significant difference in the model following rate between conditions, $F(1, 107) = 0.272$, $p = 0.603$.

> RQ2: *In the Learning-oriented Condition, how is following the recommendation of the open learner model associated with learning outcomes and enjoyment?*

As the open learner model's recommendations are based on the student's in-game learning measure (i.e., the p_{learn} mastery values of the five game types), we wanted to explore the relationship between following the model, in-game learning outcomes and post-game learning outcomes, i.e., posttest and delayed posttest scores. For each student, we measured in-game learning outcomes by looking at the final mastery values of the five game types, by the time they stopped playing, and recording the minimum of these values, i.e., the *minimum final mastery*. For example, one student's final skill mastery values for the five game types are 0.99 for *Number Line*, 0.99 for *Addition*, 0.84 for *Sorting*, 0.40 for *Bucket*, and 0.30 for *Sequence*. In this case, their minimum final mastery would be the lowest of these five values, i.e., 0.30. Our rationale for this metric is that a student who followed the open learner model more would get more practice with their weak skills, and thus be able to raise their mastery in all five decimal skills. Least mastered skill areas have also been used to measure overall progress in cognitive tutors (Long & Aleven, 2017) and to guide curriculum reform (Cajimat et al., 2020) in prior studies. Furthermore, as the posttest and delayed posttest covered all five skills in roughly equal proportions, having a balanced mastery (corresponding to a high minimum final mastery) was more beneficial than being strong in certain skills but weak in others (corresponding to a low minimum final mastery).

Under this operationalization, we then conducted a mediation analysis with the model following rate as an independent variable, minimum final mastery as a mediator, and posttest/delayed posttest score as the dependent variable. Following the standard practice of controlling for prior knowledge when analyzing posttest scores (Whitley &

Kite, 2013), we also included pretest score as a covariate in the mediation model. The indirect effect's confidence interval was then estimated at the 5% significant threshold, using bias-corrected non-parametric bootstrapping with 2000 iterations (Hayes & Rockwood, 2017; Vallat, 2018). In case a mediation effect was present, we reported its effect size through the ratio of the indirect to total effect, i.e., the mediation ratio, which indicates the proportion of the total effect that is mediated (Preacher & Kelley, 2011).

Results of the mediation analysis showed that the effect of the model following rate on posttest score was mediated by the minimum final mastery (Fig. 4). The regression coefficient between the model following rate and minimum final mastery was significant, as was the regression coefficient between the minimum final mastery and posttest score. Results of bootstrapping procedures also showed that the indirect effect was significant ($ab = 6.556$, 95% CI [2.710, 12.782], $p < .001$), with a mediation ratio of 6.556 / 8.714 = 75.24%. Similarly, the relationship between the model following rate and delayed posttest score was mediated by the minimum final mastery ($ab = 6.866$, 95% CI [2.597, 12.467], $p < .001$), with a mediation ratio of 6.345 / 6.866 = 92.41%. In other words, following the open learner model more frequently led to better test performance by supporting more balanced mastery of game content.

To better visualize the significant mediation effects, we constructed a scatter plot of the independent, mediator and dependent variable in each model. Figure 5 shows that

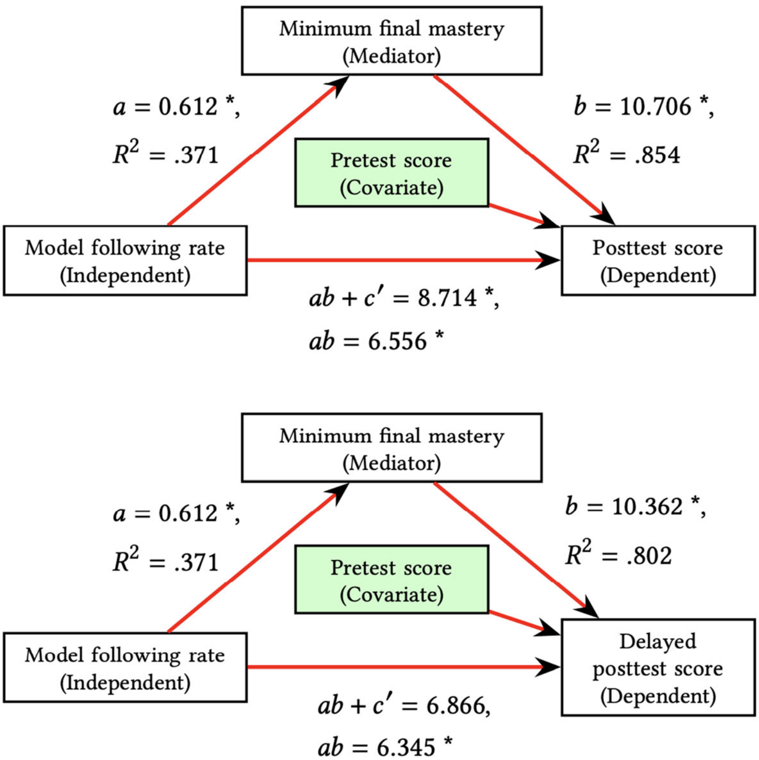


Fig. 4 Diagram of mediation model for posttest score (top) and delayed posttest score (bottom). * indicates significance at the .05 level

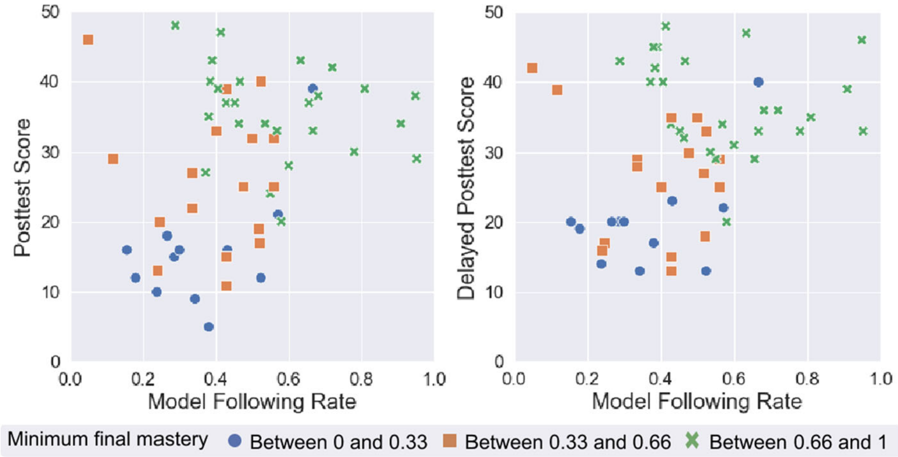


Fig. 5 Scatter plot of the relationship between the following rate and posttest score (left) and delayed posttest score (right)

students who followed the recommendations more frequently tended to attain higher levels of minimum final mastery and have higher posttest scores, while students who followed less frequently tended to attain lower levels of minimum final mastery and have lower posttest scores.

We also examined the relationship between following the open learner model and students' post-game enjoyment ratings. As the learner model did not involve any enjoyment-related information, we only conducted a Pearson correlation analysis to see if, in this context, following the model was correlated with higher ratings in any enjoyment construct. However, our results revealed no significant correlation between students' model following rates and their affective engagement ($r = .07$, $p = .594$), game engagement ($r = .03$, $p = .809$) or achievement emotion ratings ($r = .02$, $p = .884$).

Student Interaction with the Open Enjoyment Model

> RQ3: *In the Enjoyment-oriented Condition, how is following the open enjoyment model associated with learning outcomes and enjoyment?*

As the open enjoyment model's recommendations are based on the student's in-game enjoyment (i.e., the mini-game ratings from the fun-o-meter in Fig. 3), we wanted to explore the relationship between following the model, in-game enjoyment and post-game enjoyment rating in each enjoyment construct. For each student, we measured in-game enjoyment by looking at the average of all mini-game ratings that they provided in the course of game play. Our rationale for this metric is that a student who followed the open enjoyment model more frequently would play more mini-games in the types that they enjoyed the most and, in turn, assigned higher ratings to those mini-games, leading to a higher *average mini-game rating* overall.

Under this operationalization, we then conducted a mediation analysis with the model following rate as an independent variable, average mini-game rating as a mediator, and post-game enjoyment rating as the dependent variable. The resulting models indicated no significant mediation effect of average mini-game ratings in the relationship between the model following rate and affective engagement (ab = 0.051, 95% CI [−0.334, 0.715], $p = .753$), game engagement (ab = 0.090, 95% CI [−0.652, 0.922], $p = .782$) or achievement emotion rating (ab = 0.089, 95% CI [−0.782, 0.770], $p = .782$). In each dimension, the total effect, without accounting for the mediator, was likewise not significant: $c = 1.028$, $p = .265$ for affective engagement; $c = -1.261$, $p = .146$ for game engagement; $c = -0.385$, $p = .626$ for achievement emotion.

We also examined the relationship between following the open enjoyment model and post-game learning performance, based on posttest and delayed posttest scores. As the enjoyment model did not involve any learning-related information, we only conducted a Pearson correlation analysis with pretest scores as covariates to see if, in this context, following the open enjoyment model was correlated with higher test scores. However, our results revealed no significant correlation between students' model following rates and their posttest scores ($r = .03$, $p = .824$) or delayed posttest scores ($r = .14$, $p = .319$).

To better understand why the mediation effect was not present, we looked at the distribution of individual mini-game ratings, which the open enjoyment model relied on, in order to rank the game types. For each mini-game selected by a student, we recorded whether it was from their top two game types at the time of selection, and how they rated it after playing (from 1 to 5 stars). Based on Fig. 6, we found that the majority of ratings (about 60%) were 5-star ratings, regardless of whether the selected game was among the top two types or not. In other words, students were very likely to provide a maximum rating to any mini-game they played; consequently, their average mini-game ratings were heavily skewed towards the higher end, which explains why this metric did not yield a significant mediation effect.

Fig. 6 Stacked bar chart of mini-game ratings for the mini-games in the top two game types at the time of selection and for those not in the top two game types

Discussion

In this study, we investigated the impact of emphasizing the learning or enjoyment aspect of the digital learning game *Decimal Point* through a comparison of three game versions: the Enjoyment-oriented Condition (EC), the Learning-oriented Condition (LC) and the Control Condition (CC). Our analysis was motivated by whether optimizing for enjoyment would positively or negatively impact learning in a digital learning game, given that enjoyment is often posed as a trade-off to learning (Greipl et al., 2018), but can also be a conducive factor that supports learning by increasing engagement with the instructional content of the game (Giannakos, 2013; Richey et al., under review). This topic is especially relevant for in-class studies, where students' sense of enjoyment may be lacking due to the classroom environment and teacher expectations (Squire, 2005). Beyond broad intervention impacts, we also examined how students interacted with the open models, i.e., whether they used the provided information to make game play decisions that were beneficial to their learning or enjoyment. In turn, we aimed to derive actionable insights on the strengths and weaknesses of the current dashboard designs.

Overall, results from our study indicated that there were no condition differences in post-game test performance or enjoyment scores between the three conditions. There were, however, differences in game play patterns, where EC students had the least number of mini-game rounds and significantly lower replay rates than LC students. In addition, while students in EC and LC followed the open models' recommendations at similar rates (about 50% of the times), following the open learner model led to better in-game learning and post-game performance, while following the open enjoyment model did not.

Condition Effect on Learning and Enjoyment

From the learning perspective, the EC students, who did not have access to the open learner model, may not have closely monitored their learning progress (Bull & Kay, 2008; Zimmerman, 2000) and more likely wanted to explore all the mini-games offered in *Decimal Point*. In contrast, the LC students could see their skill performance and therefore were potentially more motivated to focus on mastering all of the skills, as reflected in their replay and interleaving rates. Further evidence is provided by students' post-game reflections, which indicated that the EC students liked to play all the mini-games to "see all the things filled with colors," while the LC students wanted to improve their skill masteries to "get all the decimal category skill bars full"; of particular note is how students referred to interface elements in the enjoyment- and learning-oriented dashboards as motivation for their game play behavior. More generally, our finding suggests that in a game environment where students have the agency to choose between different types of tasks, showing an open learner model can encourage re-practicing tasks, while showing an open enjoyment model may prompt students to engage in more exploration of the different tasks.

To discuss the learning implications of these two game play patterns, we would draw a connection between them and the concept of *interleaved practice*, in which exposure to learning contexts (e.g., mini-game types) are interleaved, so that consecutive questions belong to different contexts (Rohrer, 2012). Compared with the LC group, students in the EC group engaged in more interleaving at both the mini-game level (based on their replay rates) and the game type level (based on their interleaving rates). On the other hand, the LC group, who had lower interleaving rates, were potentially engaging in blocked practice, i.e., finishing all of the mini-games of one game type before moving to the next. There has been a rich literature on the benefits of interleaved practice in mathematics learning, especially when compared with blocked practice (Foster et al., 2019; Maass et al., 2015; Patel et al., 2016). However, results from our study indicated that the EC group, who demonstrated more interleaving, did not learn more than the LC group. One potential reason, as pointed out by Carvalho and Goldstone (2019), is that the effect of interleaved practice can be influenced by many factors such as students' skill level, age, and study times. Therefore, future studies that directly manipulate the levels of interleaving for students are needed to derive a more definitive finding about the boundary conditions of the interleaving effect (Carpenter, 2014).

From the enjoyment perspective, our EC design did not yield the intended effect of maximizing students' enjoyment and engaging them in the game for a longer time, compared to LC and CC. In fact, the EC students played the least number of rounds ($M = 26.65$, $SD = 4.59$) and had a low replay rate ($M = 0.15$, $SD = 0.17$), indicating that they chose to stop playing after trying most of the 24 unique mini-games once. One potential reason is that, as suggested by Lomas et al. (2017), novelty is among the primary predictors of engagement in digital learning games. EC students were able to experience all of the mini-games sooner due to their exploration behavior, and by this point, there was no other novelty factor to keep them playing. The low replay rate in the EC condition suggests students did not perceive the games to be enjoyably replayable, which could discourage them from replaying game types as much as they needed to master skills. In addition, our study was conducted in a real classroom environment, where students had limited time per day to play the game and were aware of the posttests; these factors may have negated the playful atmosphere that the Enjoyment-oriented condition was intended to induce (Osman & Bakar, 2012; Rice, 2007) or caused students not to take the enjoyment model as seriously as the learner model. Incorporating survey questions that ask students about how they used the open models would allow us to validate this conjecture.

Student Interaction with the Open Learner Model

A key result in the Learning-oriented Condition is that students who more assiduously followed the open learner model acquired a higher skill floor across all the decimal operations targeted by the five game types, as indicated by their minimum final skill mastery, which in turn led to higher posttest and delayed posttest scores. Furthermore, with the help of the dashboard, students in the Learning-oriented Condition were able to identify and practice with the mini-games that corresponded to their two weakest decimal skills quite frequently, at an average rate of 47%. While self-regulated learning (SRL) has been a driving factor in the adoption of open learner models, which aim to encourage students to take control of their learning and, in particular, encourage them to practice their least-mastered skills, it has been unclear whether this theory holds in the context of digital learning games, where students' sense of agency (Reeve et al., 2003), contextual autonomy (Deterding, 2016) and expectation (Wardrip-Fruin et al.,

2009) all play a crucial role in shaping their experience. In particular, if students feel that their game choices are not meaningful, or that their engagement with the game is mediated by external factors such as teacher control, their motivation to engage in behavioral regulation may be diminished. Therefore, our results are novel in their implication that the self-regulation support from an open learner model can indeed be beneficial in a learning game environment such as *Decimal Point*.

The follow-up question, then, is which factors may lead to the open learner model's effectiveness in *Decimal Point*. Our conjecture is that the game's relatively simple design and clear objective (i.e., help the alien friends learn about decimal numbers) were helpful in aligning the student's expectation with the game activities from the beginning of game play. Furthermore, from the perspective of self-determination theory (Przybylski et al., 2010), *Decimal Point* promotes both autonomy (students can choose which mini-games to play and when to stop) and perceived competence (through the open learner model's display of skill mastery values), which are two primary factors in fostering intrinsic motivation. In addition, the game's feedback, including the immediate corrective feedback when students solve mini-game problems and the skill meter animation (i.e., gradually increasing or decreasing) after each mini-game, also constitutes a strong support for perceived competence (Deci et al., 1991). Perceptions of competence and control may increase students' preference for challenge (Boggiano et al., 1988), which in this version of the game were clearly indicated through the open learner model recommendation system.

On the other hand, we did not observe any significant correlation between students following the model and their post-game enjoyment, in terms of affective engagement, game engagement or achievement emotion. While we expected students to enjoy games that involved more challenging decimal skills to them, such as those for which they attained lower mastery, prior evidence suggests that this is only the case if they feel like their engagement is voluntary and view it more as play than as work (Abuhamdeh & Csikszentmihalyi, 2012). In our case, students' perceptions of *Decimal Point* likely fell somewhere between play and work, because the game was played as a required in-class activity. Thus, future research could examine whether *Decimal Point*'s open learner model might lead to greater enjoyment in a different context, e.g., leisurely play, that students would perceive as more entertaining. Additionally, comparing students' behaviors, as well as learning and enjoyment outcomes, in *Decimal Point* to a non-game system (i.e., a computer tutor with the same math content, but presented in a standard, non-game manner) with the same open learner model would provide valuable insight into how an open learner model might function differently in a game versus non-game setting.

Moving forward, one potential direction for improving the learner model's effectiveness is to make its recommendations more pronounced. In particular, we could display an explanation of why certain mini-games were recommended (i.e., because the student needed more practice on them) to better convey the model's rationale and allow students to make more informed decisions. This type of explanation has been shown to increase user acceptance across many recommender systems (Adomavicius & Tuzhilin, 2005; Herlocker et al., 2000; Papadimitriou et al., 2012; Tintarev & Masthoff, 2011). In addition, adding more measures of achievement orientation (Elliot & Murayama, 2008) and self-regulated learning measures (Usher & Pajares, 2008) could help clarify whether this kind of open learner model has a positive impact on students' mastery orientations and beliefs of self-directed learning. Furthermore, the current learner dashboard design is a suitable platform for implementing the mastery-oriented features from the algebra tutor *Lynnette* (Long & Aleven, 2016), which provide explicit feedback on whether a student's problem selection choice was good or bad. As Long and Aleven have reported that these features led to more learning than a tutor version with full system control over problem sequencing, it would also be meaningful to replicate this effect in future studies of *Decimal Point*.

Student Interaction with the Open Enjoyment Model

In the Enjoyment-oriented condition, the open enjoyment model asked students to report their enjoyment rating for each mini-game round, then averaged these ratings by game types and placed the types with the highest average scores on top. We based this design decision on the assumption that playing more of the game types that students liked the most would increase their engagement (Harpstead et al., 2015; Van der Heijden, 2004). Furthermore, encouraging students to make decisions based on their own enjoyment rather than external rewards can also be seen as a form of intrinsic motivation support (Ryan & Deci, 2000). However, we found that following the open enjoyment model did not lead to higher post-game enjoyment or correlate with posttest and delayed posttest performance. This outcome was likely due to the mini-game ratings, which the enjoyment model relied on for ranking the game types, being heavily skewed towards the maximum rating of 5 stars across all mini-games. It may also be that explicitly drawing attention to enjoyment through ratings and visualizations disrupted the game environment and decreased enjoyment, thereby negating any enjoyment benefits students might have experienced by playing their preferred games.

This biased rating trend could be due to either the game interface or the classroom environment. While the fun-o-meter (Fig. 3) did provide some rating guidelines (i.e., 1 star means "not fun" and 5 stars mean "crazy fun"), these suggestions were likely insufficient in helping students discern their level of enjoyment. Furthermore, prior research has indicated a positive bias tendency in the five-star rating systems (Hu et al., 2009; Zervas et al., 2021). Providing a social contribution context for student reviews, such as telling students that their ratings will be used by other learners to make game play decisions, could motivate students to provide more varied ratings. In addition, the classroom setting where *Decimal Point* was deployed may cause students to view the game as an alternative to typical classwork. In this case, the mini-games were likely more engaging than textbook exercises and therefore earned their high ratings. To get a better sense of whether *Decimal Point* was inherently enjoyable, we should examine students' ratings as they played in their free time, when their expectation for game enjoyment would likely be higher than in the classroom.

Prior studies on the use of fun-o-meters have found that, while children tended to provide the maximum rating in most cases, there is a trend of older children being more discriminating in their ratings (Read et al., 2002; Read & MacFarlane, 2006). When comparing the age range and response variability of the participants in these studies and ours, we noted that our results are consistent with the reported trend. However, the students' ratings are still not diverse enough to serve as useful input to an adaptive recommender system, such as the open enjoyment model dashboard. Therefore, to enhance the open enjoyment model, the input bias from the fun-o-meter should be reduced, either by showing students their previous ratings as a reference to support joint evaluation (Mussweiler, 2003), or by adopting more evidence-based enjoyment questionnaires (Mekler et al., 2014).

Relationship to Digital Learning Game Research

Our results open up several directions of future research in learning game design. First, consistent with Chen et al. (2007), we have shown that incorporating an open learner model in the game is beneficial to student learning. While it could be argued that displaying the learning-related metrics would diminish the game's immersiveness and, in turn, enjoyability, this effect was not observed in *Decimal Point* – the LC students reported a similar level of enjoyment as those in EC and CC. The fact that learning can be optimized, through an open learner model and recommendation system, without sacrificing enjoyment is particularly encouraging, and would merit additional validation in other learning game studies.

Another motivating question for our study is whether enjoyment can also be optimized without undermining learning. We again saw that the EC students who interacted with this model did not learn less than those in LC or CC. An important nuance here is that, while the open learner model approach is based on metacognitive theories of self-regulated learning (Bull & Kay, 2010; Nussbaumer et al., 2014) and has been validated by numerous empirical studies (Bodily et al., 2018), the concept of an open enjoyment model is quite novel. To maintain a fair comparison between study conditions, we have set up an analogous open enjoyment model that uses the students' in-game ratings as a representation of their enjoyment. However, a feasible interpretation of our results is that learning (i.e., long-term reusable conceptions) and enjoyment (i.e., momentary affective state) are very distinct constructs and should be represented differently, rather than analogously. In its current state, the open enjoyment model's main functionality was to guide students towards the game types they presumably enjoyed the most, rather than to expose students to their own enjoyment state. In other words, the exact dimensions of how students interpret open models of their own enjoyment, and how such a model should be designed, remains a rich area for future work.

It is also possible that the learning-oriented or enjoyment-oriented recommender system should also take into account individual student characteristics that may lead to different play styles. For instance, there is a rich literature on player type models that are used in personalizing games towards individual player preferences (e.g., Bateman et al., 2011; Busch et al., 2015; Hamari & Tuunanen, 2014). *Decimal Point*, however, offers a fairly structured game experience, where students select from five game types to play, and the mini-game contents are identical across conditions. Prior work has shown that mini-game selection sequences did not impact learning or enjoyment (Harpstead et al., 2019; Wang et al., 2019), and therefore player types are likely not manifesting in the mini-game selection mechanic. Instead, we can investigate traits that may influence players' *overall* reception of the game, such as desired challenge level and aesthetic orientation (Tondello & Nacke, 2019). Identifying these traits and incorporating them in the game's recommendations is a promising next step.

Finally, our work also has implications for the design of COTS (commercial off the shelf) games. Many COTS games have systems that drive players toward challenging, rather than entertaining, tasks. For example, the game *Cut the Rope* (Cut the Rope, 2010) has three stars located in each game level. To finish the level, the player only needs to collect one star. However, the level selection screen shows whether the player has collected one, two, or three stars on each level. The player is incentivized to go back and replay the most challenging levels (e.g. where they have not yet figured out how to obtain three stars), because their total number of stars allows them to unlock new challenges. Our findings suggest that this design may drive incidental physics learning, such as that documented by (Croxton & Kortemeyer, 2017). Incorporating a similar star system in *Decimal Point* could increase perceived replayability by encouraging students to pursue three star ratings by, for example, completing a problem without any errors. In contrast, game design patterns such as *grinding* (Zagal et al., 2013) may be detrimental to incidental learning in COTS games. Grinding asks players to repeatedly complete a task, in order to accumulate rewards over time. For example, in *World of Warcraft*, a player might be tasked with collecting twenty copies of the same object, or killing a certain type of enemy ten times. While grind-based designs do not explicitly drive players toward their favorite tasks, players work together to find the easiest and most pleasurable way to complete them (Steinkuehler & Duncan, 2008). In other words, our work provides an analytic framework for subtask selection in COTS games that can be empirically evaluated in future studies.

Relationship to AI in Education Research

While the primary AI component of *Decimal Point* is the standard BKT algorithm that supports the open learner model's recommendations, there are several future opportunities to incorporate more advanced features from state-of-the-art learning analytics dashboards and recommender systems (Bodily & Verbert, 2017). In the Learning-oriented condition, we could construct more personalized learner models by applying advanced BKT algorithms that take into account differences in both skill types and individual learners (Eagle et al., 2016; Yudelson et al., 2013). The student-facing dashboards could also be enhanced with other students' data, allowing them to provide adaptive navigation functionality and social comparison support (Guerra et al., 2016). This kind of open social model (Brusilovsky et al., 2011) has also been shown to foster the relatedness component of self-determination theory (Deci et al., 1991), resulting in higher intrinsic motivation, which might in turn lead to greater enjoyment from challenging problems.

Similarly, in the Enjoyment-oriented condition, instead of asking students to self-report their mini-game ratings, we could build automated affect detectors that can infer a wide range of affective states within *Decimal Point*, based on past student data (Baker et al., 2012; Botelho et al., 2017; DeFalco et al., 2018; Paquette et al., 2014). In particular, we would distill meaningful features of student interaction from the log files, synchronize the features to field observations of student affect, and use data mining to determine which features of the log files are associated with field observations of each affective state. Results from these analyses would inform our understanding of how students interact with different features of the game and how their affective states change over the gameplay. These insights would in turn contribute to an affect-sensitive intelligent interface that can provide real-time recommendations to help students maximize their enjoyment and potentially learning (Bosch et al., 2015).

At the same time, we should note that the effectiveness of AI in education depends not only on the underlying AI techniques, but also on how students perceive and make use of the AI decisions made by the software (Cruz-Benito et al., 2019). The contribution of our work lies largely in the second area, where we have derived key insights on student interaction with the open learner and enjoyment models. In particular, we found that students have been receptive to the novel integration of these models in a learning game context, thus opening up the possibility that AI-enhanced open learner and enjoyment models would also be well received. Through presenting the skill mastery values and highlighting the skills that need more practice, the open learner model had the effect of prompting students to focus on their weakest skills, and following the model more frequently led to better learning. In contrast, by displaying students' mini-game ratings and encouraging more playing of the student's most favored game types, the open enjoyment model was oriented towards trying out different mini-games; however, the intended effect of promoting higher enjoyment was not present, due to the biased input ratings from young students. These are helpful lessons in human-AI interaction that are applicable to many other learning platforms, regardless of how sophisticated the underlying AI system may be.

Finally, we also acknowledge that student-AI interactions, especially in digital learning games, are inherently complex and multi-faceted. To further support the finding that following the open learner model more frequently leads to better learning, an important next step is to conduct a randomized control experiment that aims at testing this effect directly. In particular, we could adopt the experimental design of *Lynnette*, with a full player control condition (identical to the current Learning condition) and a joint player system control condition, where players can only select the mini-games from the top two game types. This design would directly manipulate whether or not players adhered to the open learner model by removing the reliance on their choosing to follow the recommendations. We will also complement our analyses on student-AI interaction with more concrete survey questions to capture students' thoughts about their interaction with adaptive dashboards on the whole.

Conclusion

In this work, we investigated the effects of a learning-oriented and enjoyment-oriented version of a digital learning game. We found that these versions yielded two distinct gameplay patterns, one focusing more on repeated practice (the Learning-oriented Condition) and the other on exploration (the Enjoyment-oriented Condition). Further analyses of student interaction with the open learner and enjoyment models demonstrated that following the open learner model, despite its relatively simple design, did help improve students' in-game learning and test performance. On the other hand, students who had access to the open enjoyment model did not report more enjoyment than those without. In turn, these results also raise important points about the human factors in AIED that should be considered when adopting intelligent technologies. Moving forward, we plan to enhance the existing models in terms of both assessment and functionality to amplify their potential impacts. We also encourage opening up student assessment models in digital learning games to better support students' decision making and to promote a deeper mutual human-system understanding.

Acknowledgments This work was supported by NSF Award #DRL-1238619. The opinions expressed are those of the authors and do not represent the views of NSF. Thanks to Jodi Forlizzi, Rosta Farzan, Michael Mogessie Ashenafi, Scott Herbst, Craig Ganoe, Darlan Santana Farias, Rick Henkel, Patrick B. McLaren, Grace Kihumba, Kim Lister, Kevin Dhou, John Choi, and Jimit Bhalani, all of whom made important contributions to the design of, development of and early experimentation with the *Decimal Point* game.

References

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personality and Social Psychology Bulletin, 38*(3), 317–330.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*(6), 734–749.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of Research on Learning and Instruction*, 522–560.

Anderman, E. M., & Dawson, H. (2011). Learning with motivation. Handbook of Research on Learning and Instruction, 219214.

Ang, D., & Mitchell, A. (2019). Representation and frequency of player choice in player-oriented dynamic difficulty adjustment systems. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 589–600.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.

Baldwin, A., Johnson, D., & Wyeth, P. (2016). Crowd-pleaser: Player perspectives of multiplayer dynamic difficulty adjustment in video games. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 326–337.

Bateman, C., Lowenhaupt, R., & Nacke, L. E. (2011). Player typology in theory and practice. *DiGRA Conference*.

Ben-Eliyahu, A., Moore, D., Dorph, R., & Schunn, C. D. (2018). Investigating the multidimensionality of engagement: Affective, behavioral, and cognitive engagement across science activities and contexts. *Contemporary Educational Psychology, 53*, 87–105.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. Proceedings of the 8th international Conference on learning analytics and knowledge, 41–50.

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies, 10*(4), 405–418.

Boggiano, A. K., Main, D. S., & Katz, P. A. (1988). Children's preference for challenge: The role of perceived competence and control. *Journal of Personality and Social Psychology, 54*(1), 134–141.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. Proceedings of the 20th international Conference on intelligent user interfaces, 379–388.

Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. International Conference on artificial intelligence in education, 40–51.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624–634.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction, 11*(1–2), 87–110.

Brusilovsky, P., Hsiao, I.-H., & Folajimi, Y. (2011). QuizMap: Open social student modeling and adaptive navigation support with TreeMaps. *European Conference on Technology Enhanced Learning*, 71–82.
Bull, S. (2020). There are open learner models about! *IEEE Transactions on Learning Technologies., 13*, 425–448.
Bull, S., Ginon, B., Boscolo, C., & Johnson, M. (2016). Introduction of learning visualisations and metacognitive support in a persuadable open learner model. Proceedings of the sixth international Conference on Learning Analytics & Knowledge, 30–39.
Bull, S., & Kay, J. (2010). Open learner models. In *Advances in intelligent tutoring systems* (pp. 301–322). Springer.
Bull, S., & Kay, J. (2008). Metacognition and open learner models. *The 3rd Workshop on Meta-Cognition and Self-Regulated Learning in Educational Technologies, at ITS2008*, 7–20.
Bunian, S., Canossa, A., Colvin, R., & El-Nasr, M. S. (2018). Modeling individual differences in game behavior using HMM. ArXiv Preprint ArXiv*:*1804*.00245*.
Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94–103.
Busch, M., Mattheiss, E., Orji, R., Marczewski, A., Hochleitner, W., Lankes, M., Nacke, L. E., & Tscheligi, M. (2015). Personalization in serious and persuasive games and gamified interactions. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 811–816.
Cajimat, R. T., Errabo, D. D. R., Cascolan, H. M. S., & Prudente, M. S. (2020). Cause analysis utilizing e-assessment on the least mastered contents of K-12 basic education curriculum. Proceedings of the 2020 11th international Conference on E-education, E-business, E-management, and E-learning, 199–203.
Carpenter, S. K. (2014). Spacing and interleaving of study and practice. *Applying the Science of Learning in Education: Infusing Psychological Science into the Curriculum*, 131–141.
Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning?
Charsky, D., & Ressler, W. (2011). "Games are made for fun": Lessons on the effects of concept maps in the classroom use of computer games. *Computers & Education, 56*(3), 604–615.
Chen, Z.-H., Chou, C.-Y., Deng, Y.-C., & Chan, T.-W. (2007). Active open learner models as animal companions: Motivating children to learn through interacting with my-pet and our-pet. *International Journal of Artificial Intelligence in Education, 17*(2), 145–167.
Chen, Z.-H., Liao, C., Chien, T.-C., & Chan, T.-W. (2011). Animal companions: Fostering children's effort-making by nurturing virtual pets. *British Journal of Educational Technology, 42*(1), 166–180.
Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.
Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.
Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278.
Croxton, D., & Kortemeyer, G. (2017). Informal physics learning from video games: A case study using gameplay videos. *Physics Education*, 53(1), 015012.
Cruz-Benito, J., Sánchez-Prieto, J. C., Therón, R., & García-Peñalvo, F. J. (2019). Measuring students' acceptance to AI-driven assessment in eLearning: Proposing a first TAM-based research model. International Conference on human-computer interaction, 15–25.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990). Harper & row New York.
Cut the Rope. (2010). ZeptoLab. https://en.wikipedia.org/wiki/Cut_the_Rope
Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., & Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. *International Educational Data Mining Society*.
Dascalu, M.-I., Bodea, C.-N., Mihailescu, M. N., Tanase, E. A., & Ordoñez de Pablos, P. (2016). Educational recommender systems and their application in lifelong learning. *Behaviour & Information Technology, 35*(4), 290–297.
Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26*(3–4), 325–346.
DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education, 28*(2), 152–193.
Deterding, S. (2016). Contextual autonomy support in video game play: A grounded theory. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3931–3943.

Dondlinger, M. J. (2007). Educational video game design: A review of the literature. *Journal of Applied Educational Technology, 4*(1), 21–31.

Eagle, M., Corbett, A., Stamper, J., McLaren, B. M., Baker, R., Wagner, A., MacLaren, B., & Mitchell, A. (2016). Predicting individual differences for learner modeling in intelligent tutors from previous learner activities. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 55–63.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education, 67*, 156–167.

Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition, 47*(6), 1088–1101.

Frommel, J., Fischbach, F., Rogers, K., & Weber, M. (2018). Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 163–171.

Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101–112.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 20–20.

Giannakos, M. N. (2013). Enjoy and learn with educational games: Examining factors affecting learning performance. *Computers & Education, 68*, 429–439.

Greipl, S., Ninaus, M., Bauer, D., Kiili, K., & Moeller, K. (2018). A fun-accuracy trade-off in game-based learning. International Conference on games and learning Alliance, 167–177.

Guerra, J., Hosseini, R., Somyurek, S., & Brusilovsky, P. (2016). An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement. Proceedings of the 21st international Conference on intelligent user interfaces, 152–163.

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

Hagelback, J., & Johansson, S. J. (2009). Measuring player experience on runtime dynamic difficulty scaling in an RTS game. 2009 IEEE Symposium on Computational Intelligence and Games, 46–52.

Hamari, J., & Tuunanen, J. (2014). Player types: A meta-synthesis.

Harpstead, E., Richey, J. E., Nguyen, H., & McLaren, B. M. (2019). Exploring the subtleties of agency and indirect control in digital learning games. Proceedings of the 9th international Conference on Learning Analytics & Knowledge, 121–129.

Harpstead, E., Zimmermann, T., Nagapan, N., Guajardo, J. J., Cooper, R., Solberg, T., & Greenawalt, D. (2015). What drives people: Creating engagement profiles of players from game log data. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 369–379.

Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy, 98*, 39–57.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, 241–250.

Hooshyar, D., Yousefi, M., & Lim, H. (2018). Data-driven approaches to game player modeling: A systematic literature review. *ACM Computing Surveys (CSUR), 50*(6), 1–19.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52*(10), 144–147.

Hummel, H. G., Van Den Berg, B., Berlanga, A. J., Drachsler, H., Janssen, J., Nadolski, R., & Koper, R. (2007). Combining social-based and information-based approaches for personalised recommendation on sequencing learning activities. *International Journal of Learning Technology, 3*(2), 152–168.

Isotani, S., McLaren, B. M., & Altman, M. (2010). Towards intelligent tutoring with erroneous examples: A taxonomy of decimal misconceptions. International Conference on intelligent tutoring systems, 346–348.

Jasin, H., Othman, M., Zain, N. M., & Osman, M. N. (2017). Proposed framework for combining Gamification elements with open learner model in a collaborative e-learning system for programming course. *Computing Research & Innovation (CRINN) Vol 2, October 2017*, 377.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. Proceedings of the 8th international Conference on learning analytics and knowledge, 31–40.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*(6), 1246–1252.

Khoshkangini, R., Valetto, G., & Marconi, A. (2017). Generating personalized challenges to enhance the persuasive power of gamification. *Personalization in Persuasive Technology Workshop*.

Kickmeier-Rust, M. D., & Albert, D. (2010). Personalized support, guidance, and feedback by embedded assessment and reasoning: What we can learn from educational computer games. *IFIP Human-Computer Interaction Symposium*, 142–151.

Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of Educational Data Mining, 43*, 43–56.

Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine, 34*(3), 27–41.

Leonardou, A., Rigou, M., & Garofalakis, J. D. (2019). Open learner models in smart learning environments. In *Cases on Smart Learning Environments* (pp. 346–368). IGI global.

Li, X., He, S., Dong, Y., Liu, Q., Liu, X., Fu, Y., Shi, Z., & Huang, W. (2010). To create DDA by the approach of ANN from UCT-created data. 2010 international Conference on computer application and system modeling (ICCASM 2010), 8, V8–475.

Lin, P., Van Brummelen, J., Lukin, G., Williams, R., & Breazeal, C. (2020). Zhorai: Designing a conversational agent for children to explore machine learning concepts. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(09), 13381–13388.

Liu, M., Horton, L., Olmanson, J., & Toprac, P. (2011). A study of learning and motivation in a new media enriched environment for middle school science. *Educational Technology Research and Development, 59*(2), 249–265.

Lomas, J. D., Koedinger, K., Patel, N., Shodhan, S., Poonwala, N., & Forlizzi, J. L. (2017). Is difficulty overrated? The effects of choice, novelty and suspense on intrinsic motivation in educational games. In Proceedings of the 2017 CHI conference on human factors in computing systems, 1028–1039.

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction, 27*(1), 55–88.

Long, Y., & Aleven, V. (2016). Mastery-oriented shared student/system control over problem selection in a linear equation tutor. International Conference on intelligent tutoring systems, 90–100.

Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. International Conference on artificial intelligence in education, 219–228.

Maass, J. K., Pavlik, P. I., & Hua, H. (2015). How spacing and variable retrieval practice affect the learning of statistics concepts. International Conference on artificial intelligence in education, 247–256.

Malacria, S., Scarr, J., Cockburn, A., Gutwin, C., & Grossman, T. (2013). Skillometers: Reflective widgets that motivate and help users to improve performance. *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 321–330.

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017). A computer-based game that promotes mathematics learning more than a conventional approach. *International Journal of Game-Based Learning (IJGBL), 7*(1), 36–56.

Mekler, E. D., Bopp, J. A., Tuch, A. N., & Opwis, K. (2014). A systematic review of quantitative studies on the enjoyment of digital entertainment games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 927–936.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472–489.

Nguyen, H., Harpstead, E., Wang, Y., & McLaren, B. M. (2018). Student agency and game-based learning: A study comparing low and high agency. International Conference on artificial intelligence in education, 338–351.

Nussbaumer, A., Kravcik, M., Renzel, D., Klamma, R., Berthold, M., & Albert, D. (2014). A framework for facilitating self-regulation in responsive open learning environments. ArXiv Preprint ArXiv*:*1407.5891.

Osman, K., & Bakar, N. A. (2012). Educational computer games for Malaysian classrooms: Issues and challenges. *Asian Social Science, 8*(11), 75.

Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2012). A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery, 24*(3), 555–583.

Papamitsiou, Z., Economides, A. A., Pappas, I. O., & Giannakos, M. N. (2018). Explaining learning performance using response-time, self-regulation and satisfaction from content: An fsQCA approach. Proceedings of the 8th international Conference on learning analytics and knowledge, 181–190.
Paquette, L., Baker, R. Sj., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. International Conference on intelligent tutoring systems, 1–10.
Patel, R., Liu, R., & Koedinger, K. R. (2016). When to block versus interleave practice? Evidence against teaching fraction addition before fraction Multiplication. CogSci.
Peddycord-Liu, Z., Cody, C., Kessler, S., Barnes, T., Lynch, C. F., & Rutherford, T. (2017). Using serious game analytics to inform digital curricular sequencing: What math objective should students play next? *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 195–204.
Pekrun, R. (2005). Progress and open problems in educational emotion research. *Learning and Instruction, 15*(5), 497–506.
Pittman, C. (2013). Teaching with portals: The intersection of video games and physics education. *Learning Landscapes, 6*(2), 341–360.
Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology, 105*(4), 1050–1066.
Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93–115.
Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154–166.
Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. *Proceedings of the 2006 Conference on Interaction Design and Children*, 81–88.
Read, J. C., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *Interaction Design and Children, 2*, 1–23.
Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology, 95*(2), 375–392.
Rice, J. W. (2007). New media resistance: Barriers to implementation of computer video games in the classroom. *Journal of Educational Multimedia and Hypermedia, 16*(3), 249–261.
Richey, J. E., Zhang, J., Das, R., Andres-Bray, J. M., Scruggs, R., Mogessie, M., Baker, R. S., & McLaren, B. M. (under review). Gaming and confrustion explain learning advantages for a math digital learning game.
Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*(3), 355–367.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.
Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education, 69*, 452–462.
Sarkar, A., & Cooper, S. (2018). Meet your match rating: Providing skill information and choice in player-versus-level matchmaking. Proceedings of the 13th international Conference on the foundations of digital games, 1–8.
Shute, V., Ke, F., Almond, R. G., Rahimi, S., Smith, G., & Lu, X. (2019). How to increase learning while not decreasing the fun in educational games. Learning Science: Theory, Research, and Practice, 327–357.
Squire, K. (2005). Changing the game: What happens when video games enter the classroom? Innovate: Journal of Online Education, 1(6).
Stacey, K., Helme, S., & Steinle, V. (2001). Confusions between decimals, fractions and negative numbers: A consequence of the mirror as a conceptual metaphor in three different ways. *PME Conference, 4*, 4–217.
Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17*(6), 530–543.
Tintarev, N., & Masthoff, J. (2011). Designing and evaluating explanations for recommender systems. In *Recommender systems handbook* (pp. 479–510). Springer.
Tobias, S., & Fletcher, J. D. (2007). What research has to say about designing computer games for learning. *Educational Technology*, 20–29.
Tondello, G. F., & Nacke, L. E. (2019). Player characteristics and video game preferences. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 365–378.
Usher, E. L., & Pajares, F. (2008). Self-efficacy for self-regulated learning: A validation study. *Educational and Psychological Measurement, 68*(3), 443–463.

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software, 3*(31), 1026.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly, 28*, 695–704.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *Educause Review, 41*(2), 16.

Vandewaetere, M., & Clarebout, G. (2014). Advanced technologies for personalized learning, instruction, and performance. In *Handbook of research on educational communications and technology* (pp. 425–437). Springer.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education, 26*(1), 107–112.

Wang, Y., Nguyen, H., Harpstead, E., Stamper, J., & McLaren, B. M. (2019). How does order of gameplay impact learning and enjoyment in a digital learning game? *International Conference on Artificial Intelligence in Education*, 518–531.

Wardrip-Fruin, N., Mateas, M., Dow, S., & Sali, S. (2009). Agency reconsidered. *DiGRA Conference*.

Wechselberger, U. (2013). Learning and enjoyment in serious gaming-contradiction or complement? DiGRA Conference, 26–29.

Whitley, B. E., & Kite, M. E. (2013). Principles of research in behavioral science. Routledge.

Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education, 140*, 103599.

Xu, M., Zhai, Y., Guo, Y., Lv, P., Li, Y., Wang, M., & Zhou, B. (2019). Personalized training through Kinect-based games for physical education. *Journal of Visual Communication and Image Representation, 62*, 394–401.

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

Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. International Conference on artificial intelligence in education, 171–180.

Zagal, J. P., Björk, S., & Lewis, C. (2013). Dark patterns in the design of games.

Zervas, G., Proserpio, D., & Byers, J. W. (2021). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters, 32*(1), 1–16.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91.

Zohaib, M. (2018). Dynamic difficulty adjustment (DDA) in computer games: A review. *Advances in Human-Computer Interaction, 2018, 1–12*.

Affiliations

Xinying Hou[1] · **Huy Anh Nguyen**[1] · **J. Elizabeth Richey**[1] · **Erik Harpstead**[1] · **Jessica Hammer**[1] · **Bruce M. McLaren**[1]

✉ Xinying Hou
xhou@cs.cmu.edu

Huy Anh Nguyen
hn1@cs.cmu.edu

J. Elizabeth Richey
jelizabethrichey@cmu.edu

Erik Harpstead
harpstead@cmu.edu

Jessica Hammer
hammerj@cs.cmu.edu

Bruce M. McLaren
bmclaren@cs.cmu.edu

[1] Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA