Exploring How Gender and Enjoyment Impact Learning in a Digital Learning Game

Xinying Hou[(✉)] , Huy A. Nguyen , J. Elizabeth Richey , and Bruce M. McLaren

Carnegie Mellon University, Pittsburgh, PA 15213, USA
xhou@cs.cmu.edu

Abstract. In digital learning games, do game mechanics that promote learning and those that promote enjoyment have different effects on students' experience? Do males or females learn from or enjoy games more? We explored these questions in *Decimal Point*, a digital learning game that teaches decimal numbers and decimal operations to middle school students. In this work, we conducted a classroom study with two versions of the game, one that encourages students to play to learn and one that encourages students to play for enjoyment. We compared these two conditions to a control condition that is neutral regarding learning and enjoyment. Our results indicated that the enjoyment-focused group learned more efficiently than the control group, and that females had higher learning gains than males across all conditions, particularly on the near and middle transfer learning items. Post hoc analyses also revealed that the learning-focused group engaged in re-practicing the same mini-games, while the enjoyment-focused group demonstrated more exploration of different mini-games. These findings suggest that emphasizing learning or enjoyment can result in distinctive gameplay behaviors from students, and that our game can help bridge the typical gender gap in math education.

Keywords: Digital learning game · Decimals · Learning · Enjoyment · Gender

1 Introduction

Digital learning games are instructional tools that can both engage students and promote learning [20]; however, students may be distracted from learning by the engaging game features [23]. To help students stay on track, modern learning games often incorporate learning-oriented mechanics such as collaborative problem-solving [50], instructional feedback [35] or open learner models [15]. More generally, several frameworks on how to design game features that optimize learning have been proposed [14, 26, 27].

On the other hand, an implicit expectation in digital learning games is that students' enjoyment can serve as a catalyst for their learning motivation and is positively correlated with learning outcomes [2, 19, 30]. Given this expectation, it would be interesting to compare the effects of an enjoyment-focused game environment with those of a learning-focused one, provided that the enjoyment-inducing features are also strongly tied to learning and not just superficial game activities. To our knowledge, only a handful of

prior studies have explicitly compared the learning and enjoyment constructs in the same game context. For example, [17] manipulated how undergraduate students perceived the same multimedia environment as either a learning module or a game, and found that the learning group demonstrated deeper learning while reporting the same level of motivation as the game group. On the other hand, the game group performed better than the learning group when instructional feedback was included in both conditions, implying that a game environment can be helpful if it promotes active learning. Another study by [52] adopted a similar strategy with high school students and found that enjoyment is not affected by playful or serious framing. However, rather than manipulating only the students' a priori perspective of the game as these prior studies did, we believe a more authentic comparison should take place during students' actual gameplay, with different game mechanics designed to emphasize either the learning or enjoyment aspect of the game.

In our work, we explored this idea in the context of *Decimal Point*, a math learning game for middle school students. Our study compared the learning and enjoyment-focused features through three conditions: one that displays the student's current level across different decimal skills and encourages more playing of the mini-games they are weakest at (Learning Condition - LC), one that displays the student's current enjoyment and encourages more playing of the mini-games they enjoy the most (Enjoyment Condition - EC), and one that does not show any learning- or enjoyment-related information (Control Condition - CC). Our research questions are as follows.

RQ1: *Is there a difference in learning outcomes among students in the three conditions?* As the LC design is essentially an open learner model [7, 10], which allows students to see their skill performance and helps regulate their learning, we hypothesized that LC students would achieve the highest learning outcome.

RQ2: *Is there a difference in self-reported enjoyment among students in the three conditions?* Given the emphasis on students playing their most enjoyed mini-games, we hypothesized that the EC group would report the highest enjoyment scores.

RQ3: *Is there a difference in learning outcomes between male and female students?* Given past research showing that females tend to learn more from digital learning games than males [28, 32], we hypothesized that female students would have better learning outcomes than males in our game across all three conditions.

RQ4: *Is there a difference in self-reported enjoyment between male and female students?* Prior research has suggested that males are typically drawn to video game features such as competition [4], achievement [41] and social interaction [22], whereas females tend to prefer engaging with familiar characters in a fantasy setting [49], which aligns more closely with our game environment. Therefore, we hypothesized that females would report higher enjoyment than males across all three conditions.

2   The Digital Learning Game *Decimal Point*

Decimal Point is a web-based single-player digital learning game that helps middle-school students learn about decimal numbers and their operations. The game features an amusement park metaphor (Fig. 1) with 8 theme areas and 24 mini-games that target common decimal misconceptions [25]. Each mini-game also exercises one of the following decimal skills:
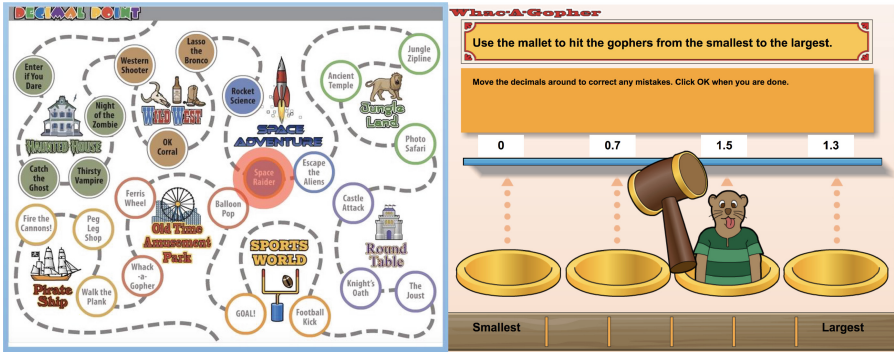
Fig. 1. The main game map where students can select among 24 mini-games to play (left), and an example mini-game in the *Sorting* skill category (right).

1. *Number Line* - locate the position of a decimal number on the number line.
2. *Addition* - add two decimals by entering the carry digits and the sum.
3. *Sequence* - fill in the next two numbers of a sequence of decimal numbers.
4. *Bucket* - compare given decimals to a threshold number and place each decimal in a "less than" or "greater than" bucket.
5. *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 [33]. Subsequent studies have integrated the element of agency into the game, by allowing students to select which mini-games to play and when to stop [24, 36]. Students who were provided agency acquired equivalent learning gains in less time than those who were not, suggesting that they could self-regulate effectively. Furthermore, a post hoc analysis by [51] reported that certain mini-game sequences which are indicative of students' exercise of agency led to higher self-reported enjoyment than others. However, no effect of agency or other game elements on test performance has been observed so far.

In addition, given an earlier finding that females benefited more from the *Decimal Point* game than males [32], we are interested in further analyses of the gender differences. As agency and learning/enjoyment-focused mechanics are integrated into the game, would the findings from [32] still hold? If a gender effect was present, which factors in the game would likely cause this effect?

3   Method

3.1   Participants and Design

196 fifth and sixth grade students in two public schools in a mid-sized U.S. city participated in our study, which was conducted during students' regular class time and lasted six days. The materials included a pretest, game play, evaluation questionnaire and posttest on the first five days, followed by a delayed posttest one week later. After the study, 35

students were removed from our analyses due to not finishing all the materials. Using the outlier criteria from a prior study in *Decimal Point* [36], we excluded two students whose gain scores from pretest to posttest were 2.5 standard deviations away from the mean. Thus, our final sample included 159 students (82 males, 77 females). Each student was randomly assigned to one of three conditions: Control (CC), Enjoyment (EC) or Learning (LC). In each condition, students could (1) select the mini-games to play in any order, and (2) choose to stop playing any time after completing at least 24 rounds. Additionally, each condition features a different dashboard attached to the main game map shown in Fig. 1. After finishing each mini-game, students were taken back to the main map, where they could see the updated dashboard and make their next mini-game selection.

Fig. 2. The dashboards shown along the game map in the Control (a), Enjoyment (b) and Learning (c) condition. The skills in the Enjoyment condition are renamed to appear more playful, e.g., *Addition* becomes *Mad Adder*.

In the CC group (30 males, 20 females), 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. 2a). After finishing the first two rounds of all 24 mini-games, students had the option to play another round of each. This condition is equivalent to the High Agency condition in [24] and [36].

In the EC group (29 males, 25 females), students played one round of mini-game per selection. After each mini-game round, students were asked to rate their enjoyment of that mini-game, on a scale from 1 ("not fun") to 5 ("crazy fun"). The dashboard showed the student's enjoyment rating of each skill, which was averaged over all the mini-game

ratings in that skill category so far (Fig. 2b). After the first three rounds, the dashboard would also recommend three mini-games to play next, chosen randomly from the two skills with the highest enjoyment; the student could follow this recommendation or make their own choice. Unlike in CC, EC students could play more rounds of a mini-game any time, without having to complete other mini-games at least once.

In the LC group (23 males, 32 females), students played one round of mini-game per selection. The dashboard showed the game's estimates of the student's mastery of each skill, from 0% to 100% (Fig. 2c), based on Bayesian Knowledge Tracing (BKT) [53]. We set the initial BKT parameters as $p(L_0) = 0.4$, $p(T) = 0.05$, $p(S) = p(G) = 0.299$ [3]. After the first three rounds, the dashboard would also recommend three mini-games to play next, chosen randomly from the two skills with the lowest mastery; the student could follow this recommendation or make their own choice. Similar to those in EC, LC students could play more rounds of a mini-game any time.

3.2 Materials

A web-based learning environment was used in this study. The materials included three tests, the game conditions outlined above, as well as questionnaire/survey items.

Pretest, Posttest, and Delayed Posttest: Each test consisted of 43 items. Most items are worth one point each, while some multi-part items are worth several points, for a total of 52 points per test. The items were designed to probe for specific decimal misconceptions, and involved either the five decimal skills targeted by the game or conceptual questions (e.g., "is a decimal number that starts with 0 smaller than 0?"). Three test forms (A, B and C) that were isomorphic and positionally counterbalanced across conditions were used. One-way ANOVAs showed no differences in terms of performance among the three versions of the test at pretest, $F(2, 156) = 0.480, p = .620$, posttest, $F(2, 156) = 1.496, p = .227$, and delayed posttest, $F(2, 156) = 1.302, p = .275$.

Questionnaires and Survey: Before and after playing the game, students were given demographic questionnaires and asked to rate several statements on a Likert scale from 1 ("strongly disagree") to 5 ("strongly agree"). These statements pertain to factors such as (1) *multidimensional engagement* (6 questions adapted from [6] with $\alpha = .775$ for the affective subscale and $\alpha = .540$ for the behavioral/cognitive engagement subscale), e.g. "I felt frustrated or annoyed," (2) *game engagement* (5 questions adapted from [8] with $\alpha = .736$), e.g., "I lost track of time," and (3) the enjoyment dimension of *achievement emotions* (6 questions adapted from [43] with $\alpha = .891$), e.g. "Reflecting on my progress in the game made me happy." In the multidimensional engagement construct, we excluded the behavioral/cognitive engagement subscale from analysis due to its low α value and only reported the results for affective engagement. After the game, students were also asked to reflect on their game play behavior, e.g. "How many mini-games did you play? Why?"

4 Results

First, a repeated-measures ANOVA showed a significant difference for all students 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 research questions. Given that gender is not a randomly assigned variable and males tend to outperform females in math performance by the end of elementary school [46], we did not expect students to be equivalent across genders at pretest. For this reason, we focused our gender analyses on gain scores [18]. In contrast, because the conditions (CC, LC and EC) were randomly assigned, we expected students to perform equally well on pretest across conditions; therefore, we used analyses of covariance (ANCOVA) to assess condition effects on posttest and delayed posttest.

Table 1. 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)

RQ1: *Is there a difference in learning outcomes among students in the three conditions?* Descriptive statistics about students' test scores in each condition are included in Table 1. From a one-way ANOVA, we observed no significant differences across conditions in pretest scores, $F(2, 156) = 1.915$, $p = .151$. With pretest scores as covariates, an ANCOVA showed no significant condition differences in posttest scores, $F(2, 155) = 0.201$, $p = .818$, or delayed posttest scores, $F(2, 155) = 0.143$, $p = .867$.

Following [34], learning efficiency was calculated for each student as the z-score of their pre-post or pre-delayed learning gains minus the z-score of total game time. As the learning efficiency data was not normally distributed, we used Kruskal-Wallis test and found a significant condition effect on learning efficiency for both posttest, H = 6.30, $p = .043$, and delayed posttest, H = 8.64, $p = .013$. Post hoc (Dunn) comparisons indicated that the EC group had significantly higher learning efficiency than CC, $p = .013$, $d = 0.28$, and delayed learning efficiency, $p = .003$, $d = 0.33$. There were no significant differences between EC and LC (pre-post: $p = .369$, pre-delayed: $p = .257$) or between CC and LC (pre-post: $p = .466$, pre-delayed: $p = .181$). In summary, there was a condition effect on learning efficiency, where EC students learned more efficiently than CC, but not on test performance, so our hypothesis that LC students would learn the most was not confirmed.

RQ2: *Is there a difference in self-reported enjoyment among students in the three conditions?* Descriptive statistics about students' enjoyment ratings by condition are included in Table 1. Based on one-way ANOVAs, 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 self-reported enjoyment, so our hypothesis that EC students would report the highest enjoyment was not confirmed.

RQ3: *Is there a difference in learning outcomes between male and female students?* Descriptive statistics about students' test scores by gender are included in Table 2. A one-way ANOVA showed no significant gender differences in pretest performance, $F(1, 157) = 0.534$, $p = .466$. A two-way ANOVA testing effects of condition and gender showed a significant main effect of gender on learning gains, $F(1, 153) = 4.351$, $p = .039$, $d = .33$, and delayed learning gains, $F(1, 153) = 4.431$, $p = .037$, $d = .35$, but no significant gender x condition interaction effect on learning gains, $F(2, 153) = 0.065$, $p = .937$, or delayed learning gains, $F(2, 153) = 0.685$, $p = .506$. Therefore, our hypothesis that females learned more than males across all conditions was confirmed. However, there were no significant gender differences in learning efficiency, $F(1, 157) = 0.259$, $p = .612$, or delayed learning efficiency, $F(1, 157) = 0.301$, $p = .584$.

RQ4: *Is there a difference in self-reported enjoyment between male and female students?* Descriptive statistics about males and females' ratings of the three enjoyment categories are included in Table 2. A two-way ANOVA testing effects of condition and gender revealed no significant main gender effect on achievement emotions, $F(1, 153) = 0.160$, $p = .690$, game engagement, $F(1, 153) = 1.689$, $p = .196$, or affective engagement, $F(1, 153) = 1.390$, $p = .240$. Similarly, there were no significant gender x condition interaction effects on these three constructs: achievement emotions, $F(2, 153) = 0.390$, $p = .678$, game engagement, $F(2, 153) = 0.345$, $p = .709$, and affective engagement, $F(2, 153) = 0.053$, $p = .948$. Thus, our hypothesis that females would enjoy the game more than males was not confirmed.

Table 2. Learning gains and self-reported enjoyment scores by gender.

Category	Males	Females
Pretest M (*SD*)	25.32 (9.74)	24.22 (9.14)
Learning gains M (*SD*)	4.34 (5.71)	6.22 (5.66)
Delayed learning gains M (*SD*)	5.80 (5.27)	7.69 (5.56)
Achievement emotion M (*SD*)	3.46 (1.04)	3.53 (0.83)
Game engagement M (*SD*)	3.01 (0.92)	3.21 (0.85)
Affective engagement M (*SD*)	3.64 (0.99)	3.46 (0.89)

Post Hoc Analyses. We conducted two follow-up analyses to better understand the condition effect on learning efficiency as well as the gender effect on learning gains. In cases where the data is not normally distributed, based on the omnibus test of normality [1], we employed the Kruskal-Wallis test [16] instead of ANOVA.

Condition Effect on Learning Efficiency. We first examined the number of mini-game rounds played in each condition. A Kruskal-Wallis test showed significant differences across conditions in number of rounds, H = 38.08, $p < .001$. Post hoc (Dunn) comparisons revealed that the CC ($M = 45.08$, $SD = 18.40$) had significantly more rounds than LC ($M = 33.20$, $SD = 9.86$), $p < .001$, $d = 0.44$, and LC had significantly more rounds than EC ($M = 26.65$, $SD = 4.59$), $p = .007$, $d = 0.33$. Furthermore, a Kruskal-Wallis test showed no significant condition differences in average game time per round, H = 2.50, $p = .286$. Therefore, EC students' higher learning efficiency than CC's could be attributed to their similar test scores but fewer mini-game rounds.

Next, we were interested in how varied the mini-games played in each condition were. For this purpose, 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 certain mini-games, while a low rate (close to 0) points to the student playing fewer rounds of more mini-games (i.e., playing a wider variety of mini-games). As CC students could not replay mini-games until after 48 rounds, their replay behaviors were necessarily different from those in LC and EC, so we focused our comparison on the LC and EC groups. 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' reflection on their gameplay behavior revealed a similar picture. Many in the EC group (25/54) and CC group (20/50) 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."

Gender Effect on Learning Gains. To see where females outperformed males in the tests, we assigned a level of learning transfer to each of the 43 test items: 20 near, 8 middle, 15 far. Following [5]'s taxonomy of transfer along the learned skill dimension [39], we classified test items as near transfer if they could be completed using identical procedures to those practiced in the game, middle transfer if they relied on practiced representations but required modification of procedures, and far transfer if they required students to understand underlying principles of practiced problems. For example, based on the sorting game in Fig. 1, a near transfer problem involves applying the same sorting procedure with new values ("Place the following list of decimals in order, smallest to largest: 0.7, 0, 1.0, 0.35"); a middle transfer problem asks students to apply representations of magnitude using a different procedure ("Which number is closest to 2.8? 2.88888, 2.91, 2.6, or 2.78"), while a far transfer problem tests abstract reasoning about decimal magnitude ("Is a longer decimal number larger than a shorter decimal number?"). Table 3 shows the results of one-way ANOVAs comparing pretest scores, learning gains and delayed learning gains between males and females at each transfer level. For the near and middle transfer items, females had lower scores than males at

Table 3. Comparison of test performance by gender at each transfer level. The mean difference (*MD*) indicates the mean value of males minus mean value of females.

Transfer level	Category	Statistical result
Near	Pretest scores	$F(1, 157) = 3.643, p = .058, MD = 1.399$
	Learning gains	$F(1, 157) = 4.541, p = .035, MD = -1.166$
	Delayed learning gains	$F(1, 157) = 4.020, p = .047, MD = -1.030$
Middle	Pretest scores	$F(1, 157) = 4.474, p = .036, MD = 0.669$
	Learning gains	$F(1, 157) = 4.695, p = .032, MD = -0.784$
	Delayed learning gains	$F(1, 157) = 2.495, p = .116, MD = -0.651$
Far	Pretest scores	$F(1, 157) = 1.569, p = .212, MD = -0.948$
	Learning gains	$F(1, 157) = 0.007, p = .932, MD = 0.047$
	Delayed learning gains	$F(1, 157) = 0.147, p = .702, MD = -0.226$

pretest but outperformed males in learning gains and delayed learning gains; however, there were no significant gender differences in performance on far transfer items.

5 Discussion and Conclusion

In this study, we investigated whether emphasizing the learning or enjoyment aspect of a digital learning game would lead to better outcomes, as well as whether males or females would benefit more from the game. We found that the Enjoyment Condition (EC) students played the least number of rounds but had higher learning efficiency than the Control Condition (CC) students. In addition, females gained more decimal knowledge than males in the posttest and delayed posttest across all conditions.

The condition effect on learning efficiency is an interesting extension to the study by [24], where students in the High Agency conditions (equivalent to our CC) learned more efficiently than those in Low Agency. In our case, since the CC and EC groups had similar test scores and average time per round, the difference in learning efficiency is due to CC students' higher number of rounds, which can be explained by their having to play two rounds per mini-game selection. Focusing on the EC and LC groups' gameplay behavior, we saw that EC students on average played 27 rounds with a replay rate of 0.15, so they chose to stop playing after trying most of the 24 mini-games once. At the same time, LC students had significantly higher number of rounds (33 on average) and replay rates (0.44 on average), indicating that the open learner model was effective in encouraging them to practice for mastery, consistent with prior literature [7, 31]. On the other hand, we found no significant differences between the LC group's learning efficiency and that of EC or CC, suggesting that replaying the mini-games past a certain point may yield diminishing returns. Our post hoc analysis of this study [37] revealed that over-practice, which could negatively impact students' learning efficiency [13], was indeed very common. Therefore, an important enhancement to the open learner model

would be to inform students when they have sufficiently practiced one skill and should move on to the next, in order to maximize their learning efficiency.

From the game play perspective, in the CC and EC settings, without the open learner model, students may not have monitored their learning progress [9, 54] and more likely wanted to explore all the mini-games that were offered. In contrast, LC students could see their skill performance and therefore were potentially more motivated to focus on mastering the skills one by one, as this is the traditional approach in school instruction [42]. These conjectures are supported by the students' reflections, which indicated that EC students liked to play all the mini-games while LC students wanted to improve their skill masteries. More generally, this finding suggests that in a game environment where students are free to choose between different types of task, showing an open learner model can encourage re-practicing the tasks one at a time (blocked practice), while not showing the model may result in students engaging in more exploration of the different tasks (interleaved practice). As the effects of these two practice modes depend on the instructional domain [11, 12, 21, 47], game designers should investigate the knowledge content of their game to see which mode is more suitable and, in turn, whether to incorporate an open learner model. In our context, the skills may be sufficiently distinct from one another and each was embedded in a unique interface, so interleaving and blocking, if present, were unlikely to yield differences in learning outcomes. A prior *Decimal Point* analysis similarly reported that students playing the mini-games in different orders still acquired the same knowledge [51].

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. One potential reason is that, while the EC and LC dashboard had similar structures, students were likely not exposed to this "open enjoyment model" before and did not use it effectively. Alternatively, students may have reported similar enjoyment levels because, despite the different dashboards, they still spent the majority of play time in the actual mini-games, which are identical across conditions. Furthermore, 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 condition was intended to induce [40, 44] or caused students to not take the enjoyment model as seriously as the learner model.

For the gender effect, we found that females outperformed males in learning gains at the near and middle transfer items, which most closely resemble those practiced in the game. In addition, females did not differ significantly from males in learning gains on far transfer problems. This outcome can be explained by the game's focus on practicing problems, which is typically beneficial for improving procedural knowledge but not necessarily for abstract knowledge or far transfer [45, 48]. Overall, our findings demonstrate that *Decimal Point* can potentially contribute to bridging the typical gender gap in math education [29, 46]. In addition, there were no gender differences in self-reported enjoyment, suggesting that *Decimal Point* was equally appealing to both genders. This is a positive outcome and likely results from the variety of mini-game themes and activities (Fig. 1), which appeal to both genders even if the game does not contain the features that we hypothesized are critical to the male students' enjoyment.

Our findings open up several avenues for future work. From the learning perspective, we could experiment with different skill mappings or model representation [7, 38] to better observe how students interact with the learner model in *Decimal Point*. From the enjoyment perspective, we plan to implement more in-game measures and survey questions to understand students' perception of game play in the classroom and to optimize enjoyment in *Decimal Point*. Finally, there is potential in further exploration of which game features are conducive to the observed gender effects, and how to extend the game's knowledge content to better support far transfer learning.

In summary, while the learning and enjoyment-focused mechanics in *Decimal Point* had similar impacts on students' outcomes, they yielded two distinct gameplay patterns, one focusing on repeated practice (the Learning Condition) and the other on exploration (the Enjoyment Condition). We also found that females improved in learning from the game more than males. These results in turn lead to the possibility of exploring the effect of emphasizing game-based learning or enjoyment in a classroom environment, as well as the game's potential in bridging the gender gap in math education.

References

1. d'Agostino, R.B.: An omnibus test of normality for moderate and large size samples. Biometrika **58**, 341–348 (1971)
2. Anderman, E.M., Dawson, H.: Learning with motivation. In: Alexander, P., Mayer, R. (eds.) Handbook of Research on Learning and Instruction, pp. 219–241. Routledge, New York (2011)
3. Baker, R.: Personal Correspondence (2019)
4. Bammel, G., Burrus-Bammel, L.L.: Leisure and Human Behaviour. Wm. C Brown Company Publishers, Dubuque (1982)
5. Barnett, S.M., Ceci, S.J.: When and where do we apply what we learn? A taxonomy for far transfer. Psychol. Bull. **128**, 612 (2002)
6. Ben-Eliyahu, A., Moore, D., Dorph, R., Schunn, C.D.: Investigating the multidimensionality of engagement: affective, behavioral, and cognitive engagement across science activities and contexts. Contemp. Educ. Psychol. **53**, 87–105 (2018)
7. Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., Verbert, K.: Open learner models and learning analytics dashboards: a systematic review. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 41–50 (2018)
8. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. J. Exp. Soc. Psychol. **45**, 624–634 (2009)
9. Bull, S., Kay, J.: Metacognition and open learner models. In: The 3rd Workshop on Meta-Cognition and Self-regulated Learning in Educational Technologies, at ITS2008, pp. 7–20 (2008)

10. Bull, S., Nghiem, T.: Helping learners to understand themselves with a learner model open to students, peers and instructors. In: Proceedings of Workshop on Individual and Group Modelling Methods that Help Learners Understand Themselves, International Conference on Intelligent Tutoring Systems. pp. 5–13. Citeseer (2002)

11. Carpenter, S.K., Mueller, F.E.: The effects of interleaving versus blocking on foreign language pronunciation learning. Mem. Cognit. **41**, 671–682 (2013)

12. Carvalho, P.F., Goldstone, R.L.: The benefits of interleaved and blocked study: different tasks benefit from different schedules of study. Psychon. Bull. Rev. **22**, 281–288 (2015)

13. Cen, H., Koedinger, K.R., Junker, B.: Is over practice necessary? Improving learning efficiency with the cognitive tutor through educational data mining. Front. Artif. Intell. Appl. **158**, 511 (2007)

14. Chen, Z.-H.: Exploring students' behaviors in a competition-driven educational game. Comput. Hum. Behav. **35**, 68–74 (2014)

15. Chen, Z.-H., Chou, C.-Y., Deng, Y.-C., Chan, T.-W.: Active open learner models as animal companions: motivating children to learn through interacting with My-Pet and Our-Pet. Int. J. Artif. Intell. Educ. **17**, 145–167 (2007)

16. Daniel, W.W.: Kruskal–Wallis one-way analysis of variance by ranks. In: Applied Nonparametric Statistics, pp. 226–234. PWS-Kent, Boston (1990)

17. Erhel, S., Jamet, E.: Digital game-based learning: impact of instructions and feedback on motivation and learning effectiveness. Comput. Educ. **67**, 156–167 (2013)

18. Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis. Wiley, Hoboken (2012)

19. Fu, F.-L., Su, R.-C., Yu, S.-C.: EGameFlow: a scale to measure learners' enjoyment of e-learning games. Comput. Educ. **52**, 101–112 (2009)

20. Gee, J.P.: What video games have to teach us about learning and literacy. Comput. Entertain. CIE. **1**, 20 (2003)

21. Goldstone, R.L.: Isolated and interrelated concepts. Mem. Cognit. **24**, 608–628 (1996)

22. Griffiths, M.D., Hunt, N.: Computer game playing in adolescence: prevalence and demographic indicators. J. Community Appl. Soc. Psychol. **5**, 189–193 (1995)

23. Harp, S.F., Mayer, R.E.: How seductive details do their damage: a theory of cognitive interest in science learning. J. Educ. Psychol. **90**, 414 (1998)

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

25. Isotani, S., McLaren, B.M., Altman, M.: Towards intelligent tutoring with erroneous examples: a taxonomy of decimal misconceptions. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6095, pp. 346–348. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13437-1_66

26. Kiili, K., Lainema, T., de Freitas, S., Arnab, S.: Flow framework for analyzing the quality of educational games. Entertain. Comput. **5**, 367–377 (2014)

27. Kiili, K., Perttula, A.: A design framework for educational exergames. In: de Freitas, S., Ott, M., Popescu, M.M., Stanescu, I. (eds.) New Pedagogical Approaches in Game Enhanced Learning: Curriculum Integration, pp. 136–158. IGI Global, Hershey (2013)

28. Klisch, Y., Miller, L.M., Wang, S., Epstein, J.: The impact of a science education game on students' learning and perception of inhalants as body pollutants. J. Sci. Educ. Technol. **21**, 295–303 (2012)

29. Leahey, E., Guo, G.: Gender differences in mathematical trajectories. Soc. Forces. **80**, 713–732 (2001)

30. Liu, M., Horton, L., Olmanson, J., Toprac, P.: A study of learning and motivation in a new media enriched environment for middle school science. Educ. Technol. Res. Dev. **59**, 249–265 (2011)

31. Malacria, S., Scarr, J., Cockburn, A., Gutwin, C., Grossman, T.: Skillometers: reflective widgets that motivate and help users to improve performance. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, pp. 321–330 (2013)
32. McLaren, B., Farzan, R., Adams, D., Mayer, R., Forlizzi, J.: Uncovering gender and problem difficulty effects in learning with an educational game. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 540–543. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_59
33. McLaren, B.M., Adams, D.M., Mayer, R.E., Forlizzi, J.: A computer-based game that promotes mathematics learning more than a conventional approach. Int. J. Game-Based Learn. IJGBL. 7, 36–56 (2017)
34. McLaren, B.M., Lim, S.-J., Koedinger, K.R.: When and how often should worked examples be given to students? New results and a summary of the current state of research. In: Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 2176–2181 (2008)
35. Moreno, R., Mayer, R.E.: Personalized messages that promote science learning in virtual environments. J. Educ. Psychol. 96, 165 (2004)
36. Nguyen, H., Harpstead, E., Wang, Y., McLaren, B.M.: Student agency and game-based learning: a study comparing low and high agency. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 338–351. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_25
37. Nguyen, H., Hou, X., Stamper, J., McLaren, B.M.: Moving beyond test scores: analyzing the effectiveness of a digital learning game through learning analytics. In: Proceedings of the 13th International Conference on Educational Data Mining (2020)
38. Nguyen, H., Wang, Y., Stamper, J., McLaren, B.M.: Using knowledge component modeling to increase domain understanding in a digital learning game. In: Proceedings of the 12th International Conference on Educational Data Mining, pp. 139–148 (2019)
39. Novick, L.R.: Representational transfer in problem solving. Psychol. Sci. 1, 128–132 (1990)
40. Osman, K., Bakar, N.A.: Educational computer games for Malaysian classrooms: issues and challenges. Asian Soc. Sci. 8, 75 (2012)
41. Paraskeva, F., Mysirlaki, S., Papagianni, A.: Multiplayer online games as educational tools: facing new challenges in learning. Comput. Educ. 54, 498–505 (2010)
42. Patel, R., Liu, R., Koedinger, K.R.: When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication. In: CogSci (2016)
43. Pekrun, R.: Progress and open problems in educational emotion research. Learn. Instr. 15, 497–506 (2005)
44. Rice, J.W.: New media resistance: barriers to implementation of computer video games in the classroom. J. Educ. Multimed. Hypermedia. 16, 249–261 (2007)
45. Richey, J.E., Nokes-Malach, T.J.: Comparing four instructional techniques for promoting robust knowledge. Educ. Psychol. Rev. 27, 181–218 (2015)
46. Robinson, J.P., Lubienski, S.T.: The development of gender achievement gaps in mathematics and reading during elementary and middle school: examining direct cognitive assessments and teacher ratings. Am. Educ. Res. J. 48, 268–302 (2011)
47. Rohrer, D., Dedrick, R.F., Burgess, K.: The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. Psychon. Bull. Rev. 21, 1323–1330 (2014)
48. Singley, M.K., Anderson, J.R.: The Transfer of Cognitive Skill. Harvard University Press, Cambridge (1989)
49. Subrahmanyam, K., Greenfield, P.M.: Computer Games for Girls: What Makes Them Play? From Barbie to Mortal Kombat: Gender and Computer Games. MIT Press, Cambridge (1998)
50. Sung, H.-Y., Hwang, G.-J.: A collaborative game-based learning approach to improving students' learning performance in science courses. Comput. Educ. 63, 43–51 (2013)

51. Wang, Y., Nguyen, H., Harpstead, E., Stamper, J., McLaren, B.M.: How does order of game-play impact learning and enjoyment in a digital learning game? In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11625, pp. 518–531. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23204-7_43
52. Wechselberger, U.: Learning and enjoyment in serious gaming-contradiction or complement? In: DiGRA Conference, pp. 26–29 (2013)
53. Yudelson, M.V., Koedinger, K.R., Gordon, G.J.: Individualized bayesian knowledge tracing models. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 171–180. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_18
54. Zimmerman, B.J.: Self-efficacy: an essential motive to learn. Contemp. Educ. Psychol. **25**, 82–91 (2000)