A Digital Learning Game for Mathematics that Leads to Better Learning Outcomes for Female Students: Further Evidence

Bruce M. McLaren[1], J. Elizabeth Richey[2], Huy A. Nguyen[1] and Michael Mogessie[1]
[1]Carnegie Mellon University, Pittsburgh, USA
[2]University of Pittsburgh, Pittsburgh, USA

Abstract. Stereotypes about men being better than women at mathematics appear to influence female students' interest and performance in mathematics. Given the potential motivational benefits of digital learning games, it is possible that games could help to reduce math anxiety, increase self-efficacy, and lead to better learning outcomes for female students. We are exploring this possibility in our work with Decimal Point, a digital learning game that scaffolds practice with decimal operations for 5th and 6th grade students. In several studies with various versions of the game, involving over 800 students across multiple years, we have consistently uncovered a learning advantage for female students with the game. In our most recent investigation of this gender effect, we decided to experiment with a central feature of the game: its use of prompted self-explanation to support student learning. Prior research has suggested that female students might benefit more from self-explanation than male students. In the new study, involving 214 middle school students, we compared three versions of self-explanation in the game – menu-based, scaffolded, and focused – each presenting students with a different type of prompted self-explanation after they solved problems in the game. We found that the focused approach led to more learning across all students than the menu-based approach, a result reported in an earlier paper. In the additional results reported in this paper, we again uncovered the gender effect – female students learned more from the game than male students, regardless of the version of self-explanation – and also found a trend in which female students made fewer self-explanation errors, suggesting they may have been more deliberate and thoughtful in their self-explanations. This self-explanation finding is a possible key to further investigation into how and why we see the gender effect in Decimal Point.

Keywords: Game-based learning, mathematics, gender effects, self-explanation

1. Introduction

Digital games have the potential to be powerful tools for learning, especially in the elementary and middle school years of education (Mayer, 2019). However, while there is evidence that both female and male students play digital games, they tend to prefer different kinds of games (Dindar, 2018; Phan et al., 2012). There is also evidence that female and male students *learn* differently from digital learning games. In particular, some studies have shown that digital learning games can be more effective for female students than male students in terms of learning (A. Khan et al., 2017; Nguyen et al., 2022; Tsai, 2017) and affective outcomes (Arroyo et al., 2013).

While conducting a study exploring the use of different types of prompted self-explanation in the context of a digital learning game, *Decimal Point* (McLaren et al., 2022), we also investigated whether the game led to gender differences, as found in prior studies with the game (McLaren, Farzan, et al. 2017; Nguyen et al., 2022). Prompted self-explanation is used to promote learning by encouraging students to self-explain what they are studying or how they have solved a problem (Wylie & Chi, 2014). Prompted self-explanation has proven to be one of the most robust and effective learning science principles over decades of research (Darling-Hammond et al., 2020). Our primary goal in this study was to further explore how prompted self-explanations might support learning with a digital game.

2. Gender Effects

Although female students have displayed very similar math achievement to male students in some meta-analyses (Lindberg et al., 2010), they have had worse outcomes among top performers and in advanced mathematics (Breda et al., 2018; Wai et al., 2010). Female students also generally report lower motivation, math self-efficacy, and interest than male students while learning mathematics (Ganley & Lubienski, 2016; Hill et al., 2016). Given that digital learning games often engage learners, they provide a unique opportunity to potentially increase confidence, motivation, and interest among female students in mathematics and thus better support their learning.

Despite this promise, research examining gender-based differences in learning games has been mixed. Some studies suggest that female students have better learning outcomes (A. Khan et al., 2017; Nguyen et al., 2022; Tsai, 2017) and enjoy and value learning games more than male students (Chung & Chang, 2017; Joiner et al., 2011). However, other research has found no gender differences in outcomes (Chang et al., 2014; Dorji et al.,

2015) or even an advantage for male students (Tawfik et al., 2009). Overall, relatively few studies have taken a rigorous approach (i.e., using a randomized, controlled design) to examine gender differences and learning outcomes with digital learning games. Even less research has been conducted to identify specific design elements of learning games that might lead to gender differences.

Differences in game preferences and gaming behaviors may provide clues about the underlying causes of gender differences in learning outcomes. For instance, while female and male students generally express a similar desire to play digital games (Hamari & Keronen, 2017), gender differences have emerged in preferences for game type, speed, and opportunities for social interaction (Aleksić & Ivanović, 2017; Romrell, 2014). In learning games research, one study showed that female students tend to rank goal clarity and social interaction as more important, while male students tend to place more importance on challenge, progress feedback, and competition (Dele-Ajayi et al., 2018). These preferences can produce meaningful differences in learning behaviors; for example, one study found that female students reported more positive feelings and increased help-seeking behaviors when a non-player “learning companion” was present, while male students did best without the learning companion (Arroyo et al., 2013).

Our exploration of gender differences has experimentally tested how different game features affect in-game behaviors and learning outcomes for female and male students (Nguyen et al., 2020; Richey et al., 2021). This will hopefully lead to insights as to why *Decimal Point* has led to gender effects favoring female students. This paper represents another step in that exploration by focusing on self-explanations, which have sometimes been hypothesized to support female students more than male students, based on young female students’ verbal learning skills and higher quality self-explanation responses (Stevenson et al., 2009).

3. Prompted Self-Explanation

Despite contentions that prompted self-explanation within a digital learning game could disrupt game flow (Killingsworth et al., 2015), induce extraneous cognitive processing (Adams & Clark, 2014; O’Neil et al, 2014), or lead learners to quickly and shallowly respond (O’Neil et al, 2014), prompted self-explanation has been shown to support learning within games. For instance, within a middle school fractions game, O’Neil and colleagues (2014) showed that menu-based self-explanations must be carefully designed and targeted to support learning. In their game, the self-explanation prompts aimed at helping learners make connections between math and game terminology were more effective for learning than no self-explanation prompts. However, not all studies have supported the idea of incorporating self-explanations into digital learning games. In a study with *Newtonian Game Dynamics*, Adams and Clark (2014) compared menu-based self-explanation with explanatory feedback and a control condition with neither self-explanation nor explanatory feedback. They found no learning differences and, in fact, students in the menu-based self-explanation condition completed fewer game levels than the condition with no self-explanation or feedback. Thus, this study shows that including prompted self-explanation in game-based learning is not always beneficial to learning.

Wylie and Chi (2014) argue that the *type* of prompted self-explanation used in digital learning environments can explain the seemingly inconsistent learning effects of self-explanation. The authors cast the various forms of prompted self-explanation that have been used in learning technology and games along a continuum from *highly constrained* to *unconstrained* self-explanations. Some learning technology studies have shown the benefits of highly constrained self-explanations, which present the learner with a small set of options to choose from (i.e., *menu-based self-explanations,* e.g., Johnson & Mayer, 2010). Highly-constrained self-explanations create the least cognitive load for learners. *Scaffolded self-explanations* create a somewhat greater cognitive load by, for instance, prompting learners to fill in the blanks in an explanation with a set of word options. Creating even more cognitive load are *focused self-explanations*, which require learners to write their own explanations but provide guidance about what to explain, such as prompting learners to explain similarities and differences. Finally, fully unconstrained self-explanations (i.e., *open-ended self-explanations*) prompt learners to create their own explanations without guidance or focus, thus presenting the highest cognitive load to learners.

While the increase in cognitive load might seem more likely to disrupt gameplay and learning, Wylie and Chi argue that less constrained prompts facilitate more active and constructive engagement, which in turn helps learners to activate and connect prior knowledge, fill in gaps in their understanding, and ultimately achieve more robust learning outcomes. It is still unclear whether highly constrained, less demanding self-explanations, which are less disruptive to game flow but could also be answered in a shallow way, are better for learning in digital games than unconstrained, more cognitively demanding self-explanations, which are more disruptive to game flow but could lead to more constructive engagement. We chose to study menu-based, scaffolded, and focused self-explanations in this study, sampling across the continuum from highly constrained to unconstrained self-explanations.

4. The Decimal Point Game

Decimal Point is a single-player, non-competitive game that scaffolds practice of decimal numbers and operations. The game is based on an amusement park metaphor (See Figure 1) and is targeted at middle-school students. The student moves to different theme areas (e.g., The Wild West, Jungle Land) playing a variety of mini-games (e.g., “Western Shooter,” "Jungle Zipline") targeted at common decimal misconceptions. Students travel through the entire amusement park playing all of the mini-games in sequence.

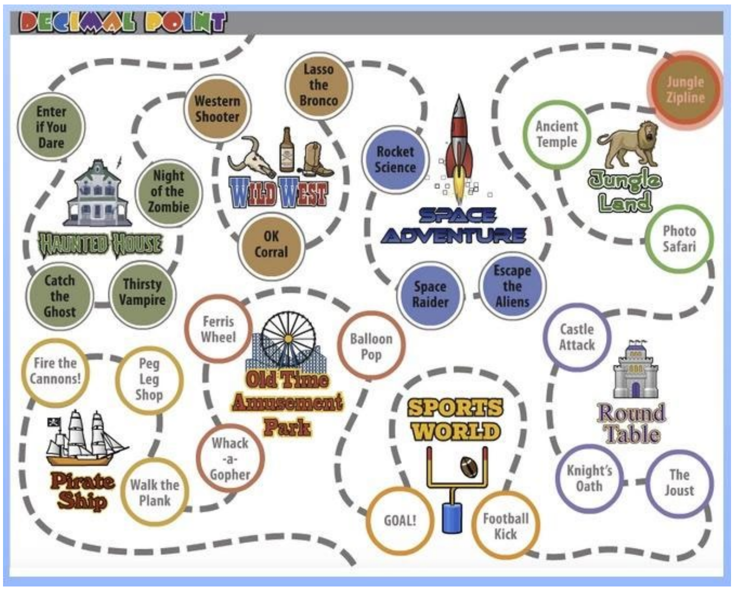


Figure 1: The *Decimal Point* Game Map

An example mini-game is “Jungle Zipline,” shown in Figure 2. This game challenges the student to zipline to trees labeled with decimal numbers (e.g., 1.0111, 1.31, 1.211, 1.1) in order from largest to smallest. When the student selects a tree, the character in the mini-game ziplines to the tree and the decimal on that tree is placed at the bottom of the screen, in left-to-right order. If the learner has made any mistakes in the decimal number order after ziplining to all of the trees, they are prompted to correct their solution by dragging and dropping the numbers at the bottom to the correct position. The student continues playing the mini-game until they correctly order all four of the decimal numbers. In Figure 2, the student is exhibiting the “longer decimals are larger” misconception by incorrectly ziplining to trees in the order of 1.0111 and 1.211.

After correctly solving two problems within each mini-game, students are prompted to answer a self-explanation question, according to one of three different self-explanation conditions illustrated in Figures 3(a), 3(b), and 3(c). The prompted self-explanation questions are intended to provide a conceptual foundation to the games' problem solving. For instance, the self-explanation question of Figure 3(a) probes the misconception that longer decimals are larger by including the incorrect distractor option, “1.0111 is bigger than 1.1 because it has more digits to the right of the decimal point.”

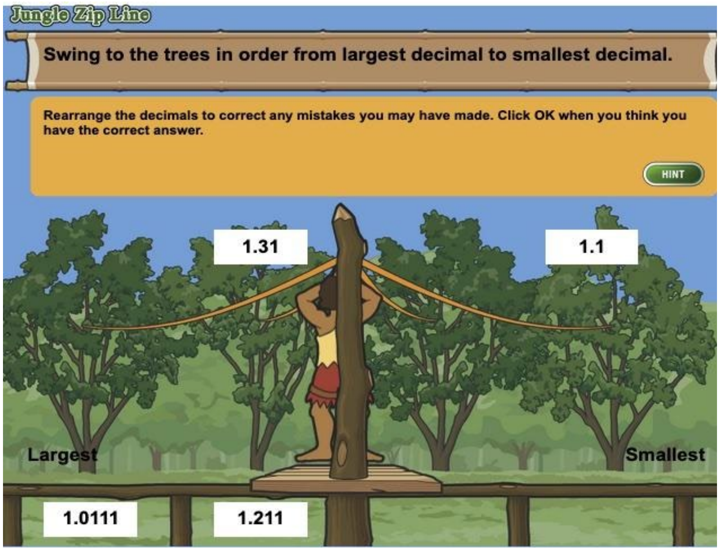


Figure 2: The "Jungle Zipline" mini-game

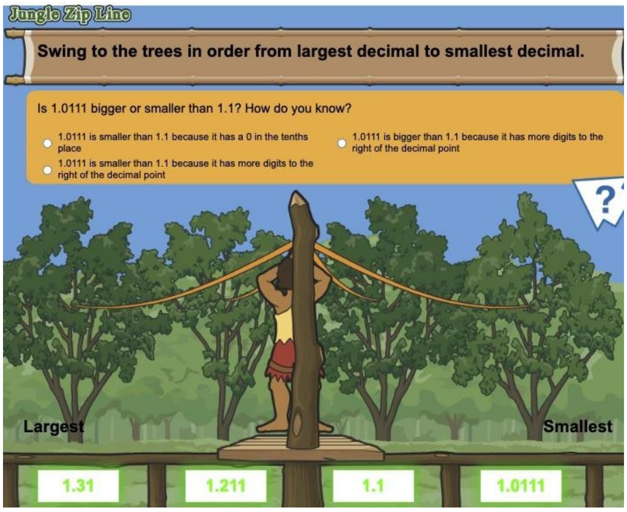
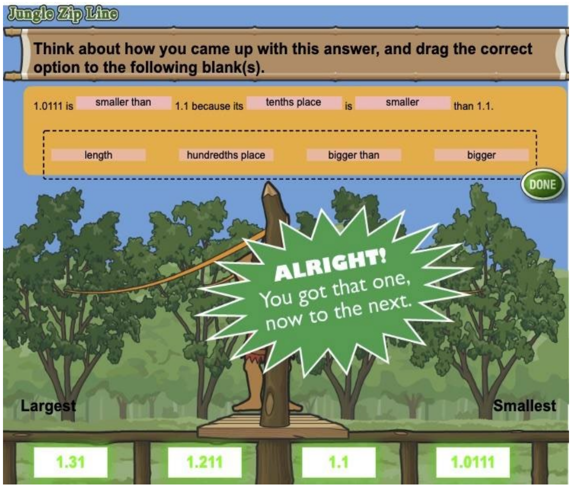


Figure 3(a): Menu-based self-explanation



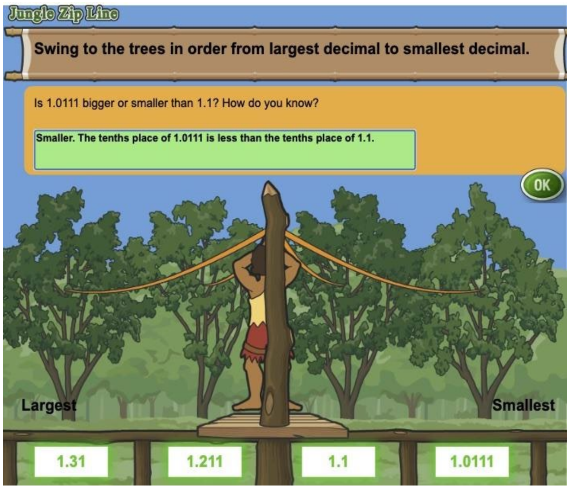


Figure 3(b): Scaffolded self-explanation

Figure 3(c): Focused self-explanation

The original version of the game employed the menu-based self-explanation approach of Figure 3(a) and led to better learning compared to a more conventional, non-game instructional technology (McLaren, Adams, et al., 2017). All studies of *Decimal Point* conducted since the initial study have employed the menu-based self-explanation approach (e.g. Hou et al., 2022; Nguyen et al., 2022), with all leading to successful learning outcomes across conditions. In this study we explored whether other forms of prompted self-explanation sampled across the Wylie and Chi (2014) continuum might also lead to successful learning and whether male and female students differed in which prompts benefitted them the most.

The three forms of self-explanation we explored were implemented as follows:

- *Menu-based self-explanations* (Figure 3(a)): Students were presented with 3 or 4 menu options providing a conceptual explanation of the correct answer to the problem. At least one incorrect option described the misconception probed by the problem. This is a *highly constrained* approach.
- *Scaffolded self-explanations* (Figure 3(b)): Students were prompted to complete a fill-in-the-blank explanation using a word bank of 4 or 5 words or phrases. This approach most closely aligns with the scaffolded approach described by Wylie and Chi (2014) and thus is a moderately constrained approach.
- *Focused self-explanation* (Figure 3(c)): Students responded to explanation prompts about specific problems and concepts by filling in an open-text field. To assure that students would expend at least minimal effort in self-explaining, we required that their self-explanations be at least four words long and contain at least one of the words from a relevant list (including common misspellings) that would legitimately be found in a correct explanation. This approach aligns with the focused self-explanations described in Wylie and Chi (2014) and is minimally constrained.

Across all conditions, students could not proceed until they provided a correct self-explanation or, in the focused condition, an explanation that met the length and relevance requirements. To control for other differences across the prompt formats, the prompts were written to be as similar as possible across all three types of self-explanation. We chose these three forms of self-explanation to sample across the full constraint continuum while making sure to include the original menu-based approach used in previous versions of the game.

5. Research Question

In this paper we address the following research question:

> *Is there a gender effect in which female students learn more than male students from the game, and does the gender effect vary based on type of self-explanation?*

This question is motivated by the consistent gender effect that has been uncovered in prior studies with *Decimal Point,* regardless of how we have altered the game intervention (McLaren, Farzan, et al. 2017; Nguyen et al., 2022). Thus, as in these earlier studies, we hypothesized for the present study that female students would learn more than male students from playing the *Decimal Point* game. Additionally, given arguments that young female students might benefit more from self-explanation, based on their more developed verbal learning strategies and more robust self-explanation responses (Stevenson et al., 2009), we hypothesized that female students' learning advantages would be more pronounced in the less constrained self-explanation conditions, which rely more on verbal reasoning.

In an earlier paper regarding the same study reported in this paper (McLaren et al., 2022), we presented and discussed the results of other research questions pertaining to the self-explanation aspect of the study. In particular, we explored which of the forms of prompted self-explanation – menu-based, scaffolded, or focused – led to the best learning and enjoyment outcomes. We wanted to examine whether any forms of prompted self-explanation would disrupt game flow, given that engagement has been shown to explain the learning benefits of *Decimal Point* compared to a non-game equivalent (Richey et al., 2021). While we focus on the gender results in this paper, we will also summarize the previously reported results.

6. Methods and Materials

Four schools, 1 rural, 2 suburban, and 1 urban, with a total of 357 5th and 6th grade students participated in the study during regular class time. One hundred and forty-three students were dropped due to (a) failing to complete part of the learning materials or any tests or (b) having participated in a study with similar materials the previous year at one of the schools. Note that the relatively high attrition rate was due, at least in part, to running the study during the COVID-19 pandemic. Some students participated in person, some at home, and some in a hybrid format. Of the remaining 214 students, 75 were in the menu-based condition (mean age: 11.6; females: 39; males: 35; 1 did not respond), 72 were in the scaffolded condition (mean age: 11.6; females: 44; males: 28), and 67 were in the focused condition (mean age: 11.6; females: 31; males: 36)

Students completed a decimal pretest, the game intervention according to condition, an immediate posttest, a questionnaire, and a delayed posttest administered one week after the posttest. The game intervention included all 24 mini-games in *Decimal Point*, with two problems and a prompted self-explanation per mini-game. All materials were deployed using an Internet-based environment.

Students worked at their own pace during 45-to-55-minute class periods for up to a full week on the pretest, game, questionnaire, and immediate posttest. Students typically took 3 to 5 days to complete the materials. Students would log out at the end of each class and were placed in the materials wherever they previously stopped when logging back in the next time. A week after the immediate posttest, students took the delayed posttest within one 45-to-55-minute class period.

Each test consisted of 43 items, with scores across these items totaling 52 points. The test items were designed to probe for decimal misconceptions and involved the decimal skills targeted by the game, as well as conceptual questions relevant to this content. Three separate test forms that were isomorphic and positionally counterbalanced across conditions were used.

The questionnaire captured responses on a Likert scale of 1-5 from "strongly disagree" to "strongly agree" and included items from the Player Experience Inventory (PXI; Abeele et al., 2020), with 3 items each from the mastery, meaning, and challenge subscales. We also included 3 items targeting affective engagement (e.g., "I felt bored") and 3 items targeting behavioral/cognitive engagement (e.g., "I tried out my ideas to see what would happen), both from Ben-Eliyahu et al.'s (2018) Multidimensional Engagement survey. Finally, we included 5 items adapted from the enjoyment subscale of Pekrun et al.'s (2011) Achievement Emotions Questionnaire (e.g., "I looked forward to playing the game").

7. Results

First, we assessed whether student performance varied based on learning location (in person, remote, or hybrid). One-way analyses of covariance (ANCOVAs) that controlled for pretest indicated no effect of location on posttest performance, $F(3, 210) = 0.12$, $p = .89$, $\eta^2_p = .001$, or delayed posttest, $F(3, 210) = 1.87$, $p = .16$, $\eta^2_p = .017$. Thus, we did not differentiate based on location in any of our analyses. Further, a one-way analysis of variance (ANOVA) showed there were no differences in pretest performance based on condition, $F(2, 211) = 1.90$, $p = .15$, $\eta^2_p = .018$, indicating that the students across all conditions had the same level of prior knowledge.

Second, we analyzed pretest, posttest, delayed posttest, errors made in problem solving, and time spent in problem solving per condition. These results are shown in Table 1.

Table 1: The mean and standard deviation of different learning and game play metrics by condition.

* - Significant differences between conditions highlighted with boldface type and asterisk

	Menu-based (*n* = 75)	Scaffolded (*n* = 72)	Focused (*n* = 67)
Pretest	26.01 (10.47)	24.63 (10.03)	27.96 (9.69)
Posttest	30.12 (9.58)	30.58 (9.88)	33.69 (9.33)
Delayed posttest (*)	30.55 (10.29)	30.97 (10.23)	34.81 (9.88)
Problem-Solving errors	75.07 (50.52)	73.22 (46.94)	65.63 (49.67)
Minutes spent in problem solving (*)	62.65 (24.68)	83.76 (27.56)	79.12 (26.09)

In McLaren et al. (2022), we reported the results of the comparison between the three self-explanation conditions. In summary, focused self-explanations led to a better learning outcome than menu-based self-explanations on the delayed posttest, with no other significant learning differences. The menu-based condition took significantly less time in problem solving than the scaffolded and focused conditions, but there was no significant difference between conditions on the number of errors during problem solving. Although not shown in Table 1, the focused condition led to significantly greater feelings of mastery than the menu-based condition, but there were no significant differences in feelings of mastery between students in the focused condition and the scaffolded condition. Finally, there was no significant effect of condition on any other questionnaire measure, indicating that students' reported experiences of engagement, enjoyment, meaning, and challenge did not differ based on condition. See the prior paper for further details and discussion of these differences between self-explanation conditions (McLaren et al., 2022).

Next, we analyzed data to answer the research question that is the focus of the current paper: whether there were gender differences in learning with *Decimal Point*. Across all conditions, female students performed marginally worse than male students on the pretest, $F(1, 211) = 2.96$, $p = .087$, $\eta^2_p = .014$. When controlling for pretest, female students performed significantly better on both the posttest, $F(1, 210) = 8.32$, $p = .004$, $\eta^2_p = .038$, and delayed posttest, $F(1, 210) = 11.25$, $p = .001$, $\eta^2_p = .051$[1]. Thus, our hypothesis that female students would benefit more than male students from playing *Decimal Point* was confirmed.

Table 2: The mean and standard deviation of test performance by gender.

* - Significant differences between female and male students highlighted with boldface type and asterisk

	Female students *M* (*SD*) N = 114	Male students *M* (*SD*) N = 99
Pretest	24.97 (10.13)	27.34 (9.93)
Posttest (*)	31.63 (9.54)	30.94 (9.77)
Delayed posttest (*)	32.52 (10.50)	31.31 (9.97)

To determine whether the different self-explanation conditions benefited female and male students differently, we also analyzed interactions between gender and self-explanation condition. A series of ANCOVAs controlling for pretest indicated no significant gender x condition interaction predicting posttest, $F(6, 206) = 0.38$, $p = .687$, $\eta^2_p = .004$, or delayed posttest, $F(6, 206) = 1.09$, $p = .34$, $\eta^2_p = .011$. Thus, our hypothesis that the less constrained self-explanation conditions of *Decimal Point* would lead to better learning for female students (i.e., Focused > Scaffolded > Menu-based) was not confirmed.

Because we suspected that how female and male students responded to the self-explanation prompts might explain differences in learning, we also looked at self-explanation errors made and time spent in self-explanation per condition (see Table 3). Note first that the number of self-explanation errors possible per condition are on different scales. In the menu-based condition, there are 3 or 4 multiple-choice options, so students can make at most 2 or 3 errors per mini-game. In the scaffolded condition, there are several blanks to fill in so students can make a much larger number of errors per mini-game. In the focused condition, the responses were coded in a post-hoc manner, so students could make at most 1 error per mini-game.

In the menu-based condition, female students had marginally fewer self-explanation errors than male students, $F(1, 72) = 3.05$, $p = .08$, $\eta^2_p = 0.04$, but female students did not spend significantly more time in minutes than male students on the self-explanation activities, $F(1, 72) = 2.04$, $p = .16$, $\eta^2_p = 0.03$. In the scaffolded condition, female students and male students did not differ significantly in their number of errors, $F(1, 72) = 0.20$, $p = .66$, $\eta^2_p < 0.01$. Female students did not spend significantly more time than male students, $F(1, 70) = 0.20$, $p = 0.65$, $\eta^2_p < 0.01$. In the focused condition, female students made significantly fewer self-explanation errors than male students, $F(1, 65) = 4.62$, $p = .04$, $\eta^2_p = 0.07$, but female and male students did not differ in time spent on the self-explanation activities, $F(1, 65) = 0.61$, $p = .44$, $\eta^2_p < .01$.

Table 3: The mean and standard deviation of time spent and errors on the self-explanation questions. * - Significant differences between female and male students highlighted with boldface type and asterisk

	Menu-based (*n* = 74)		Scaffolded (*n* = 72)		Focused (*n* = 67)	
	Female Students (n = 39)	Male Students (n = 35)	Female Students (n = 44)	Male Students (n = 28)	Female Students (n = 31)	Male Students (n = 36)
Self-Explanation Errors	11.95 (7.89)	15.03 (7.22)	41.98 (32.00)	45.18 (25.68)	**12.90 (5.23)***	**15.58 (4.96)***
Minutes spent in self-explanation	12.49 (2.94)	11.60 (2.38)	29.65 (12.38)	28.42 (9.02)	32.32 (11.13)	35.03 (16.31)

Finally, we used a series of ANOVAs to test whether male and female students differed in their reported levels of engagement or enjoyment. Results indicated that students differed by gender only in their self-reported

[1] Note that one student in the menu-based condition did not provide gender information and was thus excluded from the gender-based analyses.

behavioral/cognitive engagement, with male students (*M* = 3.36, *SD* = .80) reporting greater engagement than female students (*M* = 3.07, *SD* = .78), $F(1, 211) = 7.08$, $p = .008$, $\eta^2_p = .032$.

8. Discussion and Conclusions

The primary finding reported in this paper is that female students learned more than male students from playing the *Decimal Point* game. They had significantly higher performance on both the posttest and delayed posttest, regardless of the self-explanation version of the game they played. This gender effect is a highly robust finding, now identified in five separate studies with *Decimal Point* (Nguyen et al., 2022), including the current one. Note that this result is not due to a ceiling effect, as the mean posttest scores were in the 30 to 33 range out of a possible high score of 52.

Why do female students consistently benefit more from the *Decimal Point* game than male students? In prior work, we suggested that *Decimal Point*, which doesn't display scores or otherwise stress math achievement or competition, may have engaged female students more than male students (McLaren, Farzan, et al., 2017). Some prior work has suggested that male students may be more engaged with achievement- and competition-oriented games (Dele-Ajayi et al., 2018). However, results from the current study indicate that male students reported greater behavioral and cognitive engagement compared to female students, despite benefitting less from the game.

Another feature of *Decimal Point* that may be behind the gender effect is the focus of the current study: self-explanation. There was no interaction between self-explanation condition and gender, indicating that different versions of prompted self-explanation did not affect students differently based on gender. This contradicts some limited prior research suggesting that female students might benefit more than male students from unconstrained prompts (i.e., the focused self-explanation), given that they may tend to write more thorough and thoughtful responses (Stevenson et al., 2009). Yet, we still suspected that female and male students interacted differently with the prompted self-explanations. In particular, we hypothesized that male students might carelessly and quickly answer the self-explanation prompts more often and thus lose the potential learning benefit of considering and struggling with the explanations for their answers. We see support for this in the menu-based and focused conditions, in which female students made fewer errors than male students. The results of Nguyen et al. (2022) are consistent with these results and also suggest that careless and/or disengaged behaviors on the self-explanation step may be the key to the learning differences between female and male students.

Looking forward, there are some exciting new directions for this research. Chief among these is an investigation of how the game not only differentially interacts with the binary gender of male and female but also how it interacts with gender typicality and gender-typed occupational interests, activities, and traits (Hyde et al., 2019). Considering gender in a deeper fashion will allow us to better understand the dimensions of gender and game features that best predict learning outcomes, how they interact, and which cognitive and affective processes mediate those effects.

More broadly, we will also seek to investigate whether our findings can be distilled as design principles and transferred to other games, and whether those games produce a similar advantage for female students. We are already engaged in a project in which we will experiment with two additional learning games, *Angle Jungle* (J. Khan et al., 2017) and *Battleship Numberline* (Lomas et al., 2013), both of which are also targeted at mathematics and for roughly the same age group. If we can successfully transfer the features from *Decimal Point* that lead to better engagement and learning outcomes to these other games, we will be on a path to better understanding how equitable learning opportunities can be provided to all learners.

References

Abeele, V. V., Spiel, K., Nacke, L., Johnson, D., and Gerling, K. (2020). Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *Int'l J. of Human-Computer Studies, 135*, 102370.

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

Aleksić, V., and Ivanović, M. (2017). Early adolescent gender and multiple intelligences profiles as predictors of digital gameplay preferences. *Croatian J. of Ed.: Hrvatski časopis za odgoj i obrazovanje*, 19(3), 697-727.

Arroyo, I., Burleson, W., Tai, M., Muldner, K., and Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology*, 105(4), 957-969.

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

Breda, T., Jouini, E., and Napp, C. (2018). Societal inequalities amplify gender gaps in math. *Science*, 359(6381), 1219-1220.

Chang, M., Evans, M., Kim, S., Deater-Deckard, K., and Norton, A. (2014). Educational video games and Students' game engagement. In *2014 International Conference on Information Science & Applications* (ICISA) (pp. 1-3). IEEE.

Chung, L. Y., and Chang, R. C. (2017). The effect of gender on motivation and student achievement in digital game-based learning: A case study of a contented-based classroom. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 2309-2327.

Darling-Hammond, L., Flook, L., Cook-Harvey, C, Barron, B., and Osher, D. (2020). Implications for educational practice of the science of learning and development, *Applied Developmental Science, 24*:2, 97-140.

Dele-Ajayi, O., Strachan, R., Pickard, A., and Sanderson, J. (2018). Designing for All: Exploring Gender Diversity and Engagement with Digital Educational Games by Young People. In: *Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Dindar, M. (2018). An empirical study on gender, video game play, academic success and complex problem solving skills. *Computers and Education*, *125*, 39–52.

Dorji, U., Panjaburee, P., and Srisawasdi, N. (2015). Gender differences in students' learning achievements and awareness through residence energy saving game-based inquiry playing. *Journal of Computers in Education*, 2(2), 227-243.

Ganley, C. M., and Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences, 47*, 182-193.

Hamari, J., and Keronen, L. (2017). Why do people play games? A meta-analysis. *International Journal of Information Management*, 37(3), 125-141.

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., and Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, 48, 45-53.

Hou, X., Nguyen, H.A., Richey, J.E., Harpstead, E., Hammer, J., and McLaren, B.M. (2022). Assessing the effects of open models of learning and enjoyment in a digital learning game. *International Journal of Artificial Intelligence in Education. 32*, 120–150. https://doi.org/10.1007/s40593-021-00250-6.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., and van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171–193. https://doi.org/10.1037/amp0000307.

Johnson, C. I., and 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.

Joiner, R., Iacovides, J., Owen, M., Gavin, C., Clibbery, S., Darling, J., and Drew, B. (2011). Digital games, gender and learning in engineering: Do females benefit as much as males?. *Journal of Science Education and Technology*, 20(2), 178-185.

Khan, A., Ahmad, F. H., and Malik, M. M. (2017). Use of digital game-based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference. *Education and Information Technologies*, 22(6), 2767-2804.

Khan, J., Wang, J., Wang, X., Zhang, Y., Hammer, J., Stevens, S., and Washington, R. (2017). Angle Jungle: an educational game about angles. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (pp. 633-638).

Killingsworth, S. S., Clark, D. B., and Adams, D. (2015). Self-explanation and explanatory feedback in games: Individual differences, gameplay, and learning. *Int'l J. of Ed. in Math., Sci. and Tech., 3*, 162–186.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, 136(6), 1123.

Lomas, D., Patel, K., Forlizzi, J.L., and Koedinger, K.R. (2013). Optimizing challenge in an educational game using large-scale design experiments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 89–98. ACM.

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, *70*, 531–549.

McLaren, B.M., Adams, D.M., Mayer, R.E., and 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. doi:10.4018/IJGBL.2017010103.

McLaren, B.M., Farzan, R., Adams, D.M., Mayer, R.E., and Forlizzi, J. (2017). Uncovering gender and problem difficulty effects in learning with an educational game. In E. André, R. Baker, X. Hu, M.M.T. Rodrigo, and B. du Boulay Eds.). In: *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017)*. LNAI 10331 (pp. 540-543). Springer: Berlin.

McLaren, B.M., Nguyen, H.A., Richey, J.E., and Mogessie, M. (2022). Focused self-explanations lead to the best learning outcomes in a digital learning game. In: *Proceedings of the 16th International Conference on Learning Science (ICLS 2022).* pp. 1229-1232.

Nguyen, H., Hou, X., Richey, J.E., & McLaren, B.M. (2022). The impact of gender in learning with games: A consistent effect in a math learning game. *International Journal of Game-Based Learning (IJGBL).*

Nguyen, H.A, Hou, X., Stamper, J., and McLaren, B.M. (2020). Moving beyond test scores: Analyzing the effectiveness of a digital learning game through learning analytics. In the *Proceedings of the 13th International Conference on Educational Data Mining*, pp. 487–495.

O'Neil, H.F., Chung, G.K., Kerr, D., Vendlinski, T.P., Buschang, R.E., and Mayer, R.E. (2014). Adding self-explanation prompts to an educational computer game. *Computers in Human Behavior, 30*, 23-28.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Quest. (AEQ). *Cont. Ed. Psych., 36*(1), 36-48.

Phan, M. H., Jardina, J. R., Hoyle, S., and Chaparro, B. S. (2012). Examining the role of gender in video game usage, preference, and behavior. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1496–1500.

Richey J.E., Zhang, J., Das, R., Andres-Bray, J.M. Scruggs, R., Mogessie, M., Baker R.S.,and McLaren, B.M. (2021). Gaming and confrustion explain learning advantages for a math digital learning game. In: *Proceedings of the 22nd International Conference on Artificial Intelligence in Education (AIED 2021).*

Romrell, D. (2014). Gender and gaming: A literature review. *Annual Meeting of the AECT International Convention, Hyatt Regency Orange County, Anaheim, CA*, 170–182.

Stevenson, C. E., Resing, W. C., and Froma, M. N. (2009). Analogical reasoning skill acquisition with self-explanation in 7-8 year olds: Does feedback help? *Educational and Child Psychology*, *26*(3), 6.

Tawfik, A., He Z., and Vo, Z. (2009). Impact of video game experience and gender differences in educational video games. *Joint Conferences on Pervasive Computing (JCPC)*, Tamsui, Taipei, 2009, pp. 715-720, doi: 10.1109/JCPC.2009.5420089.

Tsai, F. H. (2017). An investigation of gender differences in a game-based learning environment with different game modes. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7), 3209-3226.

Wai, J., Cacchio, M., Putallaz, M., and Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence*, 38(4), 412-423.

Wylie, R., and Chi, M.T.H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 413–432). Cambridge Univ. Press.