Chapter 9
Decimal Point: A Decade of Learning Science Findings with a Digital Learning Game

Bruce M. McLaren

Abstract The McLearn Lab at Carnegie Mellon University (CMU) first designed and developed the artificial intelligence (AI) in education learning game, *Decimal Point*, in 2013 and 2014 to support middle school children learning decimals and decimal operations. Over a period of 10 years, the McLearn Lab has run a series of classroom experiments with the game, involving over 1,500 elementary and middle school students. In these studies, we have explored a variety of game-based learning and learning science principles and issues, such as whether the game leads to better learning—demonstrated learning gains from a pretest to a posttest and/or a delayed posttest—than a more traditional online instructional approach; whether giving students more agency leads to more learning and enjoyment; whether students benefit from hints and error messages provided during game play; and what types of prompted self-explanation lead to the best learning and enjoyment outcomes. A fascinating finding also emerged during the variety of experiments we conducted: the game consistently led to a gender effect in which girls learned more from the game than boys. In this chapter I will discuss the current state of digital learning games, how we designed and developed *Decimal Point*, the technology it is built upon—including AI techniques—and the key results of the various experiments we've conducted over the years. I conclude by discussing the important game-based learning take-aways from our studies, what we have learned about using a digital learning game as a research platform for exploring learning science principles and issues; and exciting future directions for this line of research.

Keywords Digital learning games · Research platform · Middle school mathematics · Artificial Intelligence

B. M. McLaren (✉)
Carnegie Mellon University, Pittsburgh, USA
e-mail: bmclaren@cs.cmu.edu

P. Ilic et al. (eds.), *Artificial Intelligence in Education: The Intersection of Technology and Pedagogy*, Intelligent Systems Reference Library 261,
https://doi.org/10.1007/978-3-031-71232-6_9

9.1 Introduction

Digital learning games are omnipresent and embraced by many educators and K-12 schools in the U.S. and around the world. Countless schools use digital learning games as a regular part of instruction (Juraschka, 2019). Such games span many topic areas, including math, science, language, and social science. Some examples of commonly used learning games include *Math Blaster* (https://en.wikipedia.org/wiki/Math_Blaster!), one of the oldest learning games, first publicly distributed in 1983. *Math Blaster* offers skill-building games in basic math for first- to sixth-grade students. *Legends of Learning* (https://www.legendsoflearning.com/) is a commercial organization that offers more than 2,000 math and science games for grades K-8 across more than 350 learning objectives. *Free Rice* (https://freerice.com/home) is a website with a trivia learning game, spanning a variety of content, such as English vocabulary, grammar, geography, and literature, that is designed to help kids learn and, at the same time, support donations of rice and other goods to third world and developing countries. A few of the many other learning games that are commonly seen and used in K-12 classrooms (and beyond) include *Cool Math 4 Kids* (Math), *Math Playground* (Math), *Brain Pop* (Science), *Oregon Trail* (History/Social Studies), and *Duolingo* (Language).

Taking notice of the increasing use of digital learning games during the many classroom studies my lab, the McLearn Lab, conducted with intelligent tutors in middle and high school math and science classrooms between 2006 and 2013 (Adams et al., 2012, 2013; Aleven et al., 2010; McLaren et al., 2008, 2011a, 2011b, 2014a, 2014b; Roll et al., 2011; Walker et al., 2007), I became interested in exploring the learning benefits of digital learning games. It was clear to me that teachers believe in the educational value of learning games and most of these games are highly engaging to students. However, the question I asked myself was: Do digital learning games really help students learn?

In short, the enthusiasm about and proliferation of digital learning games made me curious about their efficacy. What I found was that the evidence was very limited at the time of my initial classroom observations. There was some evidence that games can lead to more engagement and learning than conventional instructional technology, but evidence across subjects, and in particular in the domain of mathematics, was lacking (Honey & Hilton, 2011; Mayer, 2014; O'Neil & Perez, 2008; Tobias & Fletcher, 2011). In 2014, Richard E. Mayer, one of my collaborators, published a book that carefully evaluated the scientific evidence that learning games provide more learning benefits than traditional instructional approaches (Mayer, 2014—so-called *media comparison* studies). At that time there had only been five rigorous studies—those that used a controlled experiment with a comparison and measured objective learning outcomes (versus enjoyment or other subjective measures)—of digital learning games in mathematics. Of those, only three showed learning benefits for the games, with a negligible effect size of 0.03.[1]

These results—or more specifically, the lack of conclusive results—piqued my interest and led me to start a digital learning games research agenda, beginning in 2013. Given that I had already done extensive research with intelligent tutoring systems in math classrooms, it was largely a matter of switching the instructional mechanism from intelligent tutors to digital learning games for math. We focused on middle school math—and, in particular, decimals and decimal operations—as we had in prior studies with tutoring systems (Adams et al., 2012, 2013; Isotani et al., 2010; McLaren et al., 2012). My lab carefully reviewed the literature on learning games for math (Chang et al., 2012; Mayer, 2014; Van Eck & Dempsey, 2002) and game design (Schell, 2008), and began to design and develop a new math learning game. This chapter discusses that journey, the game we created, what we discovered in studies with the game, and where we, as a learning games community, are headed.

The first and primary emphasis of the chapter is an overview and discussion of the many and varied studies the McLearn Lab has conducted with *Decimal Point*, the learning game my lab designed and developed. Besides the game-based learning studies my lab has done with *Decimal Point*, our decade of experimentation has also led to some insights about using educational technology—and in particular a learning game—as a platform for exploring learning science issues and principles. That is the second emphasis of the chapter.

Because this is a long chapter, and not all readers are likely to be interested in all aspects of my lab's work with *Decimal Point*, here are some recommendations on how to read the chapter. For readers interested in learning only a bit about digital learning games, the *Decimal Point* game in particular, and the general results we've found from our many classroom studies over the past decade, I recommend reading Sect. 9.2 ("Background on Digital Learning Games"), Sect. 9.3 ("*Decimal Point*: A Digital Learning Game for Middle School Mathematics"), and Table 9.1 at the beginning of Sect. 9.4 ("Experiments with the *Decimal Point* Learning Game"), which summarizes all of the studies we've conducted. For the reader seeking a bit more depth, perspective, and understanding of the impact of the *Decimal Point* results, I suggest also reading the final three sections—5 ("Key Take-Aways: Digital Learning Game Findings"), 6 ("Key Take-Aways: Use of a Digital Learning Game as a Research Platform"), and 7 ("Conclusions")—which highlight the key take-aways about the studies and use of the game as a research platform, as well as conclusions and future directions. Finally, for readers interested in digging into details about the many results my lab has gotten with the game, I recommend reading the entire chapter, including the lengthy Sect. 9.4, which summarizes all of the classroom studies we've conducted.

[1] In the intervening years there have been many more studies of learning games in STEM subjects, and in particular mathematics, with a higher effect size in game to non-game comparisons (Hussein et al., 2022; Wang et al., 2022).

9.2 Background on Digital Learning Games

A substantial segment of the global population actively participates in digital gaming, a trend that spans various age groups. According to a report from TrueList (2023), approximately 3.26 billion people worldwide play video games with 41% of the world population estimated to be playing or have played a video game. The NPD Group, a leading market research organization, reports that video or digital gaming attracts 73% of children aged two and above (NPD, 2019). Gaming is most popular among the 18–34 age group in the United States, accounting for 36% of gamers, yet, at the same time, 24% of gamers are under the age of 18 (PlayToday, 2023), the population targeted in our work.

In summary, these findings underscore the widespread impact of digital games on recreational activities, particularly among younger people. Hence, there exists a clear rationale for leveraging games as tools to support learning, given their already widespread popularity among people in general and children in particular. Moreover, the evidence strongly suggests that digital games possess a high degree of engagement and motivation for children, which is evident in their sustained and prolonged play (Johnston, 2021). The challenge lies in seamlessly integrating instructional content into gameplay to facilitate effective learning outcomes.

In fact, there is a natural tension in digital learning games between engagement and learning. Table 9.1, taken from a Mayer and Johnson (2010) game-based learning paper, shows the pitfalls and potential of digital learning games crossed against game features and instructional features. This, to my knowledge, is one of the best depictions of the trade-offs between engagement and enjoyment, on one hand, and the educational efficacy of digital games, on the other hand.

In short, game features can be engaging and motivating but also potentially distracting, thus diminishing learning. On the other hand, instructional features, designed with a primary focus on promoting learning, can lead to student learning but can also be boring, thereby reducing motivation. Thus, the dynamic interplay between engaging game features and purposeful instructional elements presents an ongoing challenge in learning game design, demanding thoughtful design to maximize the benefits while mitigating potential drawbacks.

Table 9.1 The trade-offs between game and instructional features (Used by permission)

Potential and pitfalls of game features and instructional features in computer games for learning		
	Game features	**Instructional features**
Potential	Game features can promote motivation to learn (increasing generative processing)	Instructional features can promote learning (increasing essential and generative processing)
Pitfalls	Game features can diminish learning (increasing extraneous processing)	Instructional features can diminish motivation to learn decreasing generative processing)

Source Adapted from Mayer and Johnson (2010)

There are a variety of theories often cited for the benefits of learning games. For instance, flow theory—which posits that people can become so engaged that time passes quickly and concentration and enjoyment are deeply felt—is often cited as a reason for the benefits of games (Czikszentmihalyi, 1975, 1990; Johnston, 2021). Flow induces focused concentration and total absorption in an activity, which may in turn support better learning by enhancing engagement and persistence. Another oft-cited theorist and proponent of game-based learning, James Gee, has put forth 36 key principles of learning with games, including an "Active Learning" principle and a "Committed Learning" principle (Gee, 2003, 2007).

One of the earliest theorists of learning with games was Malone (1981), who discussed how games often trigger intrinsic motivation, employing game features such as fantasy, curiosity, and challenge. He emphasized that the immersive nature of games taps into individuals' intrinsic desires for autonomy, competence, and relatedness, aligning with Deci and Ryan's self-determination theory (Deci & Ryan, 1985; Ryan et al., 2006). Malone's insights add depth to the understanding of how games fulfill psychological needs, making them powerful tools not only for education but also for personal development. Other relevant theories are Piaget's view that play is integral to a child's cognitive development. Piaget's theory, outlined in his seminal work "Play, Dreams, and Imitation in Childhood" (1962), posits that play is not just a recreational activity but an essential component of a child's intellectual growth. Vygotsky's perspective that a child's motivation to play is related to the "Zone of Proximal Development" (ZPD—Vygotsky, 1978) further highlights the interconnectedness of play, cognitive growth and learning. These developmental theories underscore the importance of games in fostering intellectual and social skills during crucial stages of childhood. Moreover, the role of emotions in engagement and game-based learning is explored by Loderer and colleagues (2019). Their work explores the intricate connection between emotional experiences and effective learning within a gaming context, casting light on how games can be powerful tools not only for transmitting knowledge but also for shaping positive emotional responses that enhance the learning process. These diverse theories provide a comprehensive framework for appreciating the benefits of incorporating games into educational settings.

Given these learning theories about flow, motivation, and emotion as a foundation, educational technology researchers, including myself, have investigated various ways to inject the learning of traditional academic subjects, and the theory behind that learning, into digital games (see e.g., Benton et al., 2021; Cheng et al., 2017; Hooshyar et al., 2021; Lomas et al., 2013; McLaren & Nguyen, 2023; McNamara et al., 2010; Shute et al., 2019). For instance, Habgood and Ainsworth (2011) explored how to leverage intrinsic motivation (Deci, 1975) in a game context to create what has been called intrinsic integration—tightly integrating instructional content with game mechanics (Kafai, 1996). Meta-analyses of digital learning games in recent years have reported positive learning results (Clark et al., 2016; Hussein et al., 2022; Mayer, 2019; Wouters & van Oostendorp, 2017). For instance, Clark et al. (2016), in a review of 69 rigorous, empirical studies (filtered from over 1,000 studies reported in published papers), found that digital learning games were associated with a 0.33 standard deviation improvement in learning over non-game comparison conditions.

In addition, motivational and affective benefits of digital learning games have also been supported in some meta-analyses. For instance, Sitzmann (2011) found self-efficacy was higher when learning with games (average $d = 0.52$).

9.3 *Decimal Point*: A Digital Learning Game for Middle School Mathematics

The McLearn lab, along with CMU colleague Jodi Forlizzi, designed and developed the *Decimal Point* learning game, which operates using an amusement park metaphor (Forlizzi et al, 2014; McLaren et al, 2017a). We used playtesting design concepts (Walsh, 2009; Yáñez-Gómez et al., 2017) to conceptualize and design the game (Forlizzi et al., 2014). For instance, we used a co-design process in which students acted as producers, rather than consumers, in the early stages of our design work. The co-design sessions involved 32 sixth grade children over multiple sessions. Those sessions also prompted input from students on both known and established games, as well as presenting the students with preliminary game concepts we had devised for them to review. Some of the key ideas that emerged from our sessions with these children were:

- Students mentioned 54 different games, with their top choices being Minecraft, Angry Birds, Temple Run[2];
- The students particularly liked games with familiar, real-world metaphors; and
- Students liked an obstacle course concept best.

In general, the feedback we collected led us to the idea of an amusement park game (i.e., a familiar metaphor) with a series of "mini-games" (i.e., similar to obstacle courses, with multiple, different challenges).

In the *Decimal Point* game, students "travel" through a theme park playing a variety of mini-games that help them learn (and reinforce their knowledge of) decimal concepts and operations, such as place value, comparing decimal magnitude, placing decimals on a number line and adding decimals. In the base version of the game, students follow the dashed line of the amusement park map, playing mini-games in sequence, as shown in Fig. 9.1. For each mini-game, the student is prompted to play that game twice, each time with a different specific decimal problem. Across 24 mini-games, students play a total of 48 mini-game problems throughout the entire amusement park. A group of fantasy, non-player characters (NPCs) encourage students to play, congratulate them when they correctly solve problems, and provide feedback when they make mistakes (see right side of Fig. 9.1).

There are a wide variety of mini-games within *Decimal Point*, each designed to support students in learning one of following five decimal operations:

[2] Of course, this survey was conducted in 2014 when there were many different games than there are today. It would be interesting to see how this may have changed in the intervening years.

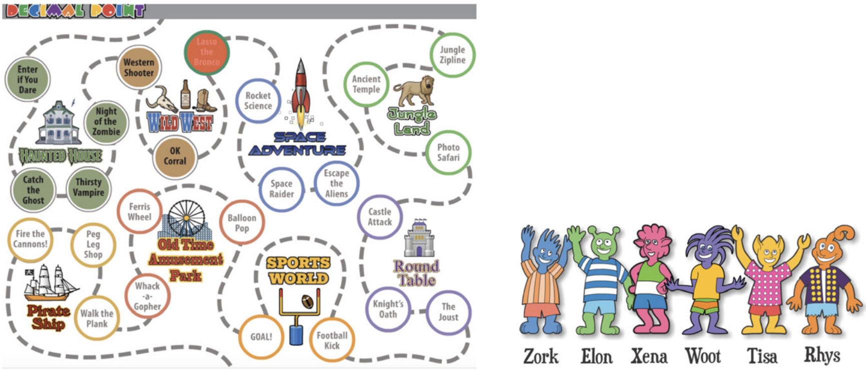


Fig. 9.1 The *Decimal Point* learning game and fantasy characters that are part of the game

- Addition of decimal numbers (*addition mini-games*—Fig. 9.2 shows an example)
- Placing decimals in less-than and greater-than "buckets" (*bucket mini-games*—Fig. 9.3)
- Placing decimals on a number line (*number line mini-games*—Fig. 9.4),
- Sorting decimals in less-than or greater-than order (*sorting mini-games*—Fig. 9.5),
- Completing sequences of decimals, given the first three numbers in a sequence (*sequence mini-games*—Fig. 9.6).

The game is embedded within a narrative designed to contextualize the math work, and the NPCs serve as guides and cheerleaders throughout the game. The narrative of *Decimal Point* provides elements of fantasy (Malone, 1981; Malone & Lepper, 1987), which is important in supporting student engagement in the early phases of interest and domain development, when students do not have enough knowledge to

Fig. 9.2 An addition mini-game

Fig. 9.3 A bucket mini-game

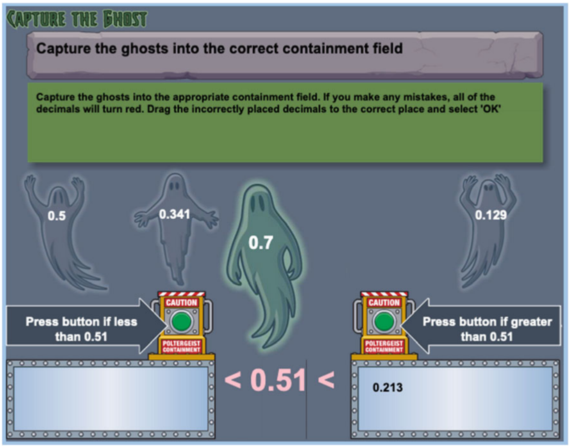
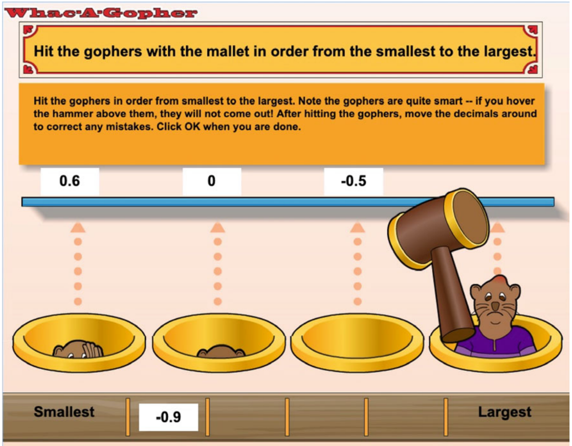


Fig. 9.4 A number line mini-game

Fig. 9.5 A sorting mini-game



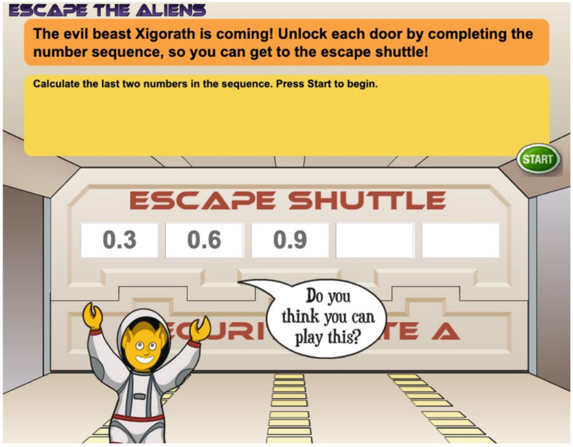


Fig. 9.6 A sequence mini-game

have developed intrinsic interest in a topic (Hidi & Renninger, 2006). This narrative also provides context for the utility of the mathematics that students are learning—and for why they are performing various problem-solving activities—a strategy that is important since it allows students to connect new knowledge to established and current knowledge (Noël et al., 2008). This connection of new to old knowledge may, in fact, be particularly important for overcoming misconceptions, which is critical for learning (Bransford et al., 2000).

All of the 24 mini-games and 48 problems in *Decimal Point* are built using learning science principles. For instance, each mini-game problem targets an established decimal misconception, such as "longer decimals are larger" (Glasgow et al, 2000; Isotani et al., 2010; Irwin, 2001; Stacey et al., 2001). This misconception occurs when kids think that, for instance, a decimal such as 0.213 is greater than 0.51 simply because the former decimal is longer (see, for instance, what the student in Fig. 9.3 has done). This misconception likely occurs because kids learn decimals after whole numbers, a domain in which the "longer is larger" heuristic works. Self-explanation, one of the most robust and widely studied learning science principles (Chi et al., 1989, 1994; Darling-Hammond et al., 2020; Rittle-Johnson & Loehr, 2017; Wylie & Chi, 2014), is also prominently used in the game. After solving the problem correctly in each of the mini-games, students are prompted to self-explain their solution. For instance, see Fig. 9.7, which shows the self-explanation step for the sorting mini-game in Fig. 9.5. The default prompted self-explanation step is a multiple-choice question, which has been argued to be minimally disruptive to gameplay while still providing the learning benefits of self-explanation (Johnson & Mayer, 2010); however, one of our studies explored other forms of prompted self-explanation, such as focused and open-ended self-explanations (McLaren et al., 2022a).

The mini-games of *Decimal Point* are essentially self-contained "intelligent tutors," built with the Cognitive Tutor Authoring Tools (CTAT), a widely used authoring tool developed by Aleven et al. (2016). The game runs on the Internet (originally implemented in Flash, and later ported to HTML/JavaScript) and takes

Fig. 9.7 Prompted self-explanation after the sorting mini-game of Fig. 9.5

advantage of tools that have been developed at Carnegie Mellon University, such as the TutorShop (Aleven et al., 2009) and DataShop (Koedinger et al., 2010). The TutorShop is how we deploy *Decimal Point* on the Internet, while the DataShop is how we capture and analyze student data. We designed and developed reusable, aggregated sets of CTAT components for *Decimal Point;* these support game mechanics that are shared across the mini-games. This approach supports shared interactions and stylistic elements across the mini-games and provides a consistency in presentation and interaction.

Artificial Intelligence techniques were used in the development of *Decimal Point* and in the studies we have run with the game. For instance, we implemented adaptive learning (Aleven et al., 2017) and Bayesian Knowledge Tracing (BKT—Corbett & Anderson, 1994) in the game for one study (Hou et al., 2020, 2022). We've also used educational data mining techniques to build detectors of cognitive, behavioral, and affective aspects of learning to analyze game play (Baker et al., 2024; Mogessie et al., 2020; Richey et al., 2021); and we recently used GPT (Ye et al., 2023) to experiment with AI-generated feedback to students (Nguyen et al., 2023b). We also recently explored using a new AI-based knowledge tracing algorithm—Deep Knowledge Tracing—and found that it performed better than BKT (Baker et al., 2023). In general, infusing AI in the context and analysis of learning games is a burgeoning and highly promising area of research (McLaren & Nguyen, 2023). I will return to this theme in the conclusions section of this chapter.

Three decimal tests were used for all studies. These tests are also designed to target common decimal misconceptions, predominantly the same ones targeted in the game, and measure near, medium, and far transfer learning. The tests have been tweaked slightly throughout the years, but generally have had between 42 and 45 items, many with multiple parts, worth a total of 52 to 61 points on each test. There are three forms of the test —A, B, and C—which are isomorphic to one another and were positionally counterbalanced in all studies, such that approximately 1/3 of the students in each condition received Test A as the pretest, 1/3 received Test B as the pretest, and 1/3 received Test C as the pretest; likewise for the posttest and delayed posttest. Some examples of test problems are: "Complete the following sequence: 0.3, 0.6, 0.9, ___, ____."; "Place 0.34 on a number line that already has 0.1, 0.3, and 0.4 on it"; "Order the following decimals, smallest to largest: 0.721, 0.3, 0.42." The test problems were largely derived from math education research and studies, with an emphasis on probing misconceptions (Brueckner, 1928; Glasgow et al., 2000; Graeber & Tirosh, 1988; Hiebert & Wearne, 1985; Irwin, 2001; Putt, 1995; Resnick et al., 1989; Sackur-Grisvard & Léonard, 1985; Stacey et al., 2001). Finally, through the various studies described in this chapter we have been able to statistically validate that the three tests are equivalent to one another.

9.4 Experiments with the *Decimal Point* Learning Game

Our overarching interest in experimenting with *Decimal Point* has always been to explore a wide and varied set of learning science-related questions in connection with game-based learning. Over the decade we have experimented with *Decimal Point*, in which a total of 1,542 students have completed our studies (See Table 9.2), we have pursued a variety of research questions, including:

- Does *Decimal Point* lead to better learning[3] and more enjoyment than a more conventional (i.e., non-game) instructional approach?
- Do female students benefit more, less, or the same as compared to male students playing the game?
- Do students learn more or less—and enjoy the game more or less—when they are given more agency in playing *Decimal Point*?
- Do students learn more or less—and enjoy the game more or less—if they are presented with a learning-focused or enjoyment-focused version of the game?
- Do students benefit from hints and error messages provided in the context of the game?
- How does the instructional context—in particular, the classroom versus remote learning—impact playing of the game and learning?
- What types of self-explanation prompts in the context of *Decimal Point* lead to the best learning and enjoyment outcomes?

[3] All mentions of 'better learning', 'more learning' or 'learning benefits' with respect to *Decimal Point* throughout this chapter means: learning gains from a pretest to a posttest and/or a delayed posttest.

- Can GPT correctly grade students' focused and open-ended self-explanations and provide correct and instructionally helpful feedback?
- Could mindfulness inductions provided in conjunction with the game enhance learning outcomes?

In effect, we have used *Decimal Point* as a *research platform* for exploring learning science questions and principles, as they relate to learning games. The game has offered us a unique opportunity to explore the nuances of learning science questions and principles. Through variations of the game and a variety of classroom studies, we have been able to probe into the underlying dynamics that define the intersection of game-based learning, education, and educational psychology.

Decimal Point has allowed us to scrutinize the effectiveness of different learning science methodologies, instructional designs, and game mechanics. Through our variations of and studies with *Decimal Point*, we have gained valuable insights into the cognitive processes, motivational factors, and emotional dimensions that contribute to the success (and failure) of learning games as educational tools. The game's customizability—which was part of the design from the start—has enabled us to test hypotheses, analyze data, and derive meaningful conclusions about the ways in which learning science principles can be applied to optimize learning with games.

The studies we have conducted over the past decade, and the key results, are summarized in Table 9.2 and discussed in the following sections.

In all of our studies we worked with a subset of the 10 public elementary and middle schools we regularly work with in a medium-size U.S. city, Pittsburgh, Pennsylvania, U.S.A. The 10 schools are distributed between urban, suburban, and rural areas. In all cases we conducted our studies in school during actual class time—except for Study 4, which was conducted during lockdown periods of the 2020–2022 worldwide pandemic, and Study 5, which was conducted in hybrid fashion due to the pandemic—over a period of approximately 6 days (5 days first week, 1 day the following week for a delayed posttest). All of the studies replaced regular instruction with our materials and online instruction. Although not mentioned in the description of every study, it is also important to note that students learned significantly from pretest to posttest and from pretest to delayed posttest in *all* conditions across *all* studies. Thus, we only report on the comparative learning benefits between conditions in each study in what follows.

Table 9.2 Studies with the *Decimal Point* learning game

Study	Description	Key research question	Key results	*N*	Relevant papers
Study 1	Comparing a learning game with a non-game digital tutor	Does *Decimal Point* lead to more learning and more enjoyment than a more conventional computer-based instructional approach (i.e., a tutor)?	*Decimal Point* led to more learning and enjoyment than the *Decimal Tutor,* a tutoring system with the same academic content. This was also the first study that uncovered a gender effect, in which girls learned more from the game (but not the tutor) than boys	153	McLaren et al., (2017a), McLaren et al., (2017b), Richey et al., (2021), Baker et al., (2023), (2024), Nguyen et al., (2022a)
Study 2	Comparing different levels of student agency in a learning game	Do students learn more or less when they are given more agency in playing *Decimal Point*?	Students did not learn more given more agency. Most students followed the canonical sequence of mini-games, rather than exercising agency	158	Nguyen et al., (2018), (2019); (2022a), Baker et al., (2023), (2024)
Study 2a (follow-up to Study 2)	Comparing students who were subject to indirect control with those not subject to indirect control in a learning game	How does the inclusion of indirect control impact students' exercise of agency in a digital learning game?	Indirect control did not lead to learning differences, but students varied in how they tried the mini-games, with some approaches leading to more enjoyment	238	Harpstead et al., (2019); Baker et al., (2023), (2024); Nguyen et al., (2019), (2022a)
Study 3	Comparing a Learning-Focused game-based learning approach to an Enjoyment-Focused approach	Do students learn more or less—and enjoy the game more or less—if they are presented with a learning-focused or enjoyment-focused version of *Decimal Point*?	Learning and enjoyment did not vary across conditions, but learning-focused students did more repeated practice, while enjoyment-focused students did more exploration	159	Hou et al., (2020), (2022), Nguyen et al., (2022a)

(continued)

Table 9.2 (continued)

Study	Description	Key research question	Key results	*N*	Relevant papers
Study 4	Comparing the use of hints and feedback in a learning game to not having hints and feedback. Also, comparing game learning in class versus at home	Do students benefit from hints and error messages provided in the context of the *Decimal Point* game? How does instructional context (i.e., classroom vs. remote) impact learning with the game?	Remote students learned more than Classroom students, but the remote drop-out rate was also very high. Surprisingly, *No-Hint* students did better in the classroom than *Hint* students on the delayed posttest	277	McLaren et al., (2022b)
Study 5	Comparing different forms of prompted self-explanation in a learning game	What form of self-explanation prompt in the context of *Decimal Point* lead to the best learning and enjoyment outcomes?	Students in the focused self-explanation condition learned more than students in the menu-based self-explanation condition at delayed posttest, with no other learning differences between conditions. Thus, it appears that focused self-explanations may be especially beneficial for retention and deeper learning	214	McLaren et al., (2022a, (2022c), Nguyen et al., (2023a), Ni et al., (2024)
Study 5a (follow-up to Study 5)	Analyzed the data from Study 5 to test whether recent advances in Large Language Models (LLMs), and in particular GPT, can support learning in a learning game	Can GPT provide instructionally meaningful feedback to incorrect student self-explanation answers in *Decimal Point*?	Results showed that GPT does very well in responding to and providing feedback for students' self-explanation errors; it also provided encouragement and flagged inappropriate language used by students, even though it was not prompted to do so. On the other hand it struggled with procedural math problems (e.g., placing points on a number line). In general, it appears teachers could gainfully use GPT, but they should stay in the loop in responding to student problems	–	Nguyen et al., (2023b)

(continued)

Table 9.2 (continued)

Study	Description	Key research question	Key results	*N*	Relevant papers
Study 6	Comparing a version of *Decimal Point* that includes prompted mindfulness, with one that prompts reading and jokes, and the base version	Can mindfulness inductions during *Decimal Point* gameplay lead to different behaviors and more learning?	Mindfulness inductions did not enhance learning or change students' gameplay behaviors. This suggests that mindfulness inductions are not beneficial for learning within digital learning games or that we may not have successfully induced mindfulness	166	Nguyen et al., (2022b), Bereczki (2024), in press; Ni et al., (2024)
Study 6a (follow-up to Study 6)	Replication of Study 6, in which we implemented a manipulation check for mindfulness to gain a better understanding of the effects of the intervention	Same basic question as Study 6, but also this: Did we manage to induce mindfulness in students in the mindfulness condition?	Once again, mindfulness inductions did not enhance learning or change students' gameplay behaviors. The only benefit detected was that students had more correct answers after listening to mindfulness reminders in the mindfulness condition as compared to listening to jokes in the story-enriched condition. The manipulation check result suggests that we did not successfully induce state mindfulness	177	Bereczki et al., (2024); Ni et al., (2024)

The N values represent the final number of students that were analyzed, i.e., those who completed all materials, in all studies. Note that in all studies some students were excluded from analyses for not having completed the materials or for having test scores too far under or over the mean

9.4.1 Study 1: Learning Game Versus Conventional Learning Technology

Our first and most fundamental study with the *Decimal Point* game was to test whether the game would lead to better engagement, enjoyment, and learning results than an equivalent, more conventional tutoring technology (McLaren et al., 2017a). As mentioned earlier, prior to this study there had only been a handful of rigorous game-based learning studies in mathematics and only three that showed learning benefits for the games, with a very small effect size of 0.03. Our research question for this first study was:

> *Study 1 RQ*: Does *Decimal Point* lead to more learning and more enjoyment than a more conventional computer-based instructional approach (i.e., a tutor)?

Study 1 was conducted during the fall of 2015 and involved students either playing and learning with the *Decimal Point* game or using a more conventional learning technology, the *Decimal Tutor.* The game and tutor share the same underlying instructional architecture and decimal content (i.e., 48 decimal problems). The decimal problems are the same across the pre-defined sequence of items in *Decimal Point* and the *Decimal Tutor,* with one example of a matching problem shown in Fig. 9.8.

One-hundred and fifty-three (153—87 female, 66 male) 6th grade students participated in and completed the initial *Decimal Point* study, from eleven classes at two middle schools. A total of two-hundred and thirteen (213) students began the study, with sixty (60) students eliminated from analyses either because they didn't finish all of the materials (52) or for having gain scores that were 2.5 standard deviations above or below the mean learning gain. Due to the potential distraction and demotivation that might have occurred with students sitting next to one another but working with very different materials, we assigned students by classroom to one of the two instructional conditions. We asked teachers to characterize classes as low, medium, and high performing and then equally distributed these across the two conditions, i.e., a quasi-random condition assignment. Seventy (70) students played and learned with

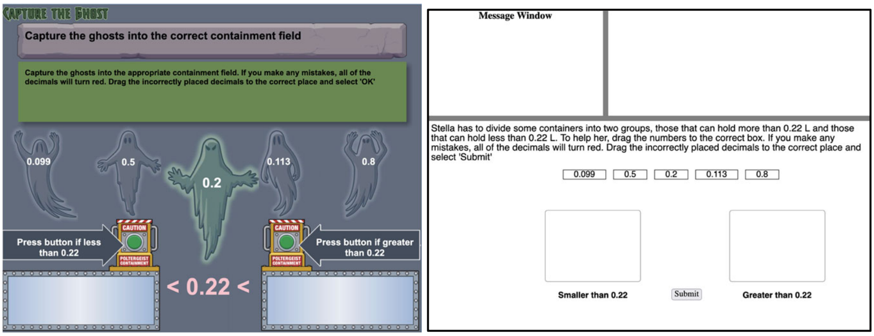


Fig. 9.8 Screenshots of the "Capture the Ghost" mini-game in *Decimal Point* (left) and the corresponding problem presented in the *Decimal Tutor* (right)

Decimal Point, while eighty-three (83) students learned the same content by using a *Decimal Tutor*. Materials included three decimal tests (pretest, posttest, delayed posttest—the tests described previously), the *Decimal Point* game as the experimental condition, the *Decimal Tutor* as the control, and two questionnaires (demographic, evaluation).

The learning results were as follows. The Game group learned more than the Tutor group, with relatively high effect sizes, on the immediate posttest ($p < 0.001$, d = 0.65 for adjusted means) and the delayed posttest ($p < 0.001$, d = 0.59 for adjusted means). In addition, the Tutor group made significantly more errors while working with the tutor (M = 273.4) than the Game group made while playing *Decimal Point* (M = 175.0). This is at least some indication that students playing the *Decimal Point* game were more engaged than those using the *Decimal Tutor* (i.e., The larger number of errors with *Decimal Tutor* likely suggests that students were guessing more frequently with the tutor). Finally, the Game group appeared to enjoy their experience more than the Tutor group, according to the evaluation questionnaire, with students expressing a significantly higher "liking" of the game than the Tutor group liked the tutor. Additional support for that finding is that the Game group expressed that the game interface was significantly easier to use than the Tutor group expressed about the Tutor. Also, the Game group expressed significantly more positive feelings about mathematics after playing than the Tutor group.

We subsequently conducted a post-hoc analysis of the data from Study 1 in which we investigated the differential impact of learning with *Decimal Point* on boys and girls (McLaren et al., 2017b). Given the established gender gap in middle school math education (Breda et al., 2018; Wai et al., 2010), where female students report higher anxiety (Huang et al., 2019; Namkung et al., 2019) and lower engagement (Else-Quest et al., 2013), we were interested in whether our game might help to address that gap. The key finding in the follow-on analysis was that female students learned more than male students from the game. This established a thread throughout all of our *Decimal Point* studies where we repeatedly investigated whether the game benefited girls or boys more. This theme is taken up and discussed in greater detail in a later section of this chapter: "The Gender Effect: A Replication Across Multiple Studies."

9.4.2 Study 2: Student Agency Versus System Control in a Learning Game

Since the beginning of our work and experiments with *Decimal Point*, we have been interested in identifying the game features that have the biggest impact on both enjoyment and learning. Agency, allowing players to make their own decisions about how to play, is one game feature that could impact both enjoyment and learning. A game might give players high agency—the ability to make many, if not all, decisions on what to do next and how to play—or low agency—where players are more restricted in what they can do, often focusing players on learning objectives they might otherwise miss if left to their own devices. Student agency is often seen as related to engagement and, consequently, learning and fun (Ryan et al., 2006). Agency is also related to self-regulated learning (SRL—Zimmerman, 2008), which depending on a student's SRL abilities, could either be helpful or harmful to learning.

Thus, the second key question we pursued in this line of research was:

> *Study 2 RQ*: Do students learn more or less—and enjoy the game more or less—when they are given more agency in playing *Decimal Point*? (Nguyen et al., 2018)

We were inspired to pursue this issue by a study of agency in the context of the *Crystal Island* learning game (Sawyer et al., 2017). *Crystal Island* is a learning game in the area of microbiology in which students try to discover the origins of an infectious disease on a remote island by interacting with key non-player characters (NPCs, e.g., a nurse, a doctor) and objects on the island. Sawyer and colleagues compared three conditions of learning*: high agency*: Students could move freely and explored throughout the island, with no guidance; *low agency*: Students investigated the infectious disease by being guided to talk to characters in a fixed order; *no agency*: Students watched a video of an expert solving the problem, essentially a worked example (Atkinson et al., 2000; Renkl, 2014; Wittwer & Renkl, 2010). They found that the low agency students attempted more incorrect submissions but at the same time learned more than the other two conditions. Interestingly, their study suggests that limiting agency can *improve* learning performance but can also lead to undesirable student behaviors, such as a propensity for guessing. Other studies have provided agency to students by allowing them to customize game features, such as icons and names in a fantasy-based arithmetic tutor (Cordova & Lepper, 1996) or customizing in-game currency, which could be spent on either personalizing the student interface or extra play in a game to improve reading comprehension skills (Snow et al., 2015). While these studies led to increased engagement and learning, student agency was essentially focused on game mechanics, and not instructional features, thus giving students a sense of control but limiting the possibility of students making poor pedagogical decisions.

Study 2 was conducted during the fall of 2017 and involved 158 5th and 6th grade students (77 female, 81 male) from two schools. Students were randomly assigned to either a high-agency (HA—79 students) or low-agency condition (LA—79 students). Thirty-nine (39) additional students were eliminated from analyses either because they didn't finish all of the materials (32) or for having gain scores that were 2.5 standard deviations above or below the mean learning gain (7). Materials included the previously discussed pretest, posttest, delayed posttest, along with two questionnaires (demographic, evaluation). The two conditions that were compared in this study—*high agency* and *low agency*—are illustrated in Fig. 9.9.

In the *high agency* condition, students were presented with a dashboard that displays the 5 different categories of mini games (e.g., Addition, Number Line), as well as the specific mini-games within each category (see the left side of the screenshot on the left of Fig. 9.9). Mini-games that had been played were marked

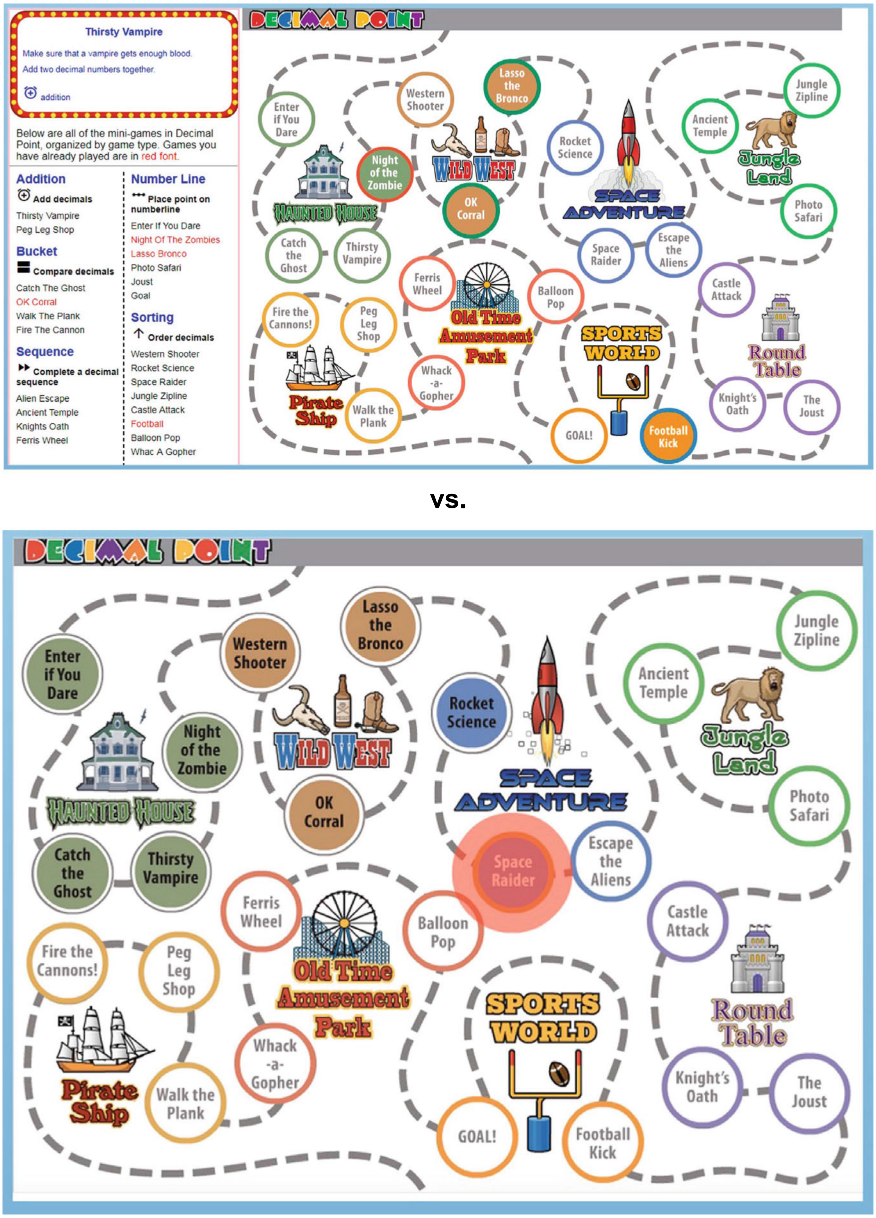


Fig. 9.9 Screenshots of the dashboard that guides the *Decimal Point* high agency condition (left) and the predefined sequence of the *Decimal Point* low agency condition, which is identical to the Game condition of Study 1 (right)

in red on the dashboard with icons filled in on the map. By mousing over the mini-game icons, students were able to learn about how each mini-game is played and what decimal skill it targets. Students could then select whatever mini-game they want to play by clicking on the corresponding mini-game icon. Students could play between 24 and 72 mini-game problems, according to their own desire. More specifically, the students could stop playing *Decimal Point* once they had finished at least 24 mini-game problems, at which point they would be presented with a dialog that contains a "Stop Playing" button. They could also play *more* than the pre-defined set of 48 mini-game problems, up to a total of 72. At any time after tackling 24 mini-game problems, the student could click on the "Stop Playing" button and thus halt game play. The *low agency* condition was the original version of *Decimal Point* from Study 1, in which students must play through all 48 mini-game problems, following the sequence shown by the dashed line on the map, starting from the upper left of the map.

There were no significant enjoyment or learning differences between the *high agency* and *low agency* groups in Study 2. In conducting post-hoc analyses, it was found that 54 of 81 *high agency* students (68%) played the same number of mini-games as the *low agency* students. Eighteen (18) of 81 *high agency* students (22%) exactly followed the canonical sequence. Also, on average, *high agency* students' sequences differed by about 10.77 edits (SD = 8.83) from the canonical sequence, a relatively minor deviation.

In short, it appears that students generally did not exercise much agency and consequently did not benefit from the high agency intervention. But why did this happen? We had several hypotheses about these results. First, students were given choices (autonomy) but may not have felt in control (agency). In particular, being in a classroom, with a teacher present, could have given many students the sense that they were not as free to make choices as we hoped. Second, perhaps the dotted line connecting all of the mini-games could have implicitly, yet unintentionally, communicated the sequence of mini-games the student should play (Schell, 2005, 2008). Finally, while we thought that students might exercise good self-regulated learning with their agency, clearly most students did not, a finding that would be predicted by some SRL research (Schunk & Zimmerman, 1998; Zimmerman, 2008). Perhaps the game environment made it even less likely that students would exercise good SRL than in other learning environments, i.e., they may have been more interested in enjoying their experience with the game than regulating their learning. The bottom line is that the hoped-for student agency and resultant benefits to enjoyment and learning did not occur. This could have been because of the teacher and/or classroom setting, the indirect control of the dotted line, or students not exercising good SRL.

9.4.2.1 Study 2a: The Impact of Indirect Control in a Learning Game

We chose to explore the second possibility, the dotted line guiding students, as the most logical next step in altering game features, as opposed to classroom or student factors. Jesse Schell has defined "indirect control" as subtle cues or design elements

Fig. 9.10 Screenshots of the three conditions of Study 2a with *Decimal Point*

that can lead players toward certain—perhaps unwanted—behaviors (Schell, 2005, 2008). Indirect control can be exerted in a variety of ways, including by presenting players with game goals, specific interface elements, characters, or visual design. Given the results of Study 2—that students generally did not exercise agency (Nguyen et al., 2018)—we wanted to more deeply explore the issue of agency and indirect control in *Decimal Point* in a follow up study (Harpstead et al, 2019). I call this Study 2a here and our question was:

> *Study 2a RQ*: Does the design of the map in *Decimal Point* exert indirect control on players by communicating an implicit order of mini-games? And does this make a difference to learning and/or enjoyment?

To explore this research question, we conducted an expanded replication of the Nguyen et al (2018) study in the spring of 2018 in which we again compared the low agency and high agency conditions, as described for Study 2, but also added a third high agency condition—what we called High Agency No Line or HANL—that operated exactly as the prior high agency condition, but with no line shown on the map (see Fig. 9.10, right). Thus, we had one condition—HAL—that exerted indirect control and another—HANL—that did not.

Two-hundred thirty-eight (238) 5th and 6th grade students from two schools (130 females, 107 males, 1 declined to respond) participated in and completed Study 2a. Forty-nine (49) other students were eliminated because: (a) they didn't finish all of the materials (35), (b) they had login errors during at least one session (13), or (c) they spent an exceptionally long time in the instructional intervention (1). Students were randomly assigned to one of three conditions: low-agency (LA—88 students), high-agency with line (HAL - 78 students) or high-agency condition with no line (HANL—72 students). Materials included the previously discussed pretest, posttest, delayed posttest, along with the two questionnaires (demographic, evaluation). The three conditions that were compared in this study—*low agency (LA), high agency with line (HAL),* and *high agency with no line (HANL)*—are illustrated in Fig. 9.10. The text is very small in this figure, but the key attributes to note in each condition is that in the LA condition students would have to follow the dotted line sequence of mini-games, in the HAL condition students could randomly choose mini-games to play, but be guided by the visible dotted line, and in the HALN condition students could randomly choose mini-games to play but without the dotted line for guidance. Note that the HAL and HALN conditions also had a dashboard like the one on the left of Fig. 9.9 that would allow students to make their own choices about game play, including choosing specific mini-games to play and playing more or less than the LA condition.

The results were as follows. There were no significant differences in learning between the three conditions. However, because students in the HAL and HANL conditions could quit early—and they largely chose to do so—they learned the same amount in significantly less game playing time, i.e., they had greater learning efficiency.[4] In particular, HAL learning efficiency > LA learning efficiency ($p = 0.012$, $d = 0.45$) and HALN learning efficiency > LA learning efficiency ($p = 0.011$, $d = 0.41$). There was no learning efficiency difference between HAL and HALN. Students in HAL and HANL played significantly fewer mini-games than in the Low Agency condition, in which they had to play all of the mini-games. There were also no differences between the three groups in enjoyment. Finally, students in HANL deviated from the canonical sequence significantly more, as measured by the length-matched Dameru-Levenshtein edit distance of a student's mini-game sequence from the canonical sequence (Bard, 2007).

The basic take-aways from this study are that while agency did not improve learning it *did* improve learning efficiency. The results further suggest that indirect control can be limited through subtle game design decisions and that students can exercise agency that ultimately leads to learning more efficiently. This suggests that the game had sufficient support in place to scaffold students' self-regulated learning (a notion discussed in Sawyer et al., 2017).

To more carefully investigate the effects of the agency we provided to students, we conducted post-hoc analyses on the combined data from Studies 2 and 2a, as reported in Wang et al., 2019. In particular, we did a cluster analysis (Bauckhage, 2015) across the 160 students who were in the HAL and HANL conditions in these two studies. We clustered students' mini-game sequences by edit distance—the number of edit operations to turn one sequence into another. We found four distinct clusters of navigation behavior in the HAL and HANL conditions. *Canonical Sequence* students (89 students) stayed very close to the prescribed order of mini-games. *Initial Exploration* students (14) initially jumped around in playing mini-games but then followed the canonical sequence. *Half-on-Top* students (100) only played half the games, in particular the ones at the top of the amusement park map. *Half-on-Left* (32) students only played half the games, in particular the ones on the left of the amusement park map. There was no difference in learning across these clusters, but a key result is about differences in enjoyment. Specifically, we observed significant differences in enjoyment between some clusters, in particular, between *Half-on-Top* and *Half-on-Left,* in which HL > HT. In general, those who deviated more from the canonical order and switched more frequently between theme areas of the Decimal Point amusement

[4] Learning Efficiency was calculated as the z-score of a student's pre-post or pre-delayed test gains minus the z-score of the total amount of time they spent playing the mini-games (McLaren et al., 2008).

park reported higher enjoyment scores. While increasing enjoyment is important, it's also important, of course, to emphasize the instruction and learning aspects of game-based learning. More investigation into the amount of instructional content needed within the game to maximize learning efficiency was clearly necessary. This prompted us to pursue our next research question.

9.4.3 Study 3: Learning Focus Versus Enjoyment Focus in a Learning Game

In our next study—Study 3, conducted in the fall of 2019—our goal was to explore the trade-off of a "learning focused" version of the game with an "enjoyment focused" version of the game. That is, we wanted to answer the question:

> *Study 3 RQ*: Do students learn more or less—and enjoy the game more or less—if they are presented with a learning-focused or enjoyment-focused version of *Decimal Poin*t? (Hou et al., 2020, 2022)

In much of game-based learning the tension between game and instructional features is palpable, as earlier depicted in Table 9.2. A successful learning game skillfully straddles the boundary between engagement and learning. A challenging aspect of digital learning game design is that features that promote engagement in learning games may also disrupt the cognitive processes that are essential for learning. For instance, one study found an inverse relationship between engagement and the difficulty of the learning task (Lomas et al., 2013). Although Lomas and colleagues found that easier learning tasks were more engaging to students in the short term, easier activities also resulted in lower learning gains and less long-term engagement. Some studies have compared enjoyment and learning constructs in the *same* game play context (Erhel & Jamet, 2013; Wechselberger, 2013). In contrast, our intent in Study 3 was to separate learning focus and enjoyment focus by comparing enjoyment and learning across *different* game play contexts. In particular, we were interested in comparing one game context that explicitly emphasized the enjoyable aspects of the game and one that explicitly emphasized the instructional aspects of the game.

To conduct this study we designed three conditions that differed based on how students were presented game information and control through a dashboard attached to the main game map. These conditions are illustrated in Fig. 9.11 and are defined as follows.

- The *Learning-Focused* condition (Fig. 9.11, left) featured an open learner model (Bodily et al., 2018; Bull, 2020), where the knowledge components (i.e., skills) displayed to students were the five decimal skills targeted in *Decimal Point* (addition, bucket, number line, sorting, sequence—Figs. 9.2 through 9.6). The bars indicate the mastery probability of each skill, which is computed by Bayesian Knowledge Tracing (Corbett & Anderson, 1994). This condition also recommended 3 specific mini-games for students to pick next, chosen randomly from

Fig. 9.11 At the top are screenshots of the three conditions of Study 3 with *Decimal Point*. On the top left is the *Learning-Focused* dashboard, in the top middle is the *Enjoyment-Focused* dashboard, and on the top right is the *Control* dashboard. At the bottom is the Fun-O-Meter dialog (Read & MacFarlane, 2006) used in the *Enjoyment-Focused* condition to rate a mini-game

those related to the top 2 skills that students needed improvement on. Our intention with this condition was that it would encourage students to focus on and practice their skills that are lacking; however, they could also choose not to follow these recommendations.

- The *Enjoyment-Focused* condition (Fig. 9.11, middle) featured an analog to the open learner model by displaying the students' enjoyment level of each skill—skills that we renamed to appear more playful (e.g., Pattern Perfect, Mad Adder). Enjoyment data was collected in this condition by prompting students to rate every mini-game round that they finished, from 1 to 5 (see Fig. 9.11, bottom). The score of a skill was the average score of all mini-games belonging to that skill. Similar to the *Learning-Focused* condition, we also recommended 3 mini-games from the game types that the student liked the most so far.

- The *Control* condition (Fig. 9.11, right) simply displayed a list of all mini-games and marked the mini-games that had been played with the red text color. Thus, this design was neutral with respect to both learning and enjoyment. Another difference with the Control condition is that students had to finish all the mini-games once before they could replay more rounds. This is a feature that was present in prior studies of *Decimal Point*, so we wanted to preserve it in the *Control*.

In all three conditions of Study 3, students would have to play at least one-half of the content of the overall game—24 mini-games—but could also play additional mini-games, up to a maximum of 72 mini-games. We hypothesized that the *Learning-Focused* version of the game would lead to the best learning outcome, whereas the *Enjoyment-Focused* version of the game would lead to the best enjoyment outcome.

One-hundred and fifty-nine (159) 5th and 6th grade students (77 females, 82 males) from two schools participated. Thirty-five (35) other students were removed from our analyses due to not finishing all of the materials and two (2) other students were excluded due to their gain scores being 2.5 standard deviations from the mean. Materials included the previously discussed pretest, posttest, delayed posttest, along with the two questionnaires (demographic, evaluation). Each student was randomly assigned to one of the three conditions—*Learning-Focused* (55 students), *Enjoyment-Focused* (54 students) or *Control* (50 students).

Our results showed that there were no significant differences in learning outcomes between the three conditions. With pretest score as a covariate, an ANCOVA showed no significant condition differences in posttest scores, $F(2, 155) = 0.201, p = 0.818$, or delayed posttest scores, $F(2, 155) = 0.143, p = 0.867$. Thus, our first hypothesis that students in the *Learning-Focused* condition would learn the most from the game was not confirmed. Regarding enjoyment, we also found that there were no significant differences across conditions according to three enjoyment constructs (i.e., achievement emotion, game engagement, affective engagement); thus our second hypothesis was also not confirmed. We additionally conducted a number of post hoc analyses. For instance, we compared the number of mini-game rounds played in each condition. With respect to the number of mini-game rounds, we found that the total number of mini-games played was *Control* > *Learning-Focused* > *Enjoyment-Focused*. (Recall that students in all three conditions could choose to stop playing at any time after finishing the first 24 mini-game rounds.) With respect to mini-game replay rate, we found that the *Learning-Focused* condition had a higher replay rate than the *Enjoyment-Focused* condition.[5]

While we didn't see either a learning or enjoyment difference between conditions, our dashboards appeared to prompt students toward significantly different learning behaviors, in particular, the *Learning-Focused* students engaged in more repeated practice and the *Enjoyment-Focused* students did more exploration, behaviors that could have led to, respectively, more learning and more enjoyment. Yet a key question

[5] Mini-game replay rate was calculated by how often students would replay any mini-games.

that arises from these results is: Why were there no learning or enjoyment differences between conditions given these behaviors?

Regarding learning: since the *Learning-Focused* version of the game used the often-effective BKT algorithm and an open learning model, one might have expected that condition to have shown significantly more learning gains than the other two conditions. We have a couple conjectures as to why this did not happen. First, while the *Learning-Focused* condition clearly encouraged blocked practice (i.e., playing mini-games with the same underlying skill back-to-back) it could be, as has been shown in some prior research for different domains (Carvalho & Goldstone, 2015), that interleaved practice is equally as effective as blocked practice. This explanation seems especially likely given the limited number of skills emphasized in *Decimal Point*—essentially only five different skills. Second, although there were obvious differences in the game dashboards and choices presented to students between the conditions, they still spent the majority of their time playing the actual mini-games, which are identical across conditions. In other words, even with the choices students were allowed to make, they were exposed to similar instructional content across conditions.

Regarding enjoyment, it is important to note that our study was conducted in classrooms, where students had limited time each day to play the game, were subject to teacher and experimenter expectations, and were aware of the posttests still to come. Some prior studies have, in fact, shown that game play enjoyment can be lost in the classroom (Rice, 2007; Squire, 2005). Thus, the intended playful and more enjoyable nature of the *Enjoyment-Focused* condition may have been reduced for this reason. Alternatively—and similar to the second explanation for no learning differences—students in the *Enjoyment-Focused* condition may not have experienced more enjoyment because they still spent the majority of their time in the mini-games, which are identical across conditions. In short, both with respect to learning and enjoyment, the student experience and exposure to mini-games may have been more similar across conditions than we intended.

9.4.4 Study 4: Hints and Error Messages in a Learning Game

While hints and feedback may seem an obvious inclusion to a learning game, the research is divided on this point. On one hand, much of intelligent tutoring systems research has demonstrated the learning benefits of providing cognitive hints and feedback to students (VanLehn, 2006, 2011; Woolf, 2008; Xu et al., 2019). Timely, contextualized feedback (Ahmadi et al., 2023; Hattie & Timperley, 2007) could also be helpful to students' learning as they engage with digital learning games. On the other hand, it could be that hints and feedback might disrupt the hoped-for engagement (Bouvier et al., 2013) and flow (Czikszentmihalyi, 1990) of students during game-based learning, a key to learning with games. Some studies have, in fact, precisely uncovered this issue (Moyer-Packenham et al., 2019; O'Rourke et al., 2014). For instance, O'Rourke et al. (2014), in an experiment involving over 50,000 students with *Refraction*, a digital learning game to help students learn fractions, explored different hint types (concrete versus abstract) and hint presentation (by level versus reward). In a 2 × 2 comparison of hint type and hint presentation, plus a condition with no hints at all, they found that students in the no-hint condition learned more than students in any of the other conditions. Thus, we were intrigued with how hints and feedback would help or hurt students in the context of learning with *Decimal Point*, and Study 4 explored the question:

> *Study 4 RQ 1*: Do students benefit from hints and error messages provided in the context of the *Decimal Point* game?[6] (McLaren et al., 2022b).

In addition to the exploration of hints and feedback, the pandemic provided a rare opportunity to explore the use of learning games in the classroom versus learning games at home. While we were conducting Study 4, during the winter and spring of 2020 and after having already administered the study at two K-12 schools, the COVID pandemic forced students across the U.S. to learn from home. Thus, we conducted Study 4 at the final three schools with students playing the game online at home. While the pandemic was of course unfortunate for students in the U.S. and around the world, this change in the study context provided us with a unique possibility to contrast how hints and error messages worked in classrooms versus at home. Thus, a second research question we pursued in this study was:

> *Study 4 RQ 2:* How does instructional context (i.e., classroom vs. remote) impact learning with the game? (McLaren et al., 2022b)

To conduct this study we extended the original, low agency version of the game. In the *Hint* condition students played a version of the game that, in addition to correctness feedback, also provided on-demand hints and error messages for common student errors (i.e., when students made a common error, they received a message specifically addressing the error immediately after entering the incorrect response). The hints were developed together with a mathematics education specialist who participated on the project (Jon Star, Harvard University School of Education). Hints were context-sensitive and three levels in length: the first level reminding the student of their goal and the general procedure to solve the problem, the second taking the student through the mathematics procedure specifically applied to the current problem, and the third providing the student with the answer, also called the "bottom-out hint." In the *No-Hint* condition students played the original version of the game that provided no hints and only correctness feedback (i.e., turning correct answers green and incorrect answers red) within the individual mini-games. These conditions are depicted in Fig. 9.12. Some examples of hints and error messages in the *Hint* Condition are shown in Table 9.3.

The study effectively became a 2 × 2 design, crossing *Hint* and *No-Hint* with *Classroom* and *Remote* game play. The study was conducted in two phases. The

[6] Note that before this study, the *Decimal Point* learning game did not include hints, beyond providing correctness feedback (red and green highlighting), nor error messages focused on the common decimal misconceptions.

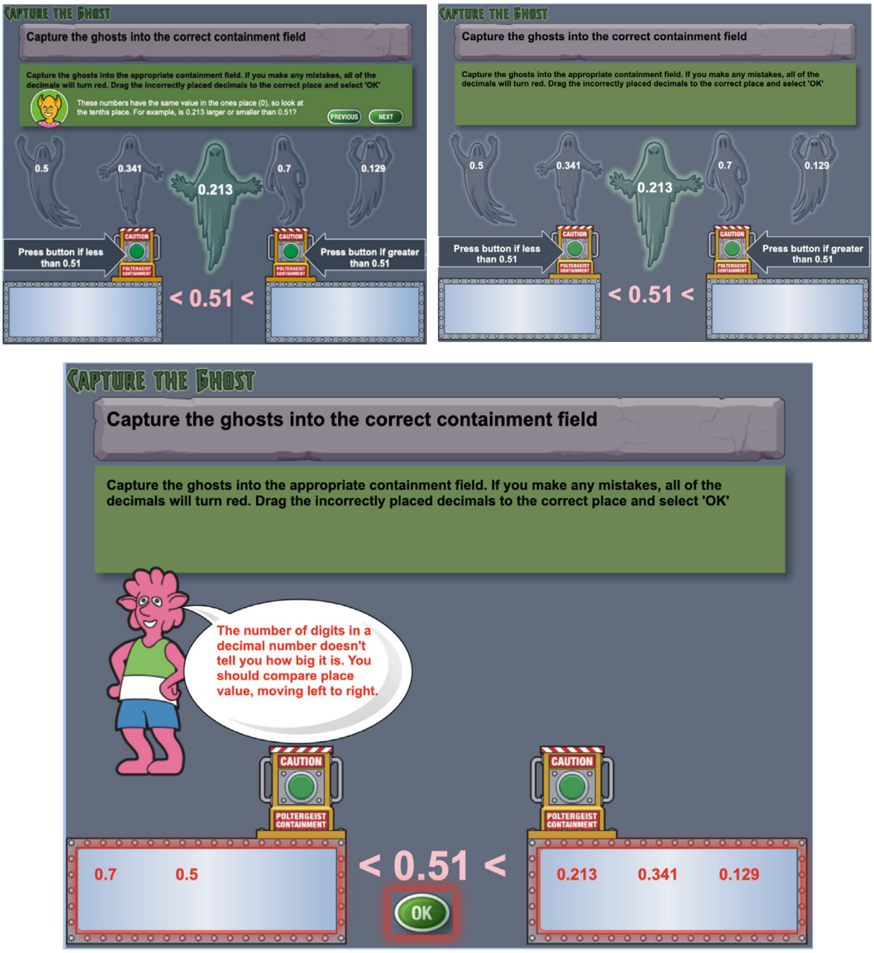


Fig. 9.12 Screenshots of the two conditions of Study 4 with *Decimal Point*. On the top left is a screenshot from the *Hint* condition–an example of an on-demand hint that was added to the game (see the "Hint" button and message in the middle of the screenshot), while on the top right shows a screenshot from the *No-Hint* condition version of the same mini-game. On the bottom is another screenshot from the *Hint* condition, an example of an error message, resulting from a student exhibiting a common misconception

first phase, conducted in the classroom pre-COVID at two schools, had a total of 151 5th and 6th grade students, sixty-seven (67—31 females, 36 males) assigned to the *Hint* condition and eighty-four (84—41 females, 43 males) assigned to the *No-Hint* condition. For this phase of the study, we assigned students to condition by class, due to concerns about students within a classroom observing one another's work and seeing differences in the game (i.e., students without hints noticing their classmates receiving hints, and students with hints might share them with classmates

Table 9.3 Example hint and error messages in *Decimal Point*

Mini-game problem type	Hint examples	Error message example
Sorting	*Level 1*: Compare digits in the same place values of the decimal numbers, moving from the leftmost digit to the rightmost *Level 2*: Since these numbers all have the same ones place (0), compare the tenths place. Which has the smallest tenths place? *Level 3*: 0.0234 has the smallest tenths place, followed by 0.14, 0.323, 0.4 *(These are the three hint levels provided when the student is given a sorting problem with the decimal numbers 0.14, 0.4, 0.0234, 0.323)*	Start by comparing the first digit to the right of the decimal point, even if the digit is 0 *(If the student is presented with sorting the decimal numbers 0.14, 0.0234, 0.323, 0.4)*
Number line	*Level 2*: If you divide the space between 0 and 1 into two pieces, 0.5 is at the end of the first piece. Is 0.456 smaller or larger than 0.5? *(Level 2 hint when the student is given a number line problem to place 0.456 on a number line running -1.0 to 1.0)*	0.456 is greater than 0, so it goes to the right of 0 *(If student clicks to the left of 0, to where the decimal number would be negative)*

not receiving hints). We asked teachers to characterize classes as low, medium, and high performers and then did quasi-random condition assignments so that we had close to the same number of classes of each level within each condition. Eighteen (18) students were excluded (8 in the *Hint* condition and 10 in the No-Hint condition) for failing to complete the materials. An additional student in the *Hint* condition was excluded for performing more than 3 standard deviations below the mean on the posttest and delayed posttest.

For the second phase of the study, when students were working from home due to COVID, three schools with a total of 126 6th grade students participated in the study remotely (64 female, 62 male), with sixty-four (64) students assigned to the *Hint* condition and sixty-two (62) students assigned to the *No-Hint* condition. For this phase, we randomly assigned students to condition, since there was no longer a concern about students seeing one another's work. Ninety-seven (97) students (51 in the *Hint* condition and 46 in the *No-Hint* condition) were excluded from analyses for failing to complete the materials in the allotted time. In summary, the numbers for each of the 2 × 2 conditions is shown in Table 9.4.

Table 9.4 Conditions in 2 × 2 study 4

Condition	N	Female	Male
Hint/classroom	67	31	36
No-hint/classroom	84	41	43
Hint/remote	64	33	31
No-hint/remote	62	31	31

The key results of Study 4 were as follows. Regarding completion rate, the different instructional settings led to significantly different completion rates: *Classroom* students completed the materials at a rate of 88.8%; *Remote* students completed at a rate of only 56.5%. Regarding learning, the *Remote* students learned significantly more than the *Classroom* students, likely due to the fact that in the *Remote* condition more of the students with lower prior knowledge (and/or students with less at-home support) failed to complete the materials. In addition, the two versions of the game, *Hint* and *No-Hint*, led to different *Classroom* versus *Remote* results. In particular, on the delayed posttest, students in the *No-Hint* condition did significantly better than the *Hint* condition in the classroom, while there was no significant difference between conditions at home. Another finding was that female students learned more in the classroom than male students, but the same effect did not occur remotely. We also conducted some post-hoc analyses, finding that students in the *Hint* group used significantly more hints in the *Classroom* than *Remotely*. In addition, higher prior knowledge students used hints more productively, with a significant negative correlation between hints and learning gains.

Some interesting conclusions emerge from these results. First, the different completion rates, as well as better test performance for *Remote* students, were likely due to more and better supervision and guidance in the classroom than at home. The students in the classroom ($N = 151$) were monitored by experimenters and teachers. On the other hand, students at home ($N = 126$), especially because this was at the beginning of the pandemic, may have been unmotivated and not pushed to work by their caretakers at home. The higher performing students working from home likely persevered more, completed the materials more frequently, and thus performed better. Second, why did students in the *No-Hint* condition do better in the classroom on the delayed posttest? While at first this may seem counter-intuitive, in light of the Interactive-Constructive-Active–Passive (ICAP) framework from Chi and Wylie (2014), this is perhaps not so surprising. *No-Hint* students may have worked harder, struggling harder to construct their knowledge, and thus learned more. In support of this, a learning curve analysis showed us that *No-Hint* students initially did worse than *Hint* students, but eventually caught up with their *Hint* counterparts. Finally, why did female students in the *Classroom* condition do better than male students, but not remotely? This was due to girls performing the same in both contexts, but, interestingly, boys did much better at home. This finding aligns with some prior research that girls tend to outperform boys in classroom settings (Dwyer & Johnson, 1997; Entwisle et al., 1997).

9.4.5 Study 5: Comparing Different Forms of Prompted Self-Explanation in a Learning Game

Another learning science principle that intrigued us with respect to game-based learning was self-explanation. Thus, for Study 5, conducted in the spring of 2021, we set about exploring the best approach to prompt self-explanation within *Decimal Point* (McLaren et al., 2022a; Nguyen et al., 2023a). Prompted self-explanation is a feature of instructional technology in which learners are induced to explain their work; it is one of the most robust of learning science principles, supported by decades of research (Chi et al., 1989, 1994; Darling-Hammond et al., 2020; Rittle-Johnson & Loehr, 2017; Wylie & Chi, 2014). Self-explanation supports learners in a number of ways; it helps them fill in gaps in their understanding, revise errors in their prior knowledge, and connect fragmented and disconnected knowledge (Chi et al., 1989; Nokes et al., 2011). When paired with problem-solving, prompted self-explanation can help learners connect problem-solving steps with principles and application conditions (Ainsworth & Burcham, 2007; Aleven et al., 2003). Prompted self-explanation has been shown to be effective in a variety of empirical studies, for instance, in prompting students to explain the principles behind steps in solving geometry problems in a cognitive tutor (Aleven & Koedinger, 2002) and prompting and coaching of self-explanations in a physics tutor (Conati & VanLehn, 2000).

Thus, a key question is:

Study 5 RQ: What form of self-explanation prompt in the context of *Decimal Point* leads to the best learning and enjoyment outcomes?

A variety of approaches have been attempted within instructional technology. Wylie and Chi (2014) cast these various forms of prompted self-explanation along a continuum between unconstrained, on one extreme, and highly constrained self-explanations, on the other extreme. Unconstrained self-explanations allow learners to freely create their own explanations, while presenting the greatest cognitive challenge to learners (i.e., *open-ended self-explanations*). Highly constrained self-explanations, on the other hand, present the learner with a small set of options to choose from to self-explain and thus create the least cognitive challenge for learners (i.e., selecting self-explanations from a menu, *menu-based self-explanations*). Between the two extremes Wylie and Chi cite three other types of prompted self-explanation: *focused self-explanations,* which are constructive but focused in a specific way, such as prompting learners to identify relationships between mental models; *scaffolded self-explanations*, which provide support and/or feedback as learners construct explanations or fill in blanks of an explanation sentence; and *resource-based self-explanations*, in which explanations are selected by learners with the support of a resource, such as a glossary. Chi and Wylie's (2014) ICAP framework for cognitive engagement predicts that students will learn more from cognitively engaging tasks, meaning that *constructive* self-explanations, such as open-ended self-explanations and focused self-explanations, should be more effective for learning than active self-explanations, such as menu-based self-explanations.

Prompted self-explanation has been minimally explored in digital learning games, in which a study by Johnson and Mayer (2010) found that menu-based prompts led to better learning than open-ended prompts. This work was, in fact, the inspiration for us to explore the issue of self-explanation in the context of *Decimal Point*. In other work, Hsu and Tsai (2011) found that prompting learners to explain their errors from a menu of choices led to better learning gains than not prompting for error explanations.

Yet, not all studies have shown learning benefits through prompted self-explanation in 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 between the three conditions 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, in Study 5, we set out to experiment with different versions of prompted self-explanation after problem solving in the game (Fig. 9.13). We decided to experiment with three types of prompted self-explanations across the Wylie and Chi continuum from unconstrained to highly constrained, starting with *focused self-explanations*, in which students must create their own explanations, but with prompting text to focus their attention on a particular aspect of the problem they are explaining. For instance, in Fig. 9.13 at the bottom left, the student is prompted to self-explain just one comparison—9.2111 compared to 9.222—of the sorting problem of four decimal numbers. Next, we created a *scaffolded self-explanation* condition, which essentially presents students with sentence builders that provide all of the components of a correct self-explanation but prompts students to correctly piece together those components into a self-explanation response. Finally, *menu-based self-explanations*—the default self-explanation approach of *Decimal Point*—prompts students to select a self-explanation from a multiple-choice list of predefined options. Note that this approach is essentially what Johnson and Mayer (2010) showed led to the best learning in the context of their game, in contrast to the prediction of the ICAP theory (Chi & Wylie, 2014).

Study 5 involved 214 5th and 6th grade students (114 females; 99 males; 1 did not report) from 4 schools (1 rural, 2 suburban, and 1 urban), with students randomly assigned to condition. Seventy-five (75) were in the menu-based condition, 72 were in the scaffolded condition, and 67 were in the focused condition. An additional one hundred and forty-three (143) students were dropped due to (a) failing to complete part or all of the learning materials or any tests and (b) having participated in one of our studies the previous year. (Note that the relatively high attrition rate was due, at least in part, to running the study during the COVID-19 period. Some students participated in person, some at home, and some in a hybrid format.)

The results of Study 5 showed that students in the *focused self-explanation* group learned more on the delayed posttest than the *menu-based self-explanation* group. There were no other significant effects. Regarding time on task, the *menu-based self-explanation* group spent significantly less time than the *focused* or *scaffolded self-explanation* groups. This indicates that at least the menu-based approach takes less time, i.e., it is more efficient. The only significant effect of engagement was that students in the *focused self-explanation* group reported a significantly higher sense of player mastery.

So, in conclusion, the key finding of this study was that focused self-explanations led to better learning than menu-based self-explanation, without any loss of engagement. This result is in line with Chi and Wylie's ICAP theory (2014), but in contrast to the Johnson and Mayer study (2010) that found menu-based self-explanations led

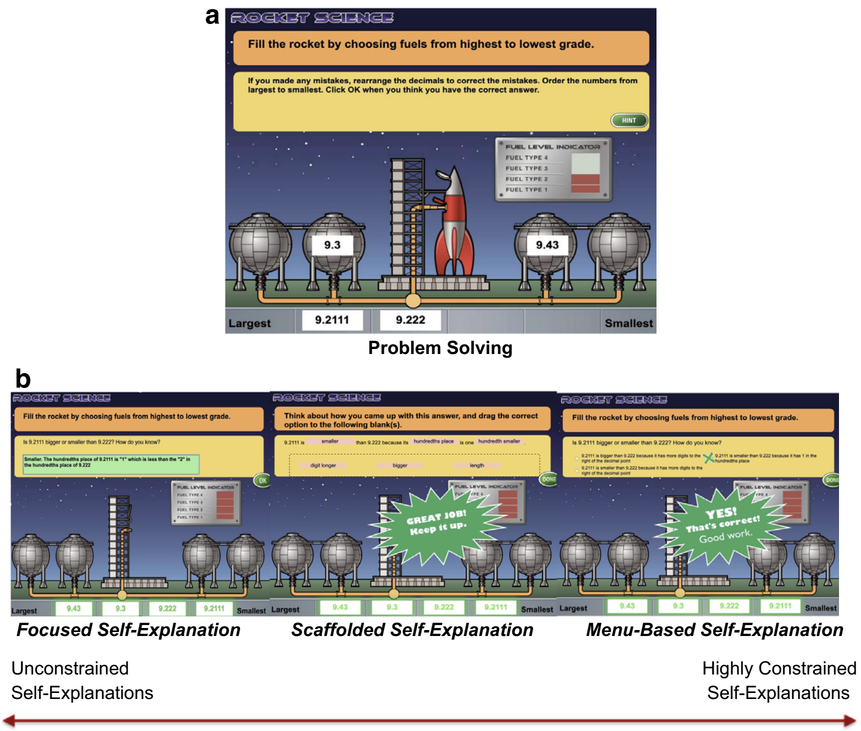


Fig. 9.13 Screenshots of the three conditions of Study 5 with *Decimal Point,* from unconstrained self-explanations to highly constrained self-explanations. On the top is an example problem solving step, within the *Rocket Science* mini-game of *Decimal Point*. On the bottom left is the subsequent prompted self-explanation step, a focused self-explanation. On the bottom middle is a scaffolded self-explanation. On the bottom right is a menu-based self-explanation, the default self-explanation approach of *Decimal Point*, used in all other studies described in this chapter

to better learning than open-ended self-explanations in a game context. This indicates that focused self-explanations used in the context of a digital learning game may be better for deeper, more conceptual learning than other forms of prompted self-explanation, without accompanying loss of game play engagement.

9.4.5.1 Study 5a: Using a Large-Language Model to Assess and Provide Feedback for Self-Explanations in a Learning Game

The recent emergence and advances with large language models (LLMs), and in particular ChatGPT (Ye et al., 2023), intrigued my lab and I, as it did many other researchers. When a commonly available version of ChatGPT appeared in November of 2022, we decided to do a post-hoc study of the data from Study 5 to explore whether

ChatGPT/GPT[7] could provide instructionally meaningful feedback to the focused self-explanations of students (Nguyen et al., 2023b). Given over 5,000 focused self-explanations from students in Study 5, we conducted analyses to assess GPT's capability to (1) solve the in-game exercises of the *Decimal Point* game, (2) determine the correctness of students' self-generated self-explanations, and (3) provide instructionally helpful feedback to incorrect self-explanations.

Study 5a was conducted completely off-line, using the 5,142 focused self-explanations collected from 117 students in Study 5. We had three specific research questions for this study:

> *Study 5a RQ1*: Can GPT correctly answer the problem-solving and self-explanation questions in the game *Decimal Point*? (i.e., Is GPT a good student in this domain?)
>
> *Study 5a RQ2*: Can GPT accurately assess the correctness of students' self-explanation answers? (i.e., Is GPT a good grader in this domain?)
>
> *Study 5a RQ3*: Can GPT provide instructionally meaningful feedback to incorrect self-explanations? (i.e., Is GPT a good teacher in this domain?)

Three coders manually graded as correct, incorrect, or off-topic all of the focused self-explanations from the 117 students, using an iterative process which included inter-rater reliability as a means of assessing coding agreement, as described in (Nguyen et al., 2023a). This resulted in 1000 correct answers, 4076 incorrect answers, and 66 off-topic answers that did not address the question. For the purpose of Study 5a's analysis, we treated off-topic answers as incorrect.

The general approach of our analyses was, for each decimal problem and student self-explanation, to send GPT the question and, in the case of the self-explanations, the student's response and a grading rubric. We developed a script to automatically send all of the prompts to GPT and then harvested all of its answers. We used GPT 3.5 for Study 5a, as that was the current version of GPT when we conducted the study.

For RQ1 we wanted to see how well GPT could solve the *Decimal Point* math problems and self-explanations. Since GPT gives a unique answer each time it is queried, we sent each math question and self-explanation prompt to GPT 10 times to assess how correctly and consistently it handled each. The correctness of GPT's responses to both the problem-solving and self-explanation items were assessed by a math expert on the research team, with the results shown in Fig. 9.14. As can be seen on the left side of Fig. 9.14, GPT was excellent at solving sorting and sequence problems, very good at solving bucket problems, but quite poor at both addition and number line problems. GPT had a much better overall performance for self-explanations than for the problem-solving activities (right side of Fig. 9.14). The only problem type where GPT's explanations were occasionally incorrect was sorting, where it sometimes slipped at assessing decimal place values.

[7] ChatGPT is the chat interface that enables sending data to and receiving data from the underlying GPT model. While ChatGPT is the commonly used term, in fact, we used the GPT API in this study; thus, from this point on I will only use the more precise term: "GPT."

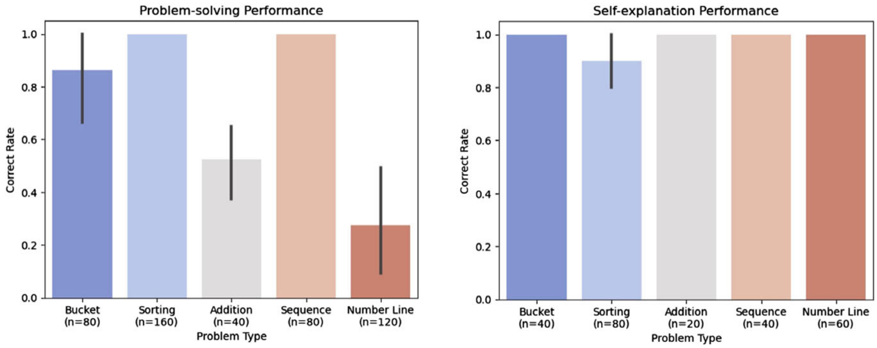


Fig. 9.14 Results of Study 5a's RQ1. GPT's problem solving (left) and self-explanation (right) performance

For RQ2 we wanted to see how well GPT could assess student self-explanations. To do this, we prompted GPT to provide a response of correct or incorrect, per self-explanation, given the self-explanation prompt, the student's self-explanation, and the grading rubric for self-explanations. GPT's performance compared to that of the human coders is shown in Table 9.5. Notice that there were a relatively small number of false negatives (lower left cell, in boldface font), but a much larger number of false positives (upper right cell, in boldface font). Most of these were due to GPT grading an incorrect answer as correct, suggesting that it did not follow the grading rubrics as closely as the human graders did. For instance, for bucket and sorting items, we found that the presence of comparison keywords such as "bigger" or "smaller" was sufficient to get a correct rating from GPT. For example, if the student just wrote "A is smaller than B because it is smaller"—clearly an example of fallacious circular reasoning—GPT would rate it as correct. Similar errors based on shallow keyword matching occurred across all problem types.

For RQ3 we wanted to know whether GPT could provide accurate and instructionally meaningful feedback to students. To generate feedback per incorrect student self-explanation we provided GPT with the self-explanation prompt, the rubric items specific to that self-explanation, and the student's self-explanation. We then coded GPT's feedback according to six relevant categories with results as shown in Table 9.6.

In summary, GPT did much better as a teacher—providing feedback to incorrect self-explanations—than it did as a student—solving and self-explaining the math

Table 9.5 Results of Study 5a's RQ2

	Human: Correct	Human: Incorrect
GPT: Correct	830	**1,118**
GPT: Incorrect	**170**	3,024

GPT's assessment of student self-explanation compared to the assessment of human coders

Table 9.6 Results of Study 5a's RQ3

Category	Description	Results
Accuracy	Does GPT distinguish between partially and fully correct self-explanations?	GPT assigned correct partial credit 75% of the time
Fluency	Is GPT's feedback grammatical and natural sounding?	GPT was 100% proficient in English
Regulation	Does GPT's feedback address all decimal misconceptions reflected in the self-explanation?	Chat GPT was very effective at identifying and addressing student misconceptions
Solution	Does GPT's feedback tell the student the correct answer?	GPT did not provide solutions to 794 incorrect and low-effort self-explanations (e.g., "idk," "by adding up")
Rationale	Does the feedback provide a rationale?	GPT demonstrated good understanding of 85% of self-explanations and provided a range of nuanced explanations (e.g., "your answer is not specific enough")
Encouragement	Does the feedback provide any form of encouragement?	GPT provided encouragement to 20% of the answers ("great job," "keep practicing") and detected 9 cases of inappropriate language used by students

Assessment of GPT's feedback on incorrect self-explanation responses according to six relevant categories

problems—or a grader—assessing the correctness of student self-explanations. In providing feedback to the incorrect student self-explanations, GPT's feedback was high quality and nuanced; it provided encouragement and flagged inappropriate language, even though it was not prompted to do so. It also did very well understanding student answers but provided incorrect feedback more frequently than a teacher likely would have. GPT did less well in solving math problems; it had difficulty with the nuances of math, such as carrying when performing addition and placing points on a number line. (Note that this shortcoming of GPT is now widely recognized, with at least some preliminary suggestions for how to correct it (Wolfram, 2023)). It also struggled a bit in assessing the correctness of student self-explanations, likely due to shallow keyword matching with the grading rubric, which led to many false positives. It appeared not to detect all of the nuances in the grading rubric. Overall, our assessment at the conclusion of Study 5a was that GPT, at least the version 3.5 current at the time of this study, is more suited for conceptual analyses (e.g., giving feedback to self-explanations) than procedural math questions. In short, at the time of Study 5a, GPT was still in a state where a teacher should remain in the loop, double checking answers before they are presented to students.

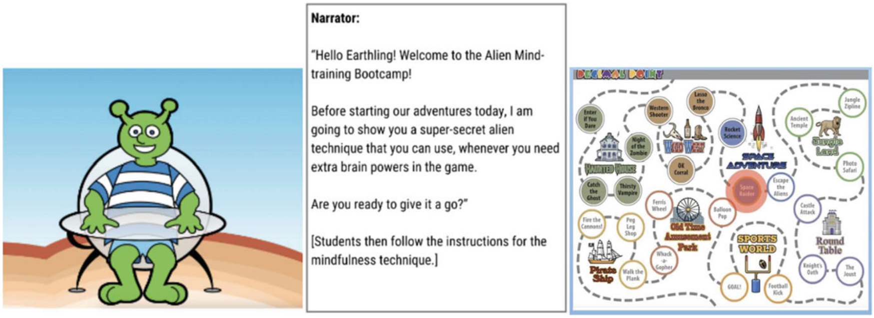
9.4.6 Study 6: Mindfulness Induction When Learning with an Online Game

Study 6, conducted in the fall of 2021, involved an investigation of the interaction of mindfulness—attending to the present moment with focus and without judgment—with game-based learning (Bereczki et al., 2024; Nguyen et al., 2022b). Mindfulness meditation has been shown to support self-regulated learning (Dunning et al., 2019; Takacs & Kassai, 2019), improve attention skills (Dunning et al., 2019, 2022; Takacs & Kassai, 2019), and reduce math anxiety (Samuel & Warner, 2021). On the other hand, the role of mindfulness in children's academic achievement and outcomes is less clear. There have been several studies that have assessed the efficacy of mindfulness-based interventions, but those studies have shown a non-significant, small average effect on learning (Maynard et al., 2017). More promising results have been found with older students and those with ADHD, but still the results are inconclusive (Güldal & Satan, 2020; Singh et al., 2018).

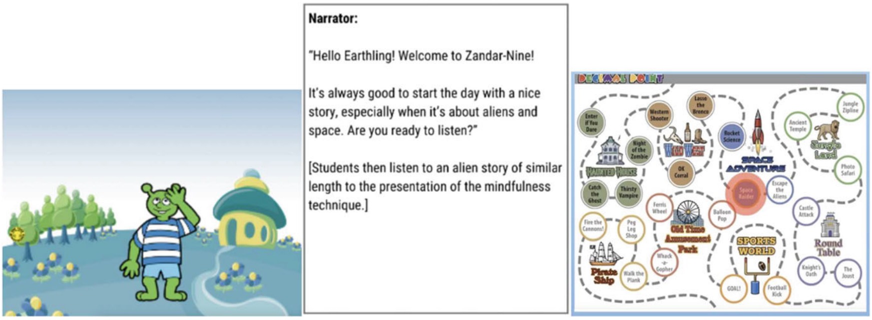
More specific to math skills, it has been shown that executive function—which entails working memory, inhibitory control, and cognitive flexibility—is key to learning math skills (Cragg & Gilmore, 2014). For instance, prior research has found that kindergarten-age children with higher executive function skills but lower math skills are more likely to catch up with their higher-performing peers by the 5th grade than those students with lower executive function skills (Ribner et al., 2017). Furthermore, mindfulness appears to play a role in supporting executive function (Dunning et al., 2019, 2022; Takacs & Kassai, 2019). Yet, even with the promising connection between mindfulness, executive function, and math learning, Study 6 is, to the best of our knowledge, the first to explore the benefits of employing mindfulness as a means for boosting learning in the context of a digital learning game for mathematics. The question we asked in this study is:

> *Study 6 RQ*: Can mindfulness inductions during *Decimal Point* gameplay lead to different behaviors and more learning?

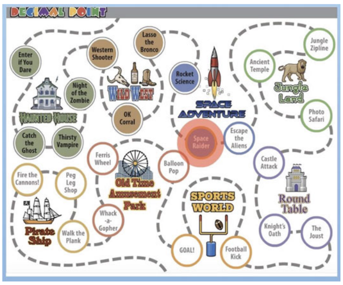
To explore this issue, we created three *Decimal Point* conditions, *Mindfulness*, *Story*, and *Control*, as shown in Fig. 9.15. The order and content of the mini-games within *Decimal Point* during gameplay was identical across all three conditions. The key differences between conditions were as follows. In the *Mindfulness* and *Story* conditions students would listen to a five-minute audio session at the start of each day of the study, prior to playing and learning with *Decimal Point*. In the *Mindfulness* condition (Fig. 9.15, top), the audio content involved an alien character prompting students to be mindful by asking them to close their eyes, focus on their breath and sounds in the environment, and let go of passing thoughts (Vekety et al., 2022). In the *Story* condition (Fig. 9.15, middle), the audio content was age appropriate, emotionally neutral (i.e., not emotionally arousing or upsetting) science fiction stories that were unrelated to the learning content. This condition was created to control for time with respect to the mindfulness condition, but with material that was not designed to induce mindfulness. Both the *Mindfulness* and *Story* conditions also featured in-game, minute-long reminders that would appear when the student had made three consecutive errors in a mini-game. In the *Mindfulness* condition, students would be encouraged to slow down, close their eyes, and focus on their breath for a moment. In the *Story* condition, students would listen to a joke from an alien character. Each reminder would appear at most once every 10 min to avoid overwhelming the students. Finally, students in the *Control* condition (see Fig. 9.15, bottom) were not presented with any opening audio material before starting gameplay each day, nor did they receive reminders when they made errors.



***Mindfulness* condition, each day begins with a mindfulness induction**



***Story* condition, each day begins with a science fiction story**



***Control* condition, students play the standard version of the *Decimal Point* game**

Fig. 9.15 Screenshots of the three conditions of Study 6 with *Decimal Point*

We hypothesized that students in the *Mindfulness* condition would learn the most, due to the expected additional benefits of mindfulness when added to game-based learning. We also hypothesized that students in the *Mindfulness* condition would take more time and make fewer errors during game play than the other two conditions.

The final analyzed set of participants in Study 6 included 166 students (90 females; 76 males) from three schools, with 53 students randomly assigned to the *Mindfulness* condition, 56 to the *Story* condition and 57 assigned to the *Control* condition. A total of 77 students were excluded from our analyses because they did not complete all of the materials.[8] Note, importantly, that students were randomly assigned to condition, meaning that every class would have a mix of students in all three conditions. (We return to this point in the discussion of the next study, Study 6a.)

The results of Study 6 found no differences in learning outcomes across the three conditions (neither pre-to-posttest nor pre-to-delayed posttest), time spent on the game, or error rates while playing. In other words, our hypothesis of the benefits of mindfulness was not confirmed, i.e., embedding mindfulness prompts within the game did not enhance learning nor change students' gameplay behaviors. Thus, at least this particular study suggests that a mindfulness induction does not enhance learning within digital learning games. Alternatively, we may not have successfully induced a state of mindfulness in the students; we explored this topic in the next study, Study 6a.

9.4.6.1 Study 6a: Mindfulness Induction When Learning with an Online Game, with a Manipulation Check to Test for the Impact of the Mindfulness Induction

Because we were unsure whether our online approach to inducing mindfulness in Study 6 had the desired effect, we ran another study—Study 6a, conducted in the spring of 2022—in an attempt to replicate the findings of Study 6 but to examine whether we had, in fact, induced mindfulness in students (Bereczki, et al., 2024). Thus, besides the question we had already explored in Study 6, we also explored the following question:

Study 6a, RQ: Did we manage to induce mindfulness in students in the mindfulness condition?

We hypothesized that students in the *Mindfulness* condition would report higher state mindfulness immediately after the mindfulness manipulation than those in the *Story* and *Control* conditions. The materials and procedures of Study 6a were the same as in Study 6, except that at the beginning of each game session, after students in the *Mindfulness* and *Story* conditions engaged in the initial mindfulness manipulation and heard a story, respectively, the students completed a state mindfulness measure.

---

[8] Note that the final population of students reported in Bereczki, et al., 2024 – 227 – is larger than what is reported here and in Nguyen et al., 2022b – 166. This is because Bereczki, et al., 2024 applied a less stringent exclusion criteria: students were excluded from the analyses if they did not complete at least 80% of the intervention game (versus 100% completion of pretest, intervention, posttest, and delayed posttest, as reported in Nguyen et al., 2022b).

Students in the *Control* condition did not have any intro procedure, so they completed the state mindfulness measure at the beginning of each of their game sessions. The state mindfulness check was measured with a 5-item scale adapted from the MAAS-A (Brown et al., 2011), so that statements would reflect students' experience at the moment. Example items of the scale include: "Right now I find it difficult to stay focused on what's happening." or "Right now I'm doing things automatically, without being aware of what I'm doing." Items were answered on a seven-point scale (1 = Not at all; 7 = Very much so).

Study 6a was also conducted in 5th and 6th grade classrooms across 3 additional public chools. A total of 193 students originally participated in the study, but 16 were excluded from the analyses because they did not complete at least 80% of the games. Thus, the final sample included 177 students (86 females, 91 males), with 62 students randomly assigned to the *Mindfulness* condition, 61 to the *Story* condition and 54 to the *Control* condition.

Similar to our results in Study 6, we found no evidence that students in the *Mindfulness* condition learned more from pretest to posttest or from pretest to delayed posttest than those in the other two conditions. We also found no difference in problem-solving duration and errors made among the three conditions. We did find a marginally significant condition effect on correctness after reminder between the *Mindfulness* and *Story* conditions: Students in the *Mindfulness* condition made more correct steps after reminders than those in the *Story* condition. Finally, a univariate ANOVA showed no significant effect of condition on students' state mindfulness after inductions (*Mindfulness* or *Story*) or at the beginning of the game sessions in the *Control* condition, $F(2, 174) = 0.51$, $p = 0.60$, $\eta p^2 = 0.006$. Also, neither of the planned comparisons were significant: *Control* vs. rest ($p = 0.65$) and *Story* vs. *Mindfulness* treatment ($p = 0.37$). These results show that we did not manage to induce mindfulness.

In conclusion, the lack of a mindfulness effect in both Study 6 and Study 6a may be due to the classroom context. First, we conducted mindfulness as an online, self-guided activity, as opposed to the more common instructor-led group activity. It is also likely that the presence of classmates who were engaging with different versions of the game—recall that *Mindfulness*, *Story*, and *Control* students were mixed together in classrooms—introduced distractions that may have reduced mindfulness. It may also have been that students were self-conscious about closing their eyes and following the mindfulness instruction. Given these possibilities, we don't see our findings as conclusive with respect to whether mindfulness can enhance learning with a digital learning game; further research is needed, with changes made to the way mindfulness is induced.

9.4.7 The Gender Effect: A Replication Across Multiple Studies

As mentioned in the discussion of Study 1, we became interested in whether girls or boys benefited more from playing *Decimal Point*. This interest arose from our knowledge of the gap between girls and boys in math achievement (Breda et al., 2018; Wai et al., 2010) and a desire to lessen this gap, at least in a small way, with our learning game. The question we asked ourselves was:

> RQ across studies: "Do female students benefit more, less, or the same as their male counterparts playing the game?" (Nguyen et al, 2022a)

The math gender gap may be attributed to *stereotype threat,* in which reminders of social group stereotypes can impact the behavior and performance of members of that group (Spencer et al., 1999). Despite a reduction in gender-based differences in math achievement over recent decades (Lindberg et al., 2010; Reardon et al., 2019), early-emerging stereotypes, such as the perception that males excel in math, can persist from childhood through adulthood (Cvencek et al., 2011; Furnham et al., 2002; Nosek et al., 2002; Passolunghi et al., 2014). Consequently, these perceptions may impact the performance of female students in mathematics and influence their interests and, eventually, their career choices (Adams et al., 2019; Bian et al., 2017; Ochsenfeld, 2016). Ultimately, addressing stereotype threat involves the complex task of promoting self-efficacy, interest, and achievement among female students, while simultaneously mitigating math anxiety and stereotype threat.

Through data from six of our *Decimal Point* studies—in particular, Studies 1, 2, 2a, 3, 4, and 5—involving approximately 1,100 students, we identified a consistent gender effect that was first seen in Study 1 and then replicated across five other studies: male students tended to do better than female students at pretest, while female students tended to learn more from the game, catching up to their male counterparts by posttest. (The first 4 of these gender effect studies, involving more than 600 students, are reported in Nguyen et al, 2022a). In addition, female students were more careful in answering the self-explanation questions, which significantly mediated the relationship between gender and learning gains in two of the first four studies (Nguyen et al, 2022a). More specifically, we found that female students made less errors and "gamed" the self-explanation step of *Decimal Point* mini-games significantly less than male students, resulting in more learning for female students, less for male students (Baker et al., 2024). These findings show that digital learning games, in combination with prompted self-explanation, can be effective tools for bridging the gender gap in middle school math education, which in turn could lead to the design and development of more personalized and inclusive learning games. Given the complexity of gender and the need to conduct research that goes beyond a binary approach to gender (Hyde et al., 2019), we are currently conducting research that measures multiple dimensions of gender, including gender identity, gender typicality, and gender-typed interests, activities, and traits (Hyde et al., 2019), to understand which aspects of gender explain the differences we have observed in learning behaviors and outcomes (Liben & Bigler, 2002). Preliminary results suggest that this multidimensional approach of using gender-typed scales may better explain students' feelings toward and preferences about digital learning games than the binary gender (Nguyen et al., 2023c).

9.5 Key Take-Aways: Digital Learning Game Findings

The decade-long research program that the McLearn Lab has conducted with the *Decimal Point* learning game has led to some important, some not so important, but always intriguing learning science results. The wide variety of studies, all conducted with *Decimal Point* as the centerpiece, has afforded the opportunity to investigate many and varied issues. In this section I will highlight the most noteworthy findings of the McLearn Lab's research program with *Decimal Point*.

Our most *fundamental* research finding, from Study 1 and reported in McLaren et al. (2017a), uncovered that digital learning games can surpass conventional online methods in improving engagement and learning outcomes. Prior to our Study 1, educational technology research had presented mixed results regarding the comparative advantages of learning games for mathematics and more traditional learning technologies (Mayer, 2014). Thus, given the state of game-based learning science as of the publication of our seminal 2017 paper, this was an important and novel finding.

Our most *robust* finding has been that female students have exhibited greater learning gains from the *Decimal Point* game as compared to their male counterparts. This finding, again, first found in Study 1 and then replicated across five other studies—Studies 2, 2a, 3, 4, and 5—featuring diverse versions of the *Decimal Point* game, serves as the focal point of my student Huy Nguyen's PhD thesis and is extensively discussed in Nguyen et al., 2022a. That paper covers the first 4 of the gender effect studies. We continue to pursue this issue in our most recent studies, including two that have not yet been published. For one of those studies, we created a new game, *Ocean Adventure*, which has exactly the same decimal content and instruction as *Decimal Point*, but with an entirely new, masculine-oriented narrative (see Fig. 9.16), which we designed based on a survey conducted with 333 students, designed to probe the preferences of male and female students (Nguyen et al., 2023c). The goal was to see whether boys would be more engaged in the new game and thus learn as much, or more than girls. While there was some evidence that boys were more engaged in the new game, they did not learn more. Ultimately, we hope an important practical outcome that will emerge from this line of inquiry will be the identification of game-based learning guidelines for alleviating the stereotype threat in female students, thus resulting in better math learning outcomes—and eventually better career prospects—for female students.

Perhaps our most *surprising* finding in this line of research—although the oft-replicated gender effect would also be a good choice—emerged from Study 4 and McLaren et al. (2022b) where we reported that hints and error messages within

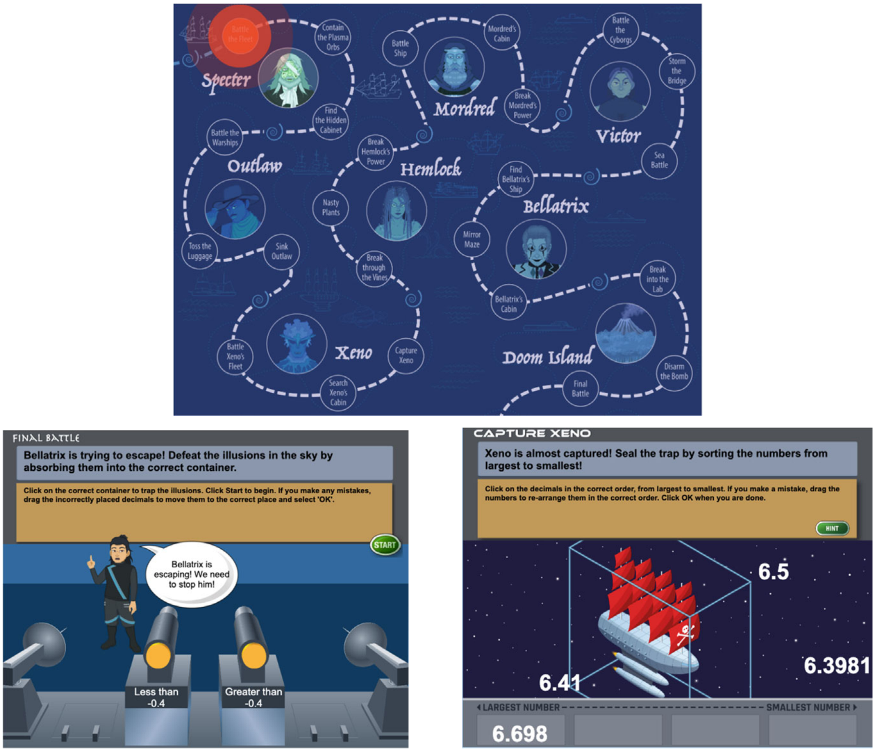


Fig. 9.16 Screenshot of the new *Ocean Adventure* learning game, along with screenshots of two of its mini-games. This game was specifically designed to be more male-oriented, according to a survey we conducted with over 300 middle school students

a digital learning game do *not* necessarily enhance learning outcomes. Given the fundamental use of hints and error feedback in other forms of educational technology, most especially intelligent tutoring systems (VanLehn, 2006, 2011; Xu et al., 2019), this was particularly unexpected. Our conjecture is that the interruption of game flow (Czikszentmihalyi, 1990) through hints and feedback may have led to unintended and negative learning consequences. This unexpected finding, on the other hand, provides evidence for the ICAP Framework (Chi & Wylie, 2014), suggesting that the struggle students may encounter—for instance, by not having hints and feedback to lean on—might contribute to deeper learning in the context of a digital learning game.

The most *central contribution to learning science* of this line of research comes from Study 5 and our findings reported in (McLaren et al., 2022a) regarding the advantages of prompted self-explanation within the context of game-based learning. Specifically, focused self-explanation — a form of prompted self-explanation in which students must generate their own explanations — emerged as the most effective form of self-explanation that we studied. This finding is important, as it presents a counter to the findings of Johnson and Mayer (2010) in which menu-based self-explanation led to better learning outcomes in the context of a digital learning game. They conjecture that minimizing impact to student game play and flow—as a menu-based approach surely does—led to their findings, whereas we conjecture that the constructive approach inherent in a focused, open-ended self-explanation led to productive student struggle (Chi & Wylie, 2014) and thus to our findings. Further exploring these different outcomes is an excellent direction for future studies. Perhaps most importantly, we discovered that prompted self-explanation likely holds the key to understanding the gender effect that we have found in many of our studies (Nguyen et al., 2022a).

Finally, our most *forward-looking* finding comes from Study 5a and Nguyen et al (2023b), in which we investigated the contribution that GPT could make in providing feedback to students who play and learn from *Decimal Point.* With AI, and especially large language models, providing an inflection point for how technology will be used and contribute to many aspects of society, it was important and timely for us to investigate how AI could impact learning with educational technology generally and our game more specifically. While the study we conducted was preliminary—done completely in post-hoc fashion with off-line data—it provided some key insights into how students might benefit from large language models.

9.6 Key Take-Aways: Use of a Digital Learning Game as a Research Platform

In essence, *Decimal Point* has functioned not only as a research tool but has become a more general *research platform*, pushing the boundaries of our understanding of how learning science can be effectively integrated into the design and implementation of learning games. We've discovered that a digital learning game can provide a rich environment for experimenting with many aspects of learning. The many and varied features of online games—both for learning and playing—furnish an excellent framework for systematic exploration, encompassing learning aspects such as the potential of student agency during game play, the tension between enjoyment and learning in game-based learning, and the benefits of hints and feedback in game-based learning, among other facets. We have leveraged the game as a platform for exploring all of these issues—and more.

A key to *Decimal Point* acting as a research platform has been its overall architecture and design. For instance, we've discovered that a learning game can be built with an underlying tutoring system engine and ITS principles (Aleven et al., 2016). The ITS model and approach has helped to structure instructional aspects of the game. Principles of ITSs, such as providing immediate feedback and on-demand hints, influenced the design of both the game and our studies with the game. While "gamification"—attempting to improve learners' engagement and experience with educational technology through, for instance, the inclusion of badges, points, leaderboards, and interactive playful agents (Landers & Landers, 2014; Landers et al., 2017)—is a popular approach to studying how game techniques can make learning more enjoyable and effective (Long & Aleven, 2014, 2018; Tahir et al., 2020), in this line of research we have shown what is possible when a game is built *from scratch* with underlying ITS principles, a more fundamental design approach than gamification. Essentially, we have shown that the degrees of freedom for experimentation are ultimately much wider (and arguably richer, as well) when a learning game is designed originally as a game, rather than as a gamified tutoring system.

Another important question that arose during the use of *Decimal Point* as a research platform is whether games are better suited for learning at home or in a classroom. Students are used to digital games being a fun, at-home activity. In contrast, they know that in school activities are more structured, less free and perhaps less fun. So, can we fully engage students in school with a perceived out of school activity? This is an excellent, open question. This issue came up in multiple instances over the years of experimenting with *Decimal Point.* For instance, in Studies 2 (Nguyen et al., 2018) and 2a (Harpstead et al., 2019)—in which we essentially studied the trade-offs between student autonomy and game system control, students in the classroom may not have really felt in control of their learning, due to the influence of the teacher and classroom context. Perhaps autonomy and agency would have been more greatly felt at home? We didn't have the opportunity to explicitly explore this, but it would be an interesting topic still to investigate. *Decimal Point*'s infrastructure and Internet implementation would allow for such a study. As another example, Study 4 (McLaren et al., 2022b)—the hints study that ended up being ½ conducted at school, ½ at home—was a step toward exploring this dichotomy that may lead to further contrasting studies.

A key aspect of intelligent tutoring systems, first articulated by Kurt van Lehn (2006), is the distinction between the "outer loop", in which problem ordering and selection is handled, and the "inner loop", in which student interactions within problems occur, is another way in which *Decimal Point* has acted as an excellent research platform. In our studies, student agency, indirect control, and mindfulness—all outer loop activities—did not yield significant differences between conditions. Conversely, the inner loop, which involves elements we tested such as hints and errors and self-explanation, emerged as a locus of noteworthy variations in learning outcomes. This is likely due to the learning aspect of *Decimal Point* being more prominent than the game aspect, which makes it harder for individual tweaks on the game mechanics to significantly change learning, but also makes the game "safer" and more robust to changes—we have never seen a condition that did not lead to significant pre-post learning gains.

Finally, a very interesting and important observation—since it has meaningful implications for game design and for how we should approach future game-based learning research—is that many of our interventions *did not actually show learning differences between conditions.* Our most significant learning difference was found in Study 1 when we compared the game to a conventional learning technology (McLaren et al., 2017a). There are surely different reasons for the lack of condition differences in each of the game versus game studies; this could be evidence that it is tricky to significantly alter learning outcomes by tweaking individual features of a game. This further suggests that perhaps students are more consistent in how they play learning games—or more resistant to our efforts to change their ways of playing—than we might think. This may have been due, at least in part, to the mostly-unchanging basic instructional approach of *Decimal Point* being more prominent than the game aspect. Throughout the decade of the game being used as a research platform, the basic precepts of *Decimal Point*'s instructional approach remained (a) a focus on decimal misconceptions and (b) an underlying ITS instructional approach. This surely made it difficult for individual tweaks to the game mechanics to significantly change learning. At the same time, it also likely made the game "safer" for and more robust to changes, as mentioned, we have so far not seen a condition that did not lead to significant pre-post learning gains. In short, a lesson for future game-based research platforms might be to create a more modifiable instructional component for experimentation.

9.7 Conclusions

In conclusion, I will propose a few possible future directions for the McLearn lab's continuing work with *Decimal Point* more specifically and for digital learning games research more generally. One direction that could be further explored in connection with digital learning games is the "Assistance Dilemma" (Koedinger & Aleven, 2007), reaching beyond the standard textual hints and feedback support we investigated in Study 4. The Assistance Dilemma raises the question of the trade-offs between giving and withholding help in the context of instructional technology. Giving help can move students forward who are stuck; it can also lead to shallow learning. Withholding help can push students to think and learn more deeply; it can also lead to frustration when they are truly stuck. The trade-offs in a game-based learning context may differ from other educational technology, however, given how games are intended to promote flow and engagement. Our Study 4 results, in which the students who received hints learned less, seemed to indicate that withholding help was the correct choice for learning with *Decimal Point,* perhaps because of how the particular help we provided might have disrupted student engagement. One aspect of the Assistance Dilemma that could be further investigated would be the value of using a different model of providing help than allowing students to simply request it and to receive standard textual hints. Perhaps, for instance, instead of providing on-demand hints, students could be prompted to ask for help when they have clearly demonstrated they need support. Such an approach might involve less disruption to a student's engagement with a game, yet still provide timely assistance. Another aspect of the Assistance Dilemma that could be explored—and which would fit the context of game-based learning well—would be the use of non-textual hints, such as animations (Berney & Bétrancourt, 2016; Nathan, 1998; Scheiter et al., 2010) or visual representations (Nagashima et al., 2021). Given the highly visual and engaging nature of learning games, not to mention evidence that visual models can support the learning of mathematics (Hegarty & Kozhevnikov, 1999; Luzón & Letón, 2015), animated or visual hints might provide better, easier to process, and more engaging help in digital learning games than standard textual hints.

Another intriguing avenue worthy of investigation involves the incorporation of learning from erroneous examples, which have been shown to be an effective learning technique in a variety of studies (Adams et al., 2012; Durkin & Rittle-Johnson, 2012; Grosse & Renkl, 2007; McLaren et al., 2012; Tsovaltzi et al., 2012), in the context of a learning game. Erroneous examples are worked examples of problem solving in which one or more of the steps has an error, typically a common error made by students. This is, in fact, how the McLearn Lab started this line of research with learning games (although we early on departed from this exploration). In particular, we originally set out to see if we could create a learning game around erroneous examples, which have a natural interactivity or playfulness associated with them in presenting students with the challenge of errors to fix. One could imagine a version of *Decimal Point* in which students don't (always) directly solve problems themselves but instead are challenged to find and fix errors made by the fantasy characters in a gameful way. Furthermore, providing badges and prizes to students as they manage to find and fix the errors could provide an even more gameful aspect to *Decimal Point*—and perhaps create a blueprint for a new type of learning game.

Of course, as discussed, the recent rise and huge steps forward in large language models and artificial intelligence raises some intriguing possibilities for AI applications in the context of digital learning games. In a recent book chapter (McLaren & Nguyen, 2023) we described the many ways that AI has already been used in digital learning games, including adapting game play and problems, AI-powered dashboards, educational data mining for game improvement and identifying cognitive, behavioral, and affective aspects of learning, and AI-powered non-player characters (NPCs). As discussed earlier, we have experimented with the first three of these AI approaches within the *Decimal Point* game, using older AI techniques than LLMs. LLMs present new and exciting opportunities to create and extend learning games with intelligent capabilities. Our Study 5a (Nguyen et al., 2023b) was a very promising first step toward incorporating the latest advancements in AI into digital learning games, but there are many other directions that could be pursued. For instance, a large language model could be called upon to not only provide cognitive feedback, as per our recent study, but also meta-cognitive (Hattie & Timperley, 2007) and affective (Howard, 2021) feedback, both of which are valuable to learners and for which there is extensive information on the Internet from which an LLM could generate feedback. An illuminating study would be one that compares a learning game that has manually-created feedback, the typical case, to feedback generated by an LLM. Another possibility for LLMs in the context of learning games would be replacing NPCs, which are currently implemented with earlier generation natural language processing (NLP) techniques, with LLMs. Given the superior language capabilities of LLMs, this could potentially be one of the most significant applications within learning games. A final suggestion for how LLMs could be employed in support of digital learning games is that they could be used as "helpers" in designing and developing new games. More specifically, LLMs could be used to rapidly generate new game ideas and narratives that game designers could build upon and to provide feedback on game ideas and early prototypes of game developers. Work in this direction has, in fact, already begun (Gatti Junior et al., 2023).

Finally, an important facet of *Decimal Point*—and digital games more generally—that warrants further investigation is the potential presence of unconscious bias embedded in the game's mechanics and artistic elements. The designers of learning games, who are typically and predominantly White—which is true for the designers of *Decimal Point* and cited as 78% the case for the game design industry more generally (Kumar et al., 2022)—are usually well meaning but often unaware of how their own biases frequently lead to design choices that subtly (or even overtly) create biased functionality, blatantly stereotypical game characters and environments, and player identities that turn away children of color, or more specifically, Black children (Peckham, 2020; Rankin & Henderson, 2021; Richard, 2017). In fact, *Decimal Point* has provided at least preliminary evidence of implicit bias. In a recent analysis of more than 700 students using the game, spanning three of the classroom studies reported in this chapter (i.e., Studies 5, 6, and 6a) and a new study, we found that well-represented students (White and Asian; $n = 578$) showed more engagement and less anxiety in using *Decimal Point* than under-represented students (Black, Hispanic or Latino, Indigenous, and multiracial; $n = 158$) (Ni et al., 2024). Unpacking potential biases is crucial for a nuanced understanding of how learning games can be designed and redesigned to support diverse learners. To address this, we recently proposed a project to the National Science Foundation in which we will engage 120 Black middle school students in co-design sessions with *Decimal Point* and in the analysis of 10 other STEM learning games, including *Math Blaster*, *Math Playground,* and *BrainPop.* By scrutinizing these games, and redesigning *Decimal Point* if and where necessary, we could contribute to the ongoing discourse on diversity, equity, and inclusion in digital learning games, paving the way for more informed and culturally sensitive game design practices.

The McLearn Lab's ten-year research program with *Decimal Point* has been thrilling, with some prominent successes, such as the gender effect and the self-explanation findings, but also some disappointing failures, such as the lack of impact of agency and mindfulness inductions in the context of the game. *Decimal Point* as a research platform has facilitated much of the work described in this chapter. The McLearn Lab looks forward to continuing this line of research not only with *Decimal Point*, but with two new games the lab has designed and developed: *Angle Jungle*, a game to help elementary and middle school students learn about angles that was reimplemented and extended for classroom use from a prior implementation (Khan et al., 2017) and *Ocean Adventure*, a game that is a "reskinning " of *Decimal Point* with precisely the same content and instructional approach but with a completely different narrative and art assets. The future possibilities of learning game design, development, and research are myriad, and we intend to pursue many of these possibilities with our various learning games.

Acknowledgements I have been fortunate to have had many incredible collaborators throughout the years. This work has been greatly influenced and guided by conversations and collaborations with many of these colleagues and students, including Ryan Baker, Rosta Farzan, Jodi Forlizzi, Jessica Hammer, Erik Harpstead, Xinying Hou, Richard E. Mayer, Michael Mogessie, Huy Nguyen, Jaclyn Ocumpaugh, Eleanor O'Rourke, Liz Richey, Jon Star, and Yeyu Wang. Thank you to Ilona Buchem, Imogene Casebourn, Huy Nguyen, Liz Richey, and Rupert Wegerif for reviewing this chapter. I would especially like to thank Huy Nguyen and Liz Richey who have been my closest and most significant collaborators throughout the many years of this research. ChatGPT 4.0 was used both to draft and to revise some paragraphs of the chapter, as well as to collect relevant information about learning games and past research. The work has been supported by three National Science Foundation grants (NSF Award #s 1238619, 1661121, and 2201796). The opinions expressed are those of the author and do not represent the views of NSF.

References

Adams, D., McLaren B.M., Durkin, K., Mayer, R.E., Rittle-Johnson, B., Isotani, S., & Van Velsen, M. (2012). Erroneous examples versus problem solving: Can we improve how middle school students learn decimals? In *Proceedings of the 34th Meeting of the Cognitive Science Society (CogSci 2012).* Sapporo, Japan: Cognitive Science Society, pp. 1260–1265.

Adams, D., McLaren, B.M., Mayer, R.E., Goguadze, G., & Isotani, S. (2013). Erroneous examples as desirable difficulty. In Lane, H.C., Yacef, K., Mostow, J., & Pavlik, P. (Eds.). *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED 2013).* LNCS 7926. Springer, Berlin, pp. 803–806.

Adams, R.B., Barber, B.M., & Odean, T. (2019). The math gender gap and women's career outcomes. Available at SSRN 2933241.

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

Ahmadi, A., Noetel, M., Parker, P., Ryan, R. M., Ntoumanis, N., Reeve, J., Beauchamp, M., Dicke, T., Yeung, A., Ahmadi, M., Bartholomew, K., Chiu, T., Curran, T., Erturan, G., Flunger, B., Frederick, C., Froiland, J. M., González-Cutre, D., Haerens, L., … Lonsdale, C. (2023). A classification system for teachers' motivational behaviors recommended in self-determination theory interventions. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000783

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction, 17*(3), 286–303.

Aleven, V.A.W.M.M., Koedinger, K. R., & Popescu, O. (2003). A tutorial dialog system to support self-explanation: Evaluation and open questions. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, pp. 39–46.

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147–179. https://doi.org/10.1016/S0364-0213(02)00061-7

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105–154. https://doi.org/10.1109/DIGITEL.2010.55

Aleven, V., McLaren, B. M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., & Koedinger, K. R. (2016). Example-tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education, 26*(1), 224–269. https://doi.org/10.1007/s40593-015-0088-2

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 522–560). Routledge.

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action, assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*(4), 224–233. https://doi.org/10.1080/00461520.2010.517740

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214.

Baker, R.S., Richey, J.E., Zhang, J., Karumbaiah, S., Andres-Bray, J.M., Nguyen, H.A., Andres, J.M.A.L., & McLaren, B.M. (2024). Gaming the system mediates the relationship between gender and learning outcomes in a digital learning game. *Instructional Science*. DOI: 10.1007/s11251-024-09679-3

Baker, R., Scruggs, R., Pavlik, P. I., McLaren, B. M., & Liu, Z. (2023). How well do contemporary knowledge tracing algorithms predict the knowledge carried out of a digital learning game? *Educational Technology Research & Development*. https://doi.org/10.1007/s11423-023-10218-z

Bard, G.V. (2007), Spelling-error tolerant, order-independent pass-phrases via the Damerau–Levenshtein string-edit distance metric, *Proceedings of the Fifth Australasian Symposium on ACSW Frontiers : 2007, Ballarat, Australia, January 30 - February 2, 2007*, Conferences in Research and Practice in Information Technology, vol. 68, Darlinghurst, Australia: Australian Computer Society, Inc., pp. 117–124, ISBN 978-1-920682-49-1.

Bauckhage, C. (2015). Numpy/scipy recipes for data science: k-medoids clustering. https://doi.org/10.13140/2.1.4453.2009.

Benton, L., Mavrikis, M., Vasalou, A., Joye, N., Sumner, E., Herbert, E., Revesz, A., Symvonis, A., & Raftopoulou, C. (2021). Designing for "challenge" in a large-scale adaptive literacy game for primary school children. *British Journal of Educational Technology, 52*, 1862–1880. https://doi.org/10.1111/bjet.13146

Bereczki, E., Takacs, Z. K., Richey, J. E., Nguyen, H., Mogessie, M., & McLaren, B. M. (2024). Mindfulness in a digital math learning game: Insights from two randomized controlled trials. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12971

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education, 101*, 150–167. https://doi.org/10.1016/j.compedu.2016.06.005

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389–391.

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. In: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pp. 41–50.

Bouvier, P., Lavoué, E., Sehaba, K., & George, S. (2013). Identifying learner's engagement in learning games: A qualitative approach based on learner's traces of interaction. In *5th International Conference on Computer Supported Education (CSEDU 2013)*, May 2013, Aachen, Germany, pp. 339–350.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

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

Brown, K. W., West, A. M., Loverich, T. M., & Biegel, G. M. (2011). Assessing adolescent mindfulness: Validation of an adapted mindful attention awareness scale in adolescent normative and psychiatric populations. *Psychological Assessment, 23*(4), 1023–1033. https://doi.org/10.1037/a0021338

Brueckner, L. J. (1928). Analysis of difficulties in decimals. *Elementary School Journal, 29*, 32–41.

Bull, S. (2020). There are open learner models about! *IEEE Transactions on Learning Technologies, 13*(2), 425–448. https://doi.org/10.1109/TLT.2020.2978473

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22*(1), 281–288. https://doi.org/10.3758/s13423-014-0676-4. PMID: 24984923.

Chang, K.-E., Wu, L.-J., Weng, S.-E., & Sung, Y.-T. (2012). Embedding game-based problem-solving phase into problem-posing system for mathematics learning. *Computers & Education, 58*, 775–786.

Cheng, M. T., Rosenheck, L., Lin, C. Y., & Klopfer, E. (2017). Analyzing gameplay data to inform feedback loops in The Radix Endeavor. *Computers & Education, 111*, 60–73.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self explanations: How students study and used examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 25*(4), 471–533.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 398–415.

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.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82*(3), 766–779.

Czikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. Jossey Bass.

Czikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

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

Deci, E. L. (1975). *Intrinsic motivation*. Plenum Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4899-2271-7

Dunning, D. L., Griffiths, K., Kuyken, W., Crane, C., Foulkes, L., Parker, J., & Dalgleish, T. (2019). Research review: The effects of mindfulness-based interventions on cognition and mental health in children and adolescents–a meta-analysis of randomized controlled trials. *Journal of Child Psychology and Psychiatry., 60*, 244–258.

Dunning, D., Tudor, K., Radley, L., Dalrymple, N., Funk, J., Vainre, M., Ford, T., Montero-Marin, J., Kuyken, W., & Dalgleish, T. (2022). Do mindfulness-based programmes improve the cognitive skills, behaviour and mental health of children and adolescents? An updated meta-analysis of randomised controlled trials. *Evidence Based Mental Health, 25*(3), 135–142. https://doi.org/10.1136/ebmental-2022-300464

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206–214. https://doi.org/10.1016/j.learninstruc.2011.11.001

Dwyer, C., & Johnson, L. (1997). Grades, accomplishments, and correlates. In W. Willingham & N. Cole (Eds.), *Gender and Fair Assessment* (pp. 127–156). Erlbaum.

Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly, 37*(3), 293–309.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, Schools, and Inequality*. Westview Press.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education, 67*, 156–167. https://doi.org/10.1016/j.compedu.2013.02.019.

Forlizzi, J., McLaren, B. M., Ganoe, C. H., McLaren, P. B., Kihumba, G., & Lister, K. (2014). Decimal Point: Designing and developing an educational game to teach decimals to middle school students. In Busch, C. (Ed.) *Proceedings of the 8th European Conference on Games Based Learning* (*ECGBL-2014*). Academic Conferences and Publishing International Limited, Reading, U.K, pp. 128–135.

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology, 163*(1), 24–39.

Gatti Junior, W., Marasco, E., Kim, B., Behjat, L., & Eggermont, M. (2023). How ChatGPT can inspire and improve serious board game design. *International Journal of Serious Games, 10*(4), 33–54. https://doi.org/10.17083/ijsg.v10i4.645

Gee, J. P. (2007). *Good video games and good learning Collected essays on video games, learning and literacy*, 2nd Edition. Peter Lang International Academic Publishers. https://doi.org/10.3726/978-1-4539-1162-4.

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

Glasgow, R., Ragan, G., Fields, W. M., Reys, R., & Wasman, D. (2000). The decimal dilemma. *Teaching Children Mathematics, 7*(2), 89–93.

Graeber, A., & Tirosh, D. (1988). Multiplication and division involving decimals: Preservice elementary teachers' performance and beliefs. *Journal of Mathematics Behavior, 7*, 263–280.

Grosse, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*, 612–634. https://doi.org/10.1016/j.learninstruc.2007.09.008

Güldal, Ş., & Satan, A. (2020). The effect of mindfulness-based psychoeducation program on adolescents' character strengths, mindfulness and academic achievement. *Current Psychology*, pp. 1–12.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029

Harpstead, E., Richey, J.E., Nguyen, H., & McLaren, B. M. (2019). Exploring the subtleties of agency and indirect control in digital learning games. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK'19)*, pp. 121–129). ACM. https://doi.org/10.1145/3303772.3303797.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–11.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684–689. https://doi.org/10.1037/0022-0663.91.4.684

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hiebert, J., & Wearne, D. (1985). A model of students' decimal computation procedures. *Cognition and Instruction, 2*, 175–205.

Honey, M.A., & Hilton, M.L. (2011). Learning science through computer games and simulations. The National Academies Press. (http://www.nap.edu/openbook.php?record_id=13078&page=R1).

Hooshyar, D., Malva, L., Yang, Y., Pedaste, M., Wang, M., & Lim, H. (2021). An adaptive educational computer game: Effects on students' knowledge and learning attitude in computational thinking. *Computers in Human Behavior, 114.*

Hou, X., Nguyen, H.A., Richey, J.E., & McLaren, B.M. (2020). Exploring how gender and enjoyment impact learning in a digital learning game. In: Bittencourt, I., Cukurova, M., Muldner, K., Luckin, R., & Millán, E. (Eds.) *Proceedings of the 21st International Conference on Artificial Intelligence in Education. AIED 2020.* Lecture Notes in Computer Science (LNCS, Vol. 12163). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_21.

Hou, X., Nguyen, H. A., Richey, J. E., Harpstead, E., Hammer, J., & 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

Howard, N. R. (2021). "How did i do?": Giving learners effective and affective feedback. *Educational Technology Research and Development, 69*(1), 123–126. https://doi.org/10.1007/s11423-020-09874-2

Hsu, C.-Y., & Tsai, C.-C. (2011). Investigating the impact of integrating self-explanation into an educational game: A pilot study. In Edutainment tech, pp. 250–254.

Huang, X., Zhang, J., & Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: The gender moderating effect. *European Journal of Psychology of Education, 34*(3), 621–640.

Hussein, M. H., Ow, S. H., Elaish, M. M., & Jensen, E. O. (2022). Digital game-based learning in K-12 mathematics education: A systematic literature review. *Education and Information Technologies, 27*(2), 2859–2891. https://doi.org/10.1007/s10639-021-10721-x

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & 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.

Irwin, K. C. (2001). Using everyday knowledge of decimals to enhance understanding. *Journal for Research in Mathematics Education, 32*(4), 399–420.

Isotani, S., McLaren, B. M., & Altman, M. (2010). Towards intelligent tutoring with erroneous examples: A taxonomy of decimal misconceptions. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS-10)*, Lecture Notes in Computer Science, 6094. Berlin: Springer, pp. 346–348.

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.

Johnston, K. (2021). Engagement and immersion in digital play: Supporting young children's digital wellbeing. *International Journal of Environmental Research and Public Health., 18*(19), 10179. https://doi.org/10.3390/ijerph181910179. PMID: 34639481 PMCID: PMC8507672.

Juraschka, R. (2019). How digital game-based learning improves student success. https://www.prodigygame.com/main-en/blog/digital-game-based-learning/

Kafai, Y. B. (1996). Learning design by making games: Children's development of strategies in the creation of a complex computational artifact. In Y. B. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking and learning in a digital world* (pp. 71–96). Erlbaum.

Khan, J., Wang, J., Wang, X., Zhang, Y., Hammer, J., Stevens, S., & 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.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.

Koedinger, K. R., Baker, R. S. J. D., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A Data Repository for the EDM community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. D. Baker (Eds.), *Handbook of Educational Data Mining.* CRC Press.

Kumar, S., Kwan, E., Weststar, J., & Coppins, T. (2022). *International Game Developers Association (IGDA), Developer Satisfaction Survey 2021: Diversity in the Game*

Industry Report. https://igda-website.s3.us-east-2.amazonaws.com/wp-content/uploads/2022/11/15161607/IGDA-DSS-2021-Diversity-Report_Final.pdf.

Landers, R., Armstrong, M., & Collmus, A. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. *Serious Games and Edutainment Applications*. https://doi.org/10.1177/1046878114563660

Landers, R., & Landers, A. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming., 45*(6), 769–785. https://doi.org/10.1177/1046878114563662

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67*(2), vii–147. https://doi.org/10.1111/1540-5834.t01-1-00187

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

Loderer, K., Pekrun, R., & Plass, J. L. (2019). Emotional foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of Game-Based Learning* (pp. 111–151). MIT.

Lomas, D., Patel, K., Forlizzi, J.L., & 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.

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (Eds.) *Intelligent Tutoring Systems. ITS 2014*. Lecture Notes in Computer Science, vol. 8474. Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_47.

Long, Y., & Aleven, V. (2018). Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *Proceedings of CHI*. https://doi.org/10.1145/3057889

Luzón, J. M., & Letón, E. (2015). Use of animated text to improve the learning of basic mathematics. *Computers & Education, 88*, 119–128. https://doi.org/10.1016/j.compedu.2015.04.016

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*, 333–369. https://doi.org/10.1207/s15516709cog0504_2

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, Learning, and Instruction, 3*(1987), 223–253.

Mayer, R.E., & Johnson, C.I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*(3), 241–265. https://journals.sagepub.com/doi/https://doi.org/10.2190/EC.42.3.a.

Mayer, R. E. (2014). Computer games for learning: An evidence-based approach. Cambridge, MA: MIT Press. ISBN: 9780262027571

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

Maynard, B. R., Solis, M. R., Miller, V. L., & Brendel, K. E. (2017). Mindfulness-based interventions for improving cognition, academic achievement, behavior, and socioemotional functioning of primary and secondary school students. *Campbell Systematic Reviews., 13*, 1–144.

McLaren, B.M., Lim, S., & Koedinger, K.R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In Love, B. C., McRae, K. & Sloutsky, V. M. (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society. pp. 2176–2181.

McLaren, B.M., Adams, D., Durkin, K., Goguadze, G. Mayer, R.E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S., & Van Velsen, M. (2012). To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In Ravenscroft, A., Lindstaedt, S., Delgado Kloos, C. & Hernández-Leo, D. (Eds.), *Proceedings of EC-TEL 2012: Seventh European Conference on Technology Enhanced Learning,* LNCS 7563. Springer, Berlin, pp. 222–235.

McLaren, B.M., Timms, T., Weihnacht, D., Brenner, D., Luttgen, K., Grillo-Hill, A., & Brown, D.H. (2014a). A web-based system to support inquiry learning: Towards determining how much assistance students need. In Zvacek, S., Restivo, M.T., Uhomoibhi, J. and Helfert, M. (Eds.) *Proceedings of the Sixth International Conference on Computer-Supported Education (CSEDU-2014).* SCITEPRESS – Science and Technology Publications. Vol. 1, pp. 43–52.

McLaren, B.M., van Gog, T., Ganoe, C., Yaron, D. & Karabinos, M. (2014b) Exploring the assistance dilemma: Comparing instructional support in examples and problems. In Trausan-Matu, S. et al. (Eds.) *Proceedings of the Twelfth International Conference on Intelligent Tutoring Systems* (ITS-2014). LNCS 8474. Springer International Publishing Switzerland, pp. 354–361.

McLaren, B. M., Farzan, R., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017b). 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.). In *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017).* LNAI 10331. Springer: Berlin, pp. 540–543.

McLaren, B.M., Richey, J.E., Nguyen, H.A., & Mogessie, M. (2022c). A digital learning game for mathematics that leads to better learning outcomes for female students: Further evidence. In: *Proceedings of the 16th European Conference on Game Based Learning (ECGBL 2022).* pp. 339–348.

McLaren, B.M., Nguyen, H.A., Richey, J.E., & Mogessie, M. (2022a). 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.

McLaren, B. M., & Nguyen, H. A. (2023). Digital learning games in Artificial Intelligence in Education (AIED): A review. *Handbook of Artificial Intelligence in Education.* Chapter 20.

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017a). A computer-based game that promotes mathematics learning more than a conventional approach. *International Journal of Game-Based Learning (IJGBL), 7*(1), 36–56. https://doi.org/10.4018/IJGBL.2017010103

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011a). Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers & Education, 56*(3), 574–584. https://doi.org/10.1016/j.compedu.2010.09.019

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011b). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human Computer Studies, 69*(1–2), 70–79. https://doi.org/10.1016/j.ijhcs.2010.09.001

McLaren, B. M., Richey, J. E., Nguyen, H., & Hou, X. (2022b). How instructional context can impact learning with educational technology: Lessons from a study with a digital learning game. *Computers & Education.* https://doi.org/10.1016/j.compedu.2021.104366

McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2010). Intelligent tutoring and games (ITaG). In Y. K. Baek (Ed.), *Gaming for classroom-based learning: Digital role-playing as a motivator of study* (pp. 44–65). IGI Global.

Mogessie M., Richey J. E., McLaren B. M., Andres-Bray J. M. L., & Baker R. S. (2020). Confrustion and gaming while learning with erroneous examples in a decimals game. In *Proceedings of the 21st International Conference on Artificial Intelligence in Education. AIED 2020.* Lecture Notes in Computer Science (LNCS, Vol. 12164). Springer, Cham. https://doi.org/10.1007/978-3-030-52240-7_38.

Moyer-Packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E. K., Roxburgh, A. L., & Jordan, K. (2019). How design features in digital math games support learning and mathematics connections. *Computers in Human Behavior, 91*, 316–332.

Nagashima, T., Bartel, A. N., Yadav, G., Tseng, S., Vest, N. A., Silla, E. M., Alibali, M.W., & Aleven, V.A. (2021). Scaffolded self-explanation with visual representations promotes efficient learning in early algebra. *Annual Meeting of the International Society of the Learning Sciences (ISLS 2021).*

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459–496.

Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments, 5*, 135–159.

Nguyen, H., Harpstead, E., Wang, Y., & McLaren, B.M. (2018). Student agency and game-based learning: A study comparing low and high agency. In C. Rosé, R. Martínez-Maldonado, H.U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren and B. du Boulay (Eds.). *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018).* LNAI 10947. Springer: Berlin, pp. 338–351.

Nguyen, H., Wang, Y., Stamper, J., & McLaren, B.M. (2019). Using knowledge component modeling to increase domain understanding in a digital learning game. In *Proceedings of the 12th International Conference on Educational Data Mining* (EDM 2019), pp. 139–148.

Nguyen, H. A., Takacs, Z.K. Bereczki, E., Richey, J. E., & Mogessie, M. & McLaren, B. M. (2022b). Investigating the effects of mindfulness meditation on a digital learning game for mathematics. In: *Proceedings of the 23rd International Conference on Artificial Intelligence in Education (AIED 2022).* pp. 762–767. https://doi.org/10.1007/978-3-031-11644-5_80.

Nguyen, H., Hou, X., Stec, H., Di, S., Stamper, J., & McLaren, B.M. (2023a). Examining the benefits of prompted self-explanation for problem-solving in a decimal learning game. In *Proceedings of 24th International Conference on Artificial Intelligence in Education (AIED 2023).*

Nguyen, H., Else-Quest, N., Richey, J.E., Hammer, J., Di, S., & McLaren, B.M. (2023c). Gender differences in learning game preferences: Results using a multi-dimensional gender framework. In *Proceedings of 24th International Conference on Artificial Intelligence in Education (AIED 2023).* pp. 553–564.

Nguyen, H., Stec, H., Hou, X., Di, S., & McLaren, B.M. (2023b). Evaluating ChatGPT's decimal skills and feedback generation to students' self-explanations in a digital learning game. *Proceedings of Eighteenth European Conference on Technology Enhanced Learning (ECTEL 2023).*

Nguyen, H., Hou, X., Richey, J. E., & McLaren, B. M. (2022a). The impact of gender in learning with games: A consistent effect in a math learning game. *International Journal of Game-Based Learning (IJGBL)., 12*(1), 1–29. https://doi.org/10.4018/IJGBL.309128

Ni, X., Nguyen, H.A., Else-Quest, N., Pagano, A., & McLaren, B.M. (2024). Investigating racial and ethnic differences in learning with a digital game and tutor for decimal numbers. *The Nineteenth European Conference on Technology Enhanced Learning (ECTEL 2024).* Krems, Austria, September 16-20, 2024

Noël, M. P., Grégoire, J., Meert, G., & Seron, X. (2008). The innate schema of natural numbers does not explain historical, cultural, and developmental differences. *Behavioral and Brain Sciences, 31*(6), 664–665.

Nokes, T. J., Hausmann, R. G., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: The case of self-explanation prompts. *Instructional Science, 39*(5), 645–666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology, 83*(1), 44.

O'Neil, H. F., & Perez, R. S. (2008). *Computer games and team and individual learning*. Elsevier.

O'Rourke, E., Ballweber, C., & Popovic, Z. (2014). Hint systems may negatively impact performance in educational games. In *Proceedings of the First Annual ACM Conference on Learning @ Scale (L@S '14)*, pp. 51–60. https://doi.org/10.1145/2556325.2566248.

Ochsenfeld, F. (2016). Preferences, constraints, and the process of sex segregation in college majors: A choice analysis. *Social Science Research, 56*, 117–132.

Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70–76.

Peckham, E. (2020). Confronting racial bias in video games. *Tech Crunch.* Downloaded from https://techcrunch.com/2020/06/21/confronting-racial-bias-in-video-games/.

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. Norton.

PlayToday (2023). Gamer demographics: 2023 Game-changing statistics worth checking https://playtoday.co/blog/stats/gamer-demographics/.

Putt, I. J. (1995). Preservice teachers ordering of decimal numbers: When more is smaller and less is larger! *Focus on Learning Problems in Mathematics, 17*(3), 1–15.

Rankin, Y. A., & Henderson, K. K. (2021). Resisting racism in tech design: Centering the experiences of Black youth. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–32.

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children*, pp. 81–88.

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender achievement gaps in US school districts. *American Educational Research Journal, 56*(6), 2474–2508.

Renkl, A. (2014). Learning from worked examples: How to prepare students for meaningful problem solving. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 118–130). Society for the Teaching of Psychology.

Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: The case of decimal fractions. *Journal for Research in Mathematics Education, 20*(1), 8–27.

Ribner, A. D., Willoughby, M. T., Blair, C. B., & Family Life Project Key Investigators (2017). Executive function buffers the association between early math and later academic skills. Frontiers in Psychology 869 https://doi.org/10.3389/fpsyg.2017.00869

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), July 2007 ISSN 1055–8896 Publisher: Association for the Advancement of Computing in Education (AACE).

Richard, G. T. (2017). Video games, gender, diversity, and learning as cultural practice: Implications for equitable learning and computing participation through games. *Educational Technology*, pp. 36–43.

Richey J. E., Zhang, J., Das, R., Andres-Bray, J. M. Scruggs, R., Mogessie, M., Baker R. S., & 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).*

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin Review, 24*, 1501–1510. https://doi.org/10.3758/s13423-016-1079-5

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In Biswas, G., Bull, S., Kay, J. & Mitrovic, A. (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education (AIED-2011).* Lecture Notes in Computer Science, 6738. Berlin: Springer, pp. 288–295.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 347–363. https://doi.org/10.1007/s11031-006-9051-8

Sackur-Grisvard, C., & Léonard, F. (1985). Intermediate cognitive organizations in the process of learning a mathematical concept: The order of positive decimal numbers. *Cognition and Instruction, 2*, 157–174.

Samuel, T. S., & Warner, J. (2021). "I can math!": Reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Community College Journal of Research and Practice, 45*, 205–222. https://doi.org/10.1080/10668926.2022.2050843

Sawyer, R., Smith, A., Rowe, J., Azevedo, R., Lester, J. (2017). Is more agency better? The impact of student agency on game-based learning. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) AIED 2017. LNCS, vol. 10331, pp. 335–346. Springer, Cham. https://doi.org/10.1007/978-3-319-61425-0 28.

Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science, 38*, 487–502. https://doi.org/10.1007/s11251-009-9114-9

Schell, J. (2008). Story and game structures can be artfully merged with indirect control. *The Art of Game Design: A Book of Lenses*. Taylor & Francis, pp. 317–334.

Schell, J. (2005). Understanding entertainment. *Computers in Entertainment., 3*, 6. https://doi.org/10.1145/1057270.1057284

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press.

Shute, V. J., Rahimi, S., & Smith, G. (2019). Chapter 4: Game-based learning analytics in physics playground. In Tlili, A. & Chang, M. (Eds.), *Data Analytics Approaches in Educational Games and Gamification Systems, Smart Computing and Intelligence*, https://doi.org/10.1007/978-981-32-9335-9_4.

Singh, N. N., Lancioni, G. E., Nabors, L., Myers, R. E., Felver, J. C., & Manikam, R. (2018). Samatha meditation training for students with attention deficit/hyperactivity disorder: Effects on active academic engagement and math performance. *Mindfulness, 9*, 1867–1876.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education, 82*, 378–392.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28.

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. In Heuvel-Panhuizen, M. V. D. (Ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education.* Utrecht: PME., vol. 4, pp. 217–224.

Tahir, F., Mitrovic, A., & Sotardi, V. (2020). Investigating the effects of gamifying SQL-Tutor. In: So, H. J. et al. (Eds.) *Proceedings of the 28th International Conference on Computers in Education.* Asia-Pacific Society for Computers in Education. pp. 416–425, ISBN978-986-97214-5-5.

Takacs, Z. K., & Kassai, R. (2019). The efficacy of different interventions to foster children's executive function skills: A series of meta-analyses. *Psychological Bulletin, 145*, 653.

The NPD Group (2019). Retail tracking service, 2019 Entertainment Survey. https://www.npd.com/news/press-releases/2019/according-to-the-npd-group-73-percent-of-u-s-consumers-play-video-games/

Tobias, S., & Fletcher, J. D. (2011). *Computer games and instruction.* Charlotte NC: Information Age. https://eric.ed.gov/?id=ED529495.

TrueList (2023). 33 Evolutionary Gaming Statistics of 2023. https://truelist.co/blog/gaming-statistics/.

Tsovaltzi, D., Melis, E., & McLaren, B. M. (2012). Erroneous examples: Effects on learning fractions in a web-based setting. *International Journal of Technology Enhanced Learning (IJTEL).* V4 N3/4 2012 pp. 191–230.

Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills in a computer-based instructional simulation game. *Educational Technology Research and Development, 50*, 23–41.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vekety, B., Kassai, R., & Takacs, Z. K. (2022). Mindfulness with children: A content analysis of evidence-based interventions from a developmental perspective. *The Educational and Developmental Psychologist, 39*(2), 231–244. https://doi.org/10.1080/20590776.2022.2081072

Vygotsky, L. S. (1978). In: Cole, M., John-Steiner, V., Scribner, S. & Souberman, E. (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press. ISBN 9780674576292.

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

Walker, E., McLaren, B. M., Rummel, N., & Koedinger, K. R. (2007). Who says three's a crowd? Using a cognitive tutor to support peer tutoring. In Luckin, R. Koedinger, K. R. & Greer, J. (Eds.), In *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (AIED-07), Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work. Amsterdam: IOS Press, pp. 399–406.

Walsh, G. (2009). Wii can do it: Using co-design for creating an instructional game. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pp. 4693–4698.

Wang, Y., Nguyen, H. A., Harpstead, E., Stamper, J. & McLaren, B. M. (2019). How does order of gameplay impact learning and enjoyment in a digital learning game? In: Isotani S., Millán E., Ogan A., Hastings P., McLaren B., Luckin R. (Eds). *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED 2019).* LNAI 11625. Springer, pp. 518–531.

Wang, L. H., Chen, B., Hwang, G. J., Guan, J. Q., & Wang, Y. Q. (2022). Effects of digital game-based STEM education on students' learning achievement: A meta-analysis. *International Journal of STEM Education, 9*, 26. https://doi.org/10.1186/s40594-022-00344-0

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

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22*(4), 393–409.

Wolfram, S. (2023). What is ChatGPT doing... and why does it work? Wolfram Media, Inc. ISBN-13: 978-1-57955-081-3 (paperback).

Woolf, B. P. (2008). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann.

Wouters, P., & van Oostendorp, H. (Eds.). (2017). *Instructional techniques to facilitate learning and motivation of serious games*. Springer.

Wylie, R., & 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 University Press.

Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology, 50*(6), 3119–3137.

Yáñez-Gómez, R., Cascado-Caballero, D., & Sevillano, J. L. (2017). Academic methods for usability evaluation of serious games: A systematic review. *Multimedia Tools and Applications., 76*(4), 5755–5784. https://doi.org/10.1007/s11042-016-3845-9.hdl:11441/74400.ISSN1380-7501.S2CID254833872

Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q., & Huang, X. (2023). A comprehensive capability analysis of GPT-3 and GPT-3.5 series models. arXiv preprint arXiv:2303.10420.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183.