Instructional Science
https://doi.org/10.1007/s11251-024-09679-3

ORIGINAL RESEARCH

Gaming the system mediates the relationship between gender and learning outcomes in a digital learning game

Ryan S. Baker[1] · J. Elizabeth Richey[2] · Jiayi Zhang[1] · Shamya Karumbaiah[3] · Juan Miguel Andres-Bray[1] · Huy Anh Nguyen[3] · Juliana Maria Alexandra L. Andres[1,2,3] · Bruce M. McLaren[3]

Received: 24 July 2022 / Accepted: 31 July 2024

Abstract
Digital learning games have been increasingly adopted in classrooms to facilitate learning and to promote learning outcomes. Contrary to common beliefs, many digital learning games can be more effective for female students than male students in terms of learning and affective outcomes. However, the in-game learning mechanisms that explain these differences remain unclear. In the current study, we re-analyze three retrospective data sets drawn from three studies conducted in different years. These data sets, which involved 213, 197, and 287 students, were collected from a digital learning game that teaches late elementary and middle school students decimal concepts. We re-analyzed these data sets to understand how female and male students differ in the rates of gaming the system, a behavioral measure that reflects a form of disengagement while playing the game. Rates of gaming the system are compared between female and male students within each of the game's two core instructional activities (i.e. problem-solving and self-explanation) as well as tested in a game vs. non-game condition. We found that female students game the system significantly less than male students in the self-explanation step in the game condition, in all three studies. This difference in the rates of gaming mediates the relationship between gender and learning outcomes, a pattern in which female students tend to learn more than male students, across all three studies. These results suggest that future design iterations of the game could focus on reducing gaming behaviors for male students, which might improve learning outcomes for female students as well. Understanding gender-based differences in game behaviors can inform future game design to promote better learning outcomes for all students.

Keywords Digital learning games · Affect detector · Gaming the system · Self-explanation · Gender

Extended author information available on the last page of the article

Introduction

Digital games have emerged as an effective medium to improve student engagement and learning in some learning domains (Clark et al., 2016; Mayer, 2019; Scoresby & Shelton, 2011). Although many empirical studies have reported that learning games are effective overall, it has been noted that relatively few studies have taken a rigorous empirical approach to understanding *why* these games are effective (see discussion in Richey et al., 2021). In other words, what behavioral or cognitive changes do digital learning games promote when compared to non-game instruction, and how do these changes relate to learning outcomes? With the increase in efforts to develop learning games for varied content areas and student populations (e.g., Math—Lomas et al., 2013; Khan et al., 2017a, 2017b; McLaren et al., 2017a; Riconscente, 2013; Science—Cheng et al., 2015, 2017; Harpstead et al., 2013; Lester et al., 2014; Shute et al., 2015, 2021; Computational Thinking—Hooshyar et al., 2021; Rowe et al., 2021; Tahir et al., 2020; Policy argumentation—Easterday et al., 2017; Reading—Jacovina et al., 2016), there is a need to identify the features that make digital games effective for learning and understand why these features are beneficial. Accordingly, a few recent reviews have speculated about which features and designs are most likely to lead to games being effective (e.g., Clark et al., 2016; Mayer, 2019; Wouters & van Oostendorp, 2017). However, further empirical research is needed to deeply understand the mechanisms through which digital learning games promote learning. Developing this understanding can ultimately guide the efforts of game designers in building better learning games, and teachers in selecting when and how to use them. In this paper, we focus on gaming the system—a measure of behavioral disengagement—as a mechanism that may explain differences in learning outcomes between games and other learning activities and between different subgroups of students.

The lack of a comprehensive explanation of how, when, and why games are effective also poses challenges to achieving equitable student outcomes, as design choices are often not informed by cognitive theory or clear empirical evidence regarding the psychological and behavioral effects of those choices. Compounding this issue, many studies have not considered whether digital learning games work in the same ways (and with comparable effectiveness) for different sub-groups of overall student populations. Though some studies have looked at whether games work equally well for different groups (e.g., female vs. male students—Papastergiou, 2009; Chung & Chang, 2017; McLaren et al., 2017b; Tsai, 2017; students of different races—Shin et al., 2012; Kao & Harrell, 2015), this remains a small proportion of the studies on learning in games. Furthermore, as noted by Dele-Ajayi et al. (2018), only a small number of the studies that do check for differences in learning or engagement in terms of student group membership continue on to explicitly investigate *why and how* these differences are seen. It is difficult to change a pattern of lower success for some groups of students, and to design to promote success for all learners, without understanding who is currently less supported, and how games are less successful for those learners. Thus, efforts to understand how digital games work (and how to better design them) must more explicitly investigate not only who benefits from these games, but specifically how these differences manifest—what cognitive and behavioral processes accompany the greater and lesser effectiveness of specific games for specific groups of students? Understanding this can subsequently guide the efforts of game designers in building more effective and more equitable learning games.

To contribute towards answering this question, we investigate previously documented gender differences in the effectiveness of a digital learning game by exploring how female

and male students interact with the game. Within different game activities, we compare students' propensity to game the system, a behavioral measurement of disengagement where a student misuses a learning system's properties to complete the learning activity, as opposed to engaging and learning with the material (Baker et al., 2004b). Gaming the system has been found to be associated with differences in learning outcomes in a variety of studies (Cocea et al., 2009; Fancsali, 2014; Pardos et al., 2014), with differences in student emotional experiences (Baker et al., 2010a, 2010b) and long-term academic and professional outcomes (Almeda & Baker, 2020; San Pedro et al., 2013). One prior study found that different levels of gaming the system explained differences in learning outcomes between a game and non-game control (Richey et al., 2021), with lower levels of gaming the system (i.e., greater behavioral engagement) and better learning outcomes seen in the game condition. This suggests that gaming the system may be a particularly useful behavioral measure for understanding how games affect the engagement of different populations of students differently. This behavior, specific to some types of learning activities, does not represent the full spectrum of disengaged behavior seen across learning activities, but serves as a clear and impactful indicator of disengagement in the games and other learning contexts and activities where it manifests. In this paper, we aim to extend this prior research by examining whether gender differences in learning outcomes might reflect gender differences in engagement, as measured by gaming the system, while students play a digital learning game. Specifically, we compare the difference in gaming the system within different activities between female and male students, and study whether gaming the system in specific activities plays a mediating role in the differences in learning outcomes between female and male students. This represents a step towards understanding the full range of mediating variables that explain these differences. Better understanding how games affect learners' behaviors and affective experiences—and how female and male students respond differently to the same game activity—is a critical step to inform successful game design that better promotes productive learning processes and outcomes for all students.

Background

Digital learning games

Digital learning games are increasingly used in education and there is increasing evidence that they are effective at promoting successful outcomes in mathematics and science (see reviews in Clark et al., 2016; Mayer, 2019). In math learning, for example, Riconscente (2013) studied the use of a tablet game for fractions and found a significant increase (10–15% on average) in students' learning, self-efficacy, and math interest compared to the students in the regular mathematics instruction condition. This suggests that math games may be especially beneficial for student groups with lower levels of self-efficacy (e.g., girls; Louis & Mistele, 2012). In another controlled experiment, Siew et al. (2016) reported a significant increase in algebraic thinking in students playing an android learning game compared to a conventional approach to teaching algebra based on imitation and repetition. Even beyond a general focus on domain content, a wide range of digital learning games have been developed focusing on varied skills and competencies such as creativity (Jackson, 2012), civic engagement (Easterday et al., 2017; Ferguson & Garza, 2011), and visual-spatial abilities and attention (Shute et al., 2015).

One of the key reasons for the uptake of games in education is the potential for them to be more fun and engaging than traditional learning activities. Digital learning games also have been reported to promote motivation in students, with several meta-analyses finding medium, positive effect sizes for digital learning games compared to more traditional instruction (Sitzmann, 2011; Vogel et al., 2006). A meta-analysis conducted by Vogel et al. (2006) compared games and simulated environments to traditional teaching methods and reported significantly better attitudes in students learning from games. A later meta-analysis (Sitzmann, 2011) also found a substantial overall increase in students' self-efficacy when learning with digital games, but the authors also discovered a concerning potential publication bias in this research by identifying sixteen unpublished results. A key limitation within the literature, identified by these meta-analyses, is that many of the identified studies compared digital learning games to conventional instruction rather than other learning technologies, making it difficult to conclude whether the benefits of games come specifically from their game features or more general aspects of technology-supported instruction. It is also likely that the affective benefits of games vary based on game design. In fact, one research synthesis on affect and engagement in technology-supported instruction found that games and other types of learning technology each had examples of very positive and very poor affect and engagement (Rodrigo & Baker, 2011).

There is a wide variation in the design of games, with game features drawn from and inspired by a range of sources, including theories of learning and motivation (Howard-Jones & Demetriou, 2009; Shute et al., 2014). The in-depth study of the features of games for learning dates back more than four decades (i.e., Malone, 1981), with researchers developing taxonomies of game features and using these taxonomies to study how different game features impact students' motivation to play (King et al., 2010; Malone, 1981) and their learning outcomes (Bedwell et al., 2012). Recently, meta-analyses have offered insight into which features of games are beneficial for learning and engagement (Clark et al., 2016; Ke, 2016; Mayer, 2019; Wouters & van Oostendorp, 2017). In one such meta-analysis, intentional learning supports such as explicit training or instruction, cues and feedback, in-game learning tools, and prompts for self-explanation or reflection were found to improve students' learning (Wouters & van Oostendorp, 2017). Another meta-analysis reported a higher success rate in games with more complex game mechanics, a wider variety of potential game actions, and lower degrees of contextualization (Clark et al., 2016).

Motivational theories offer a number of potential explanations for learning benefits from digital games. For example, the four-phase model of interest development suggests that attention-grabbing learning contexts such as digital learning games can trigger situational interest, which in turn may lead to more developed phases of interest if the learner returns to the content over time (Hidi & Renninger, 2006). Heightened situational interest has been associated with greater engagement with the learning content, more connections between new content and prior knowledge, and better learning outcomes (Schraw & Lehman, 2001), providing one potential pathway for digital learning games to produce better learning outcomes compared to non-game learning systems. However, more research is needed to test the motivational pathways through which digital learning games might impact learning outcomes.

Thus far, the majority of research on digital learning games has asked the question of whether they are better for learning and engagement than non-game instruction. While evidence for the overall effectiveness of digital learning games is important to assess the claims for their educational benefits, it is also important to understand the mechanisms that make games effective for some students (Richey et al., 2021). Motivation and engagement are frequently identified as likely mechanisms for explaining the learning benefits

of games, but relatively few studies have tested motivation or engagement as a mediating pathway to learning outcomes attained from digital learning games. Specifically, if digital learning games increase learning by increasing engagement, we should also see behavioral changes in how students interact with the game that reflect those effects. In this paper, we aim to do this by examining gaming the system, a measure of behavioral disengagement based on students' interactions with the game compared to their interactions in a non-game digital control. By comparing rates of gaming the system in the game and a non-game control, we can see whether gender differences in engagement appear only in digital learning games or across both game and non-game digital learning platforms.

Gaming the system

A range of behaviors occur during gameplay and during digital learning in general, with differing impacts on student outcomes. One form of behavior that emerges in a variety of learning systems is gaming the system, defined as "attempting to succeed in an educational task by systematically taking advantage of properties and regularities in the system used to complete that task, rather than by thinking through the material" (Baker et al., 2006a, 2006b). Often construed as a form of behavioral disengagement (e.g. DeFalco et al., 2014), gaming the system is associated both with cognitive and affective processes (Baker et al., 2010a, 2010b) and has been found to have substantial negative relationships with student outcomes (Baker et al., 2004b; Fancsali, 2014; Mogessie et al., 2020; Pardos et al., 2014), even several years after the gaming the system occurs (Almeda & Baker, 2020; San Pedro et al., 2013). The first automated detector that could recognize this behavior directly from student log data was developed in 2004 for the Cognitive Tutor (Baker et al., 2004a). Since then, detectors of gaming the system have been developed for a variety of other learning systems (*ASSISTments*: Pardos et al., 2014; Paquette & Baker, 2019; *Newton's Playground*: Wang et al., 2013; *Decimal Point*: Mogessie et al., 2020).

Students game the system in several different fashions; some of the most frequently reported gaming behaviors include help abuse (i.e., repeatedly and quickly asking for hints or help until the learning system provides the answer) and systematic guessing (such as trying every given value in a problem statement, trying every plausible answer, or counting). These behaviors are more strongly associated with low student learning than off-task behaviors (e.g., talking to a neighbor, surfing the web) that do not lead to systematic misuse of the learning system's features (Cocea et al., 2009; Pardos et al., 2014). Gaming the system has also been reported to be associated with experiencing frustration (Walonoski & Heffernan, 2006a) or boredom (Baker et al., 2010a, 2010b).

Several studies have shown that students' propensity to game the system is influenced considerably by the design features of the learning system they are using. Baker et al. (2009) studied the design features of Cognitive Tutors (Koedinger & Aleven, 2016), a type of intelligent tutoring system, to identify which aspects of design correlated to the different frequencies of gaming behaviors observed in different lessons in the system. They found that lessons that involved concrete materials but with limited engagement-increasing text were gamed more often, and that activities that lacked clarity in the activity or material also tended to be gamed more often. Slater et al. (2016) studied the text and linguistic features of mathematics problems in *ASSISTments* and found that several such features were associated with differences in the frequency of gaming the system. Specifically, they found that the use of complex grammar and the heavy use of pronouns led to higher gaming.

Attempts to improve system design by incorporating interventions to prevent gaming have seen partial success, although at the cost of increased complexity in student interaction. Baker et al. (2006a, 2006b) experimented with the use of supplementary exercises with the content on which students gamed and found improvements in their learning. Visualizations of student gaming behavior and meta-cognitive messages about gaming have also led to lower frequency of this behavior (Arroyo et al., 2007; Roll et al., 2007; Walonoski & Heffernan, 2006b; Xia et al., 2020). Personalizing learning content to a student's personal interests has also been shown to reduce the frequency of gaming (Walkington & Maull, 2011). However, simply making it more difficult to game the system leads to students finding new ways to game the system (Murray & VanLehn, 2005).

Adopting a game-based design might be another way to reduce disengagement, including gaming the system. Though few studies have examined whether digital learning games reduce gaming the system compared to equivalent non-game digital learning systems, this question has been studied in *Decimal Point*, the digital learning game that is the focus of this study. Though gaming the system is associated with worse outcomes in *Decimal Point* (Mogessie et al., 2020), Richey et al. (2021) reported significantly lower levels (around half as much) of gaming the system behavior in *Decimal Point* compared to a more traditional computer-based instructional system covering identical content (non-game). Furthermore, a mediation analysis showed that the better learning seen in students playing the digital game was fully mediated by their lower frequency of gaming the system behavior.

Some studies have suggested that gaming the system is not closely associated with demographic factors, but these studies have only examined a small number of demographic variables. Baker and Gowda (2010) found that the prevalence of gaming the system did not vary based on whether students lived in urban, suburban, or rural areas. In addition, Paquette and Baker (2017) did not find strong evidence that the frequency of gaming the system varied based on urbanicity, race/ethnicity, math and reading proficiency, or economic status. They found that the differences were associated more strongly with learning environments than with student populations. However, research has not yet investigated differences between female and male students in the propensity to game the system. Given the influence of design features on the choice to game the system, and *Decimal Point's* overall effect on the prevalence of gaming, there is some reason to anticipate gender differences in gaming the system within *Decimal Point.* In fact, *Decimal Point* has led to consistently better learning for female students than male students (Hou et al., 2022; McLaren et al., 2017b, 2022b; Nguyen et al., 2022). One possible hypothesis is that this may be because female students are more engaged—and thus may game the system less often—than male students when playing *Decimal Point*. Below, we discuss evidence of gender differences in gameplay experiences and outcomes from digital learning games. In the current paper, we obtained multiple data sets from the *Decimal Point* team, representing data from three studies testing different iterations of *Decimal Point*, and tested this hypothesis across those datasets. We also examined the hypothesis across different components of the game, in particular problem-solving activities and self-explanation activities.

Gender differences in learning games

Research on gender differences in digital learning game outcomes has shown mixed results, with an overall pattern suggesting female students benefit more. Female students have been shown to enjoy learning games more (Adamo-Villani et al., 2008; Chung & Chang, 2017), to be more likely to find a learning game worth playing (Joiner et al., 2011), and to achieve

better learning outcomes (Khan et al., 2017a; Klisch et al., 2012; McLaren et al., 2017b; Nguyen et al., 2022; Tsai, 2017). However, other studies report no gender differences in outcomes (Chang et al., 2014; Clark et al., 2011; Dorji et al., 2015; Manero et al., 2016; Papastergiou, 2009).

The differing effectiveness of learning games for male and female students is sometimes attributed solely to broad differences between genders, such as differences in decision-making processes and the degree of emphasis placed on interpersonal goals versus task-orientation (Koivisto & Hamari, 2014). More generally, prior research has identified gender-based differences in motivation and cognitive strategies (Wolters & Pintrich, 1998). If differences in game behaviors reflect general gender differences, then we would expect to see the same patterns of gender differences in a non-game control; on the other hand, if differences emerge specifically because of the unique features of digital learning games, then we would not expect to see the same differences in a non-game control.

Games may affect emotions and confidence differently across genders, particularly in domains like mathematics in which anxiety and stereotype threat can disproportionately affect female students by reminding them of negative stereotypes and thus consuming available working memory with distracting thoughts (Doyle & Voyer, 2016; Spencer et al., 1999). In this case, a game context might reduce the saliency of math cues and thus free up more working memory space for female students to focus on learning and practicing the academic content of the game. According to this hypothesis, games would not necessarily produce gender differences in engagement or interest in games, and they would tend to benefit female students only in domains in which they experience stereotype threat. Prior work has also found that female students sometimes report lower self-efficacy in certain academic contexts, such as mathematics (Louis & Mistele, 2012). Given that games have been found to increase self-efficacy in math, this might also provide a pathway for games to benefit female students in particular (Riconscente, 2013; Sitzmann, 2011).

Other accounts have suggested that learning differences for female and male student are caused by gender differences in the motivational appeal of learning games and how games are perceived (Ferguson & Olson, 2013; Huang, 2013; Osunde et al., 2018). If groups of students learn less from games because they find them less interesting or engaging, then we also would expect to see differences in how those students play the games, with those behavioral differences mediating the relations between individual characteristics and learning outcomes. However, few studies have tried to analyze learning game behaviors to test whether differences in male and female students' interactions with the game explain the differences in learning outcomes.

Despite being popular among both females and males (Hamari & Keronen, 2017), there are significant gender differences in preferences about digital game features such as avatar characteristics, social interaction, game speed and style (Aleksić & Ivanović, 2017; Chou & Tsai, 2007; Greenberg et al., 2010; Romrell, 2014). Similarly, there are also gender differences in preferences for *learning* games. For instance, female students tend to be more collaborative in games, while male students are more competitive (Dele-Ajayi, 2018; Garber et al., 2017). Female students also tend to prefer playing competitive games alone while male students prefer to play in the company of other male students (Jenson & de Castell, 2005). A recent game preferences survey of 333 middle school students found that girls reported more interest in the casual, music and party, and cooperative genres of games, while boys tended to prefer action, sports and racing, and battle-oriented game genres (Nguyen et al., 2023). Other studies have reported that scores and rewards are more appealing to and valued by male students (Hartmann & Klimmt, 2006; Raney et al., 2006). Furthermore, rewarding speedy play has a more negative impact on female students than male

students (Heeter & Winn, 2008). Given the gender-based differences in game preferences and learning outcomes, there have been discussions around adapting learning game design based on the students' gender to support them better, based on evidence that this can be useful within digital (non-learning) games more broadly (Boyle & Connolly, 2009; Kinzie & Joseph, 2008; Law, 2010; Steiner et al., 2009). The challenge, however, is that although a lot is known about male and female students' preferences in games, considerably less is known about how these preferences translate to differences in gameplay behaviors. Do female and male students engage in the same behaviors? Do they behave differently in the presence of specific gameplay features? We investigate these questions within the current study, with the goal of better informing game design to promote learning for all students.

The digital learning game *Decimal Point*

Decimal Point is a single-player, computer-based game designed for 5th through 7th grade students to learn about decimal numbers, operations, and concepts (McLaren et al., 2017a). The game runs on the Internet, within any standard browser, and was originally developed using Flash and later ported to HTML/JavaScript. The Cognitive Tutor Authoring Tools (or CTAT—Aleven et al., 2016) were used to develop the game, to assess and log student actions. The materials are deployed on the web-based learning management system TutorShop (Aleven et al., 2009), which logs all student actions, such as correct and incorrect steps and hint requests.

The game is set in the thematic context of an amusement park and is composed of a series of 24 mini-games. The mini-games are presented to students in a pre-defined order—at least in the base version of the game and most versions that have been studied—with each mini-game containing two problems for students to solve. In Studies 2 and 3 students were given agency to pick mini-games to play in any order, as explained below. Seventy-two problems were implemented for the game in total. Five types of problems are available in the mini-games, which include (1) ordering decimals; (2) number line placement; (3) completing decimal sequences; (4) sorting decimals into less-than and greater-than "buckets"; and (5) adding decimals. The subject matter and specific content of each problem type was selected because decimal number misconceptions are particularly robust, persisting through middle school and sometimes even into adulthood (Putt, 1995). Each problem type focuses on providing practice opportunities for a specific decimal number operation or concept aligned with specific, well-documented decimal number misconceptions (Isotani et al., 2010). The problems were designed in consultation with a mathematics education expert to specifically target decimal number misconceptions that have been well documented in the math education literature, such as the misconception that longer decimal numbers are larger in magnitude (e.g., $0.234 > 0.9$—Irwin, 2001; Isotani et al., 2010; Stacey et al., 2001).

Every problem is composed of two steps (i.e. problem-solving and self-explanation). Problem-solving and self-explanation activities are distinct but connected in the game, and each can be expected to play specific roles in the learning process (Richey & Nokes-Malach, 2015). Problem-solving practice consists of executing correct procedures to solve various decimal number problems and is essential for skill acquisition, with repeated practice leading to reduced time and greater accuracy on tasks (Singley & Anderson, 1989). Self-explanation occurs when the learner is prompted to explain what they are learning to themselves, which can involve making inferences about why something is right or wrong, developing justifications, or identifying their own lack of understanding or misunderstanding

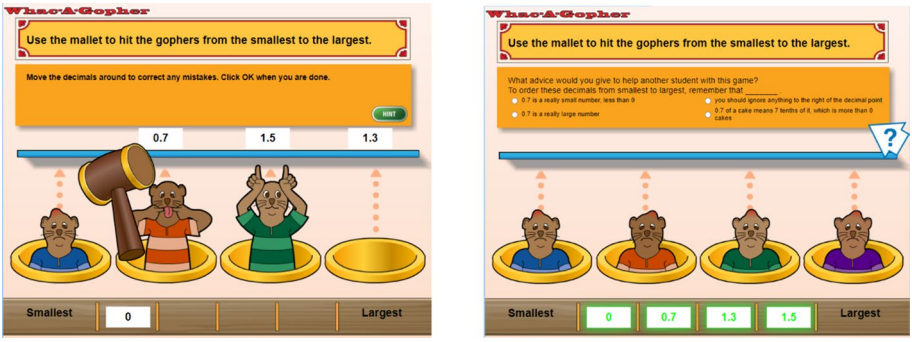

Fig. 1 Whac-a-Gopher, an example of an ordering mini-game, includes a problem-solving step (left) and a self-explanation step (right)

(Chi & Wylie, 2014). Self-explanation helps learners revise errors in prior knowledge, fill in gaps in their understanding, and connect fragmented knowledge, all of which support more robust learning and transfer (see review in Chi & Wylie, 2014; also see McLaren et al., 2022a, McNamara, 2017; Nokes et al., 2011; Richey & Nokes-Malach, 2015; Rittle-Johnson et al., 2017). Particularly when paired with problem-solving practice, self-explanation can make knowledge of problem-solving procedures more flexible by helping learners connect problem-solving steps with principles and application conditions (Ainsworth & Burcham, 2007; Aleven et al., 2003). Self-explanation activities are more common in non-game digital learning technology than in games (Bisra et al., 2018; McLaren et al., 2008; Renkl & Atkinson, 2002; Roy & Chi, 2005), but some evidence suggests that incorporating self-explanation in games may yield similar benefits. In particular, Johnson and Mayer (2010) found that self-explanation led to better learning outcomes from a digital learning game, but only when self-explanation took a multiple-choice format where students were asked to select the correct explanation. They argued that more open-ended forms of self-explanation, in which learners are prompted to type explanations, do not support learning in games because they disrupt the game flow. Similarly, *Decimal Point* incorporates self-explanation prompts in a multiple-choice format.

When students encounter a problem in *Decimal Point*, they start in the problem-solving step and are prompted to solve the problem through game play. After solving the problem, students then move on to the self-explanation step, reflecting on how they derived the answer by selecting from a multiple-choice list of possible explanations. For example, in an ordering decimals problem, students are asked to "hit the gophers from the smallest to the largest" in the problem-solving step (see the left side of Fig. 1). Once they finish solving the problem, they are presented with the self-explanation question (see the right side of Fig. 1). This self-explanation step is designed based on an extensive literature showing that self-explanation promotes deeper learning, and that multiple-choice self-explanation prompts are most effective in game contexts because they are less disruptive to game flow[1] (Bichler et al., 2022; Chi & Wylie, 2014; Johnson & Mayer, 2010; McLaren et al., 2022a; McNamara, 2017; Richey & Nokes-Malach, 2015; Rittle-Johnson et al., 2017).

[1] However, there are also versions of *Decimal Point* in which focused, open-ended and sentence builder self-explanations were studied (McLaren et al., 2022a).

Decimal Point incorporates elements of fantasy (Malone, 1981) through the amusement park context and through six alien characters who accompany the students throughout the game. The alien characters playfully incorporate accuracy feedback when students provide correct or incorrect responses, as well as providing encouragement throughout game play. Feedback is immediately provided after each step, and students must correctly answer both the problem solving and self-explanation steps in order to advance to the next problem. There are no penalties or limits on attempts for incorrect responses.

The present set of studies

The present set of studies utilizes past data from the use of a digital learning game, *Decimal Point*, obtained from the *Decimal Point* team. In this paper, we seek to answer the following research questions.

1. Do female and male students differ in how they interact with a digital learning game? Specifically, do they differ in the rates of gaming the system (a measure of behavioral disengagement) in a digital learning game as compared to a non-game control?
2. Do differences in gaming the system between female and male students occur in a specific activity (i.e., problem solving, self-explanation) within the game?
3. Do the differences in gaming the system in these specific contexts explain differences in learning outcomes between female and male students?
4. Do female and male students differ in their self-efficacy or interest in the game, and if so, does controlling for these differences eliminate mediating effects of gaming the system?

By comparing the frequency of gaming between female and male students in each step of the game (i.e., the problem solving and self-explanation steps), within the digital learning game *Decimal Point* and a non-game equivalent, we can investigate where and when differences manifest in this form of engagement between female and male students. We then further investigate whether gaming the system mediates and explains the relationships between gender and learning outcomes. Such mediation models can illuminate the specific learning processes and outcomes for different students playing *Decimal Point*, which in turn can inform instructional design to better support optimal learning processes.

To understand how female and male students differ in how often they game the system and the impact of this form of engagement on learning, we reanalyzed interaction and outcome data from three studies where students used *Decimal Point*. Each dataset was obtained from the *Decimal Point* team and contained pretest scores, immediate posttest scores, delayed posttest scores, and log data capturing students' interaction with the digital learning game or non-game tutor. We describe each dataset as a separate study below.

Study 1 method

Study 1 utilized a dataset collected in the fall semester of the 2015 school year. The original study investigated the benefit of computer-based games in digital learning and results were first reported in McLaren et al. (2017a). In this experiment, students were assigned to use either the *Decimal Point* game or a non-game tutor with equivalent problem content. This dataset allowed us to examine the differences in the proportion

Table 1 Participant demographic information across studies

Study	Initial sample	Final sample size	Age *M* (*SD*)
Fall 2015	213	153 (66 male, 87 female)	11.3 (0.52)
Fall 2017	197	165 (85 male, 80 female)	11.2 (0.60)
Spring 2018	287	237 (107 male, 130 female)	11.9 (0.47)

of gaming between female and male students in both the game and non-game conditions. This allowed us to draw conclusions about the degree to which the game changes engagement compared to a non-game control, and whether male and female students engage differently in the game compared to a non-game tutor. We also examined the effect of engagement on learning outcomes and, given prior findings of gender differences in learning outcomes (Nguyen et al., 2022), we investigated whether levels of engagement mediate the relationship between gender and learning outcomes.

Study 1 participants

In this dataset, 213 students at two middle schools in a northeastern U.S. metropolitan area used either *Decimal Point* or the non-game equivalent as part of their normal classroom math instruction. Because of the distraction and demotivation that might have occurred with students sitting next to one another but working with very different materials, the researchers assigned students by classroom to one of the two instructional conditions; teachers classified each class as a low-, medium-, or high-performing class, and classes were equally distributed based on these ratings across the two conditions. Students who did not complete the materials in time or had an incomplete pretest, posttest, or delayed posttest (N = 52) were excluded from the analysis. An additional 8 students were removed for having gain scores that were more than 2.5 standard deviations above or below the mean. Of the remaining 153 students, 70 students were assigned to play *Decimal Point*, while 83 students completed the non-game equivalent of the system covering the same content. Both conditions had similar proportions of male and female students. Specifically, 31 male and 39 female students were in the game condition, and 35 male and 48 female students were in the non-game condition. Demographic information about participants in each study is reported in Table 1.

Study 1 materials and procedure

Study 1 compared students learning in *Decimal Point* to a non-game control that presented identical learning and test problems, problem-solving mechanics, self-explanation prompts, and accuracy feedback. Figure 2 shows the equivalent non-game item as the Whac-A-Gopher problem in Fig. 1. The cover stories for the learning problems differed in the non-game context to avoid having a consistent theme. All problems in the non-game tutor were presented on a plain screen without characters.

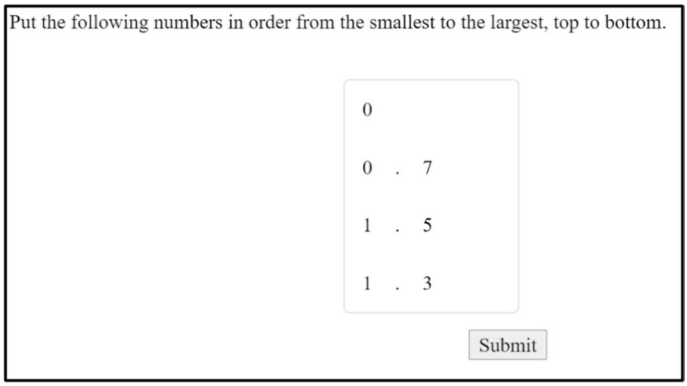

Fig. 2 The non-game equivalent of the same ordering mini-game shown in Fig. 1

Knowledge tests

Three isomorphic tests were designed to target students' learning of the decimal number operations practiced during the game and non-game tutor, as well as the underlying concepts and decimal number misconceptions addressed through the game and tutor. Tests were counterbalanced and administered as a pretest for students to take immediately before the beginning of the game or tutor. The pretest was used to assess prior knowledge; a posttest administered immediately after the end of the game or tutor was used to assess knowledge after completing the learning materials; and a delayed posttest administered one week after the end of the game or tutor was used to assess knowledge retention. Each test consisted of 42 items worth a total of 52 points, as some test items were worth multiple points.

Gaming detector construction

Models were developed to recognize gaming the system within the interaction data from *Decimal Point* and its non-game comparison condition, by first hand-labeling a subset of the *Decimal Point* data in terms of whether it involved gaming behavior, and then using machine learning to develop "detectors" that replicate those human judgments at scale (Baker et al., 2006a, 2006b).

The hand labels were obtained through text replay coding. Text replay coding has been used in many past studies, producing labels with acceptable inter-rater reliability, and in turn being used to develop automated detectors that are successful at recognizing when gaming the system is occurring (Baker & de Carvalho, 2008; Baker et al., 2006a, 2006b, 2010a, 2010b; Paquette & Baker, 2019). In text replay coding, human coders read through a clip of log data that captures a student's interaction with the learning environment, and then use their judgment to infer the learner's behaviors at the time. In the current study, we used text replay coding to identify gaming the system within the log data obtained from *Decimal Point* and the non-game tutor.

To develop text replays for coding, the research team first breaks down log data into clips, sub-segments of student behavior within the system. In general, each clip can capture a specific amount of time, a specific number of student actions, or a specific segment of an

activity. In this study, in order to understand how engagement differs in each step (whether a problem-solving step or a self-explanation step) in male and female students, we delineated clips by treating each step as its own clip. Two iterations of text replay coding were conducted. The first iteration of the text replay coding labeled the gaming behaviors in the problem-solving steps while the second iteration labeled the gaming behaviors in the self-explanation steps.

In each iteration, text replay coding was conducted in three phases. In phase 1, two human coders coded a set of clips together. By discussing their judgment and the behavioral patterns noticed in the clips, the coders established a labeling rubric (this rubric was also based on the extensive past work that has been published on understanding gaming the system—cf. Paquette et al., 2014). The rubric contains a set of behavioral patterns that indicate the student is gaming the system. Specifically, within *Decimal Point*, behaviors that were identified as gaming the system included:

- Clicking through the hints at high speed to obtain the answer, then immediately entering the answer and moving on
- Systematically and rapidly guessing numbers (i.e., 1, 2, 3, …)
- Systematically and rapidly selecting each multiple-choice option (i.e., A, B, C, …)

In phase 2, the two coders coded another set of clips separately using the rubric established. The labels from each coder were then compared and used to compute the inter-rater reliability (Cohen's Kappa). If the inter-rater reliability had been below criterion, the coders would have discussed the differences and repeated the phase 2 coding until an acceptable inter-rater reliability had been achieved before moving on to phase 3. However, in this specific case, the two coders achieved acceptably high inter-rater reliability (by the typical standards of data used as the basis for machine learning of this nature) on the first round of phase 2 coding for both the problem-solving ($k = 0.74$) and self-explanation ($k = 0.88$) steps. In phase 3, coders split the remaining clips and coded independently. Clips were stratified to equally represent schools, problem type, and experiment condition. In total, 800 problem-solving clips and 1500 self-explanation clips were coded and used to construct the automated gaming detectors. More self-explanation clips were coded than problem-solving clips, because the first 800 self-explanation clips only had a small number of positive cases for the algorithm to learn from.

To create automated gaming detectors, the labeled data was input into machine learning algorithms to replicate the coders' judgment. This approach has been used to detect gaming the system in prior published studies (see, for example, Baker & de Carvalho, 2008; Baker et al., 2010b; Paquette & Baker, 2019). After evaluating the performance of several classification algorithms on this data, an Extreme Gradient Boosting (XGBoost) classifier (Chen & Guestrin, 2016) was used to build the automated detector for each of the two types of steps, classifying whether a clip capturing either the problem-solving or self-explanation step contains a gaming behavior. XGBoost uses an ensemble technique that trains an initial, weak decision tree and calculates its prediction errors. Following the initial training, the classifier then trains subsequent trees iteratively to predict the errors in the previous trees. The final prediction represents the sum of the predictions of all the trees in the set.

The models were tested with tenfold student-level cross-validation, in which models were trained using data from a subset of students and tested on other students' data. Based on the cross-validation results, we determined that the models could reliably predict gaming in unseen students in both the problem-solving ($AUC = 0.89$, $k = 0.50$) and self-explanation ($AUC = 0.99$, $k = 0.95$) steps. The detectors were then applied to the rest of

Table 2 Correlations between test performance, gender (female = 0, male = 1), and gaming the system for Study 1

	Posttest	Delayed posttest	Gender	Gaming (PS)	Gaming (SE)
Pretest	0.78*	0.74*	0.07	−0.67*	−0.42*
Posttest		0.79*	0.02	−0.64*	−0.46*
Delayed Posttest			−0.02	−0.63*	−0.48*
Gender				−0.04	0.26
Gaming (PS)					0.57*

*$p < 0.05$

the dataset to predict gaming. We computed the rate of gaming for each student and step using the gaming labels from the detectors. The rate of gaming reflects how often a student gamed the system at either the problem-solving or the self-explanation steps.

Study 1 results

First, rates of gaming were computed for each student and step to reflect how often each student gamed the system on the problem-solving and self-explanation steps. Correlations between test scores, gender, and rates of gaming are shown in Table 2. For both the problem-solving and self-explanation steps, gaming the system was significantly, negatively correlated with test performance.

We then compared rates of gaming between female and male students on the problem-solving and self-explanation steps in both the game and non-game conditions. Means and standard deviations for test scores and rates of gaming (for female and male students, in the two conditions, across types of activities) are shown in Table 3. Analysis of variance (ANOVA) was performed to assess whether the students' rates of gaming on each step differed by gender. ANOVA was selected rather than a non-parametric test, due to lack of evidence of non-normality in the variables (skewness and kurtosis were in the acceptable range for all variables). In the non-game condition, no significant difference between male and female students' gaming the system behaviors was observed in either the problem-solving ($F(1,81) = 0.17$, $p = 0.68$) or self-explanation steps ($F(1,81) = 0.07$, $p = 0.79$). This suggests that students engaged similarly with the non-game tutor regardless of gender. However, in the game condition, male students gamed the system significantly more than female students within the self-explanation steps ($F(1,68) = 4.83$, $p = 0.031$), indicating that male students demonstrated more disengagement than female students in the game. There was no significant difference on the problem-solving step ($F(1,68) = 0.096$, $p = 0.76$).

The lack of differences in the frequency of gaming the system between female and male students in the non-game condition suggests that some aspect of playing the game triggered differences in gaming the system behaviors. To understand how gaming the system in the self-explanation step relates to learning outcomes for female students and male students playing the game, linear regression models were used to predict the immediate and delayed posttest scores by rates of gaming in the self-explanation steps, controlling for pre-test scores (Table 4). Gaming the system in the self-explanation step did not significantly predict students' scores on the immediate posttest when controlling for pretest. However, both pretest and gaming the system in the self-explanation step significantly predicted students'

Table 3 Average rates of gaming the system for problem solving (PS) and self-explanation (SE) steps and test scores by condition and gender for Study 1

	N	Gaming (PS) *M* (*SD*)	Gaming (SE) *M* (*SD*)	Prestest *M* (*SD*)	Immed. Posttest *M* (*SD*)	Delayed Posttest *M* (*SD*)
Game condition						
Female	39	0.13 (0.08)	0.19 (0.11)	31.33 (10.18)	38.64 (8.98)	40.77 (8.34)
Male	31	0.12 (0.09)	0.24 (0.11)	32.77 (11.98)	39.03 (10.39)	40.32 (10.60)
Non-game condition						
Female	48	0.28 (0.13)	0.31 (0.15)	25.96 (9.54)	29.52 (10.47)	31.23 (10.82)
Male	35	0.27 (0.14)	0.32 (0.16)	26.03 (10.02)	30.89 (10.77)	32.86 (11.80)

Table 4 Regression models predicting immediate and delayed posttest with pretest scores and rates of gaming for students in the game condition. Beta and p values are from combined models

	Immediate posttest	Delayed posttest
Overall model	$R^2 = 0.61$, $F(2,67) = 56.87$, $p < 0.001$	$R^2 = 0.58$, $F(2,67) = 46.83$, $p < 0.001$
Pretest	$\beta = 0.71$, $p < 0.001$	$\beta = 0.65$, $p < 0.001$
Gaming (SE)	$\beta = -0.16$, $p = 0.055$	$\beta = -0.21$, $p = 0.022$

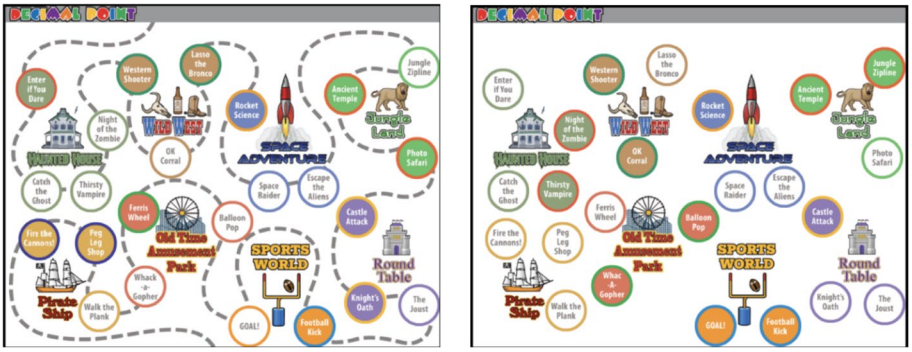

Fig. 3 The mediation model showing path standardized coefficients for a mediation analysis of gender on delayed posttest through gaming the system on self-explanation questions

scores on the delayed posttest, with both predictors statistically significant (see Table 4). A model with just pretest scores predicted 55% of the variance in posttest scores, and the overall model predicted 58% of the variance, indicating that adding the rate of gaming the system on the self-explanation step predicted an additional 3 percent of the variance over the pretest alone.

In both models predicting students' test scores, gaming in the self-explanation steps was negatively associated with the posttest performance (according to the β coefficients), suggesting that when controlling for pretest, higher rates of gaming were associated with worse learning outcomes.

Finally, we examined whether the difference in gaming the system in the self-explanation step explained any effect of gender on learning outcomes on the delayed posttest. Given the fact that gaming the system behaviors differ between genders and gaming the system significantly predicts learning outcomes on the delayed posttest, we created a mediation model (Hayes, 2017) examining the relationship between gender and delayed posttest scores with gaming in the self-explanation step as the mediator (Fig. 3). The "mediate" function in the "psych" package in R was used to build each model, using 5000 bootstrap iterations. This model generates confidence intervals to test the indirect effect of gender (female = 0, male = 1) on delayed posttest scores, with gaming the system on the self-explanation step as the mediator. Pretest scores were again included as a covariate. We focused on delayed posttest scores because gaming on the self-explanation step was a significant predictor of delayed posttest score.

Results indicated that male students had a significantly higher frequency of gaming the system in the self-explanation step, $a = 0.06$, $p < 0.008$. Gaming the system on the self-explanation step was negatively associated with performance on the delayed posttest regardless of gender, $b = -16.7$, $p = 0.03$. There was no direct effect of gender on delayed

posttest performance when controlling for gaming the system, $c' = -0.30$, $p = 0.85$, but the indirect effect of gender on posttest through gaming the system on the self-explanation step was significantly different from zero, $ab = -1.07$, 95% CI [−2.71, −0.06]. This indicates that gaming the system mediates the effect of gender on delayed posttest scores.

Study 2 method

Results from Study 1 showed that male and female students engaged differently with the game but not the non-game tutor, and that gaming the system on the self-explanation step mediated the effect of gender on the delayed posttest. In Studies 2 and 3, we examined additional *Decimal Point* datasets to see whether these effects would replicate across studies. Study 2 utilized a dataset collected in the 2017 fall semester, and results were originally reported in Nguyen et al., (2018). The original purpose of Study 2 was to investigate the effect of agency in digital learning games. Specifically, the study examined whether enabling students to choose which mini-games to play and when to quit would lead to different behaviors and learning outcomes.

Study 2 participants

In the Study 2 dataset, 197 students at one of the same middle schools from the Study 1 dataset and at an elementary school in the same northeastern U.S. metropolitan area used *Decimal Point* as part of their normal math instruction. Students who did not complete the pretest, posttest or delayed posttest were excluded from the analysis (N = 32). Seven additional students were excluded as outliers because their gains from pretest to posttest or delayed posttest were more than 2.5 standard deviations greater or less than the mean. Of the remaining 158 students, 77 were female and 81 were male. Additional demographic information about the participants in Study 2 is reported in Table 1.

Study 2 materials and procedure

Students in the Study 2 dataset used either the original version of *Decimal Point* (low-agency condition) or a version of the game in which they could select the order in which they played the mini-games and could choose to quit at any point after completing 24 rounds of mini-games (Nguyen et al., 2018). All problem content and within-game mechanics were the same across conditions, and we analyze the two conditions together in this paper. This choice enables us to focus on investigating the impact of gender on gameplay behaviors and learning outcomes, regardless of what order the games were played in. Study 2 also introduced questionnaires asking students to self-report their interest in the game and self-efficacy for decimal number operations.

Knowledge tests

The same three isomorphic tests designed for Study 1 were used to assess knowledge in Study 2. As in Study 1, tests were counterbalanced and administered as a pretest immediately before the beginning of the game to assess prior knowledge; a posttest administered immediately after the end of the game to assess knowledge after completing the game; and

Table 5 Correlations between test performance, gender (female = 0, male = 1), gaming the system, decimal self-efficacy, and interest in the game for Study 2

	Posttest	Delayed posttest	Gender	Gaming (PS)	Gaming (SE)	Self-efficacy	Interest in game
Pretest	0.86*	0.86*	0.15	−0.74*	−0.56*	−0.61*	−0.03
Posttest		0.87*	0.10	−0.70*	−0.56*	−0.52*	−0.11
Delayed Posttest			0.08	−0.66*	−0.52*	−0.50*	−0.12
Gender				−0.13	0.17*	−0.22*	−0.10
Gaming (PS)					0.48*	0.53*	0.11
Gaming (SE)						0.30*	0.11
Self-efficacy							0.08

*$p < 0.05$

a delayed posttest administered one week after the end of the game to assess knowledge retention.

Gaming detectors

The same models developed to recognize gaming the system within the interaction data from Study 1 were applied to interaction data to detect gaming in Study 2.

Self-efficacy and interest surveys

Study 2 added several additional measures of students' affective experiences: a questionnaire assessing student self-efficacy, and a questionnaire assessing interest. Self-efficacy items were administered before the start of the game (five items, $a = 0.79$). Students responded to statements such as "I do well on decimal problems in school" and "Before this lesson, I understood decimals (such as 0.235)". After completing the game, students responded to three items about their interest in the game ($a = 0.86$). Example statements included "I liked doing this lesson" and "I would like to do more lessons like this." For both questionnaires, students responded on a 1–5 Likert-type scale from 1 (strongly disagree) to 5 (strongly agree).

Study 2 results

First, rates of gaming were computed for each student and step to reflect how often each student gamed the system on the problem-solving and self-explanation steps. Correlations between test scores, gender, and rates of gaming are shown in Table 5. Similar to Study 1, there were strong, negative correlations between test performance and gaming the system; surprisingly, there were also negative correlations between test performance and decimal self-efficacy.

We then compared rates of gaming between female and male students on the problem-solving and self-explanation steps in the game. Means and standard deviations for test scores, self-efficacy, interest, and rates of gaming (for female and male students, across

Table 6 Average probabilities of gaming the system for problem solving (PS) and self-explanation (SE) activities, and average self-reported decimal self-efficacy and interest

	N	Gaming (PS) *M* (*SD*)	Gaming (SE) *M* (*SD*)	Self-efficacy *M* (*SD*)	Interest *M* (*SD*)	Pretest *M* (*SD*)	Imm. Posttest *M* (*SD*)	Delayed posttest
Female	77	0.27 (0.20)	0.27 (0.24)	2.52 (0.90)	2.16 (0.92)	34.06 (11.94)	40.40 (9.92)	41.48 (10.76)
Male	81	0.22 (0.17)	0.36 (0.28)	2.16 (0.75)	1.97 (0.98)	37.68 (12.42)	42.36 (10.03)	43.27 (10.68)

Table 7 Regression models predicting immediate and delayed posttest with pretest scores and rates of gaming. Beta and p values are from combined models

	Immediate posttest	Delayed posttest
Overall model	$R^2 = 0.75$, $F(2,155) = 226.47$, $p < 0.001$	$R^2 = 0.75$, $F(2,155) = 229.73$, $p < 0.001$
Pretest	$\beta = 0.79$, $p < 0.001$	$\beta = 0.83$, $p < 0.001$
Gaming (SE)	$\beta = -0.12$, $p = 0.015$	$\beta = -0.06$, $p = 0.22$

types of activities) are shown in Table 6. A one-way ANOVA revealed no significant difference in pretest performance between male and female students, $F(1,156) = 3.47$, $p = 0.064$. One-way ANCOVAs controlling for pretest revealed no significant effect of gender on the immediate posttest, $F(1,155) = 0.48$, $p = 0.49$, or on the delayed posttest, $F(1,155) = 1.19$, $p = 0.28$. As in Study 1, male students gamed the system significantly more than female students within the self-explanation steps $F(1,156) = 4.82$, $p = 0.030$. As in Study 1, there was not a statistically significant difference in gaming the system on the problem-solving step, $F(1,156) = 2.67$, $p = 0.10$. Prior research has indicated that apparent gender differences in math or digital learning game outcomes could result from differences in learners' self-efficacy or interest in the content. To examine this possibility, we assessed gender differences in self-reported decimal self-efficacy and interest in the *Decimal Point* digital learning game. Contrary to prior research, female students reported significantly higher decimal self-efficacy than male students, $F(1,156) = 7.63$, $p = 0.006$. However. there were no significant gender differences in students' interest in the game, $F(1,156) = 1.50$, $p = 0.22$.

As with Study 1, regression models were used to predict the immediate and delayed posttest scores by rates of gaming in the self-explanation steps, controlling for pretest scores. Both pretest and gaming the system in the self-explanation step significantly predicted students' scores on the immediate posttest, with both predictors statistically significant (see Table 7). A model with just pretest scores predicted 74 percent of the variance in posttest scores, and the overall model predicted 75 percent of the variance, indicating that adding the rate of gaming the system on the self-explanation step predicted an additional 1 percent of the variance over the pretest alone. Despite the small amount of additional variance explained by gaming the system, this predictor remained statistically significant in a combined model, $t(155) = -2.46$, $p = 0.015$.

However, while pretest significantly predicted students' scores on the delayed posttest, gaming the system in the self-explanation step did not significantly predict students' delayed posttest scores in Study 2, $t(155) = -1.22$, $p = 0.22$.

Since gender did not have a significant effect on test performance, we applied a bootstrap mediation analysis that does not require the predictor variable to significantly predict the outcome variable (Hayes, 2017). This mediation approach can detect significant indirect pathways even when the direct pathway is not significant. We built a mediation model to test the indirect effect of gender (female = 0, male = 1) on posttest scores, with gaming the system on the self-explanation step as the mediator. Pretest scores were again included as a covariate. We focused on posttest scores because gaming the system on the self-explanation step was a significant predictor of posttest score in Study 2.

Results indicated that male students gamed the system significantly more often than female students in the self-explanation step, $a = 0.14$, $p < 0.001$, and gaming was found to be significantly, negatively associated with immediate posttest scores, $b = -4.60$, $p = 0.02$. The rate of gaming in the self-explanation steps was shown to explain the relationship between gender and the immediate posttest scores. There was no direct effect of gender

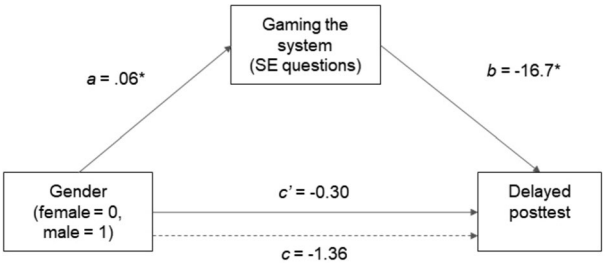

Fig. 4 The mediation model showing path standardized coefficients for a mediation analysis of gender on posttest through gaming the system on self-explanation questions, in Study 2

on posttest performance when controlling for gaming the system, $c' = 0.05$, $p = 0.95$, but the indirect effect was statistically significantly different from zero, $ab = -0.63$, 95% CI [−1.45, −0.06] (see Fig. 4). Similar to Study 1, gaming the system mediated the relation between gender and test performance.

To account for the possibility that gender differences in self-efficacy might contribute to gender differences in gaming the system or learning outcomes, we re-ran the mediation model predicting immediate posttest, this time including decimal self-efficacy as an additional covariate. The overall results of the mediation model did not change. Male students were again found to have gamed the system significantly more often than female students in the self-explanation step, $a = 0.14$, $p < 0.001$, and gaming was found to be significantly negatively associated with the immediate posttest scores, $b = -4.60$, $p = 0.02$. The rate of gaming in the self-explanation steps still mediated the relation between gender and the immediate posttest scores, as the indirect effect was statistically significantly different from zero, $ab = -0.63$, 95% *CI* [−1.47, −0.045].

Study 3 method

Results from Study 2 replicated our Study 1 findings that male and female students engaged differently with *Decimal Point,* and that these differences in engagement mediated the relation between gender and learning outcomes. In Study 3, we assessed whether we could replicate our findings again with a third *Decimal Point* dataset. Study 3 utilized a dataset collected in the 2018 spring semester; results were originally reported in Harpstead et al., (2019). As with Study 2, the original purpose of Study 3 was to investigate the effect of agency in digital learning games. Specifically, Study 3 was originally designed to examine whether enabling students to choose which mini-games to play and when to quit would lead to different behaviors and learning outcomes; it also investigated the effects of indirect control on students' gameplay choices.

Study 3 participants

In the Study 3 dataset, 285 students at two different middle schools in the same northeastern metropolitan area used *Decimal Point* as part of their normal math instruction. Students who did not complete the pretest, posttest or delayed posttest, who did not complete the learning materials, or who experienced log-in errors were excluded from the analysis (N = 48). One additional student was excluded as an outlier based on posttest scores, and another was excluded because they declined to provide gender information. In total, 237

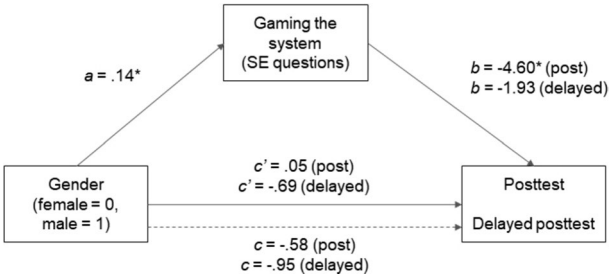

Fig. 5 The original theme park map (left) and the map without a line (right) used to compare high-agency conditions in Study 3; the line was considered a form of indirect control hypothesized to constrain learners' choices

students were included in the analysis, with 130 female students and 107 male students. Students used either the original version of *Decimal Point* or one of two modified versions of the game that allowed students to select the order in which they would play the mini-games and when to quit after completing a minimum of 24 rounds. As with Study 2, we collapsed across all conditions when analyzing the Spring 2018 data set because the within-game mechanisms and content did not vary across conditions.

Study 3 materials and procedure

The materials for Study 3 included the same two conditions used in Study 2 (the original low-agency condition and a high-agency condition that introduced student choice), as well as a third condition that removed the visual path through the amusement park map (Fig. 5). Results from the Nguyen et al. (2018) indicated that students tended to follow the same path through the amusement park even when they had the choice to play in a different order, and the authors speculated that the visual path might exert indirect control over students' selections concerning the order in which they completed the games (Harpstead et al., 2019).

Knowledge tests

The same three isomorphic tests designed for Study 1 were used to assess knowledge in Study 3. As in Studies 1 and 2, tests were counterbalanced and administered as a pretest immediately before the beginning of the game to assess prior knowledge; a posttest administered immediately after the end of the game to assess knowledge after completing the game; and a delayed posttest administered one week after the end of the game to assess knowledge retention.

Gaming detectors

The same models developed to recognize gaming the system within the interaction data from Study 1 were applied to interaction data to detect gaming in Study 3.

Self-efficacy and interest surveys

Study 3 included the same measures of affective experiences introduced in Study 2: a self-efficacy questionnaire (reduced to four items for Study 3, $a = 0.84$) and an interest

Table 8 Correlations between test performance, gender (female = 0, male = 1), gaming the system, decimal self-efficacy, and interest in the game for Study 3

	Posttest	Delayed posttest	Gender	Gaming (PS)	Gaming (SE)	Self-efficacy	Interest in game
Pretest	0.80*	0.79*	0.12	−0.60*	−0.45*	0.46*	0.05
Posttest		0.86*	0.02	−0.59*	−0.49*	0.45*	0.11
Delayed Posttest			0.04	−0.61*	−0.51*	0.42*	0.11
Gender				−0.04	0.23*	0.10	−0.12
Gaming (PS)					0.47*	−0.30*	−0.05
Gaming (SE)						−0.24*	−0.12
Self-efficacy							0.26*

*$p < 0.05$

questionnaire ($a = 0.87$). For both questionnaires, students responded on a 1–5 Likert-type scale from 1 (strongly disagree) to 5 (strongly agree).

Study 3 results

First, rates of gaming were computed for each student and step to reflect how often each student gamed the system on the problem-solving and self-explanation steps. Correlations between test scores, gender, and rates of gaming are shown in Table 8. There were strong, negative correlations between test performance and gaming the system and strong, positive correlations between test performance and decimal self-efficacy.

We then compared rates of gaming between female and male students on the problem-solving and self-explanation steps in the game. Means and standard deviations for test scores, self-efficacy, interest, and rates of gaming are reported in Table 9. A one-way ANOVA indicated no effect of gender on pretest, $F(1,235) = 3.45$, $p = 0.064$. A one-way ANCOVA controlling for pretest revealed a significant effect of gender on posttest, $F(1,235) = 3.93$, $p = 0.048$, with female students improving more than male students when controlling for pretest. There was no effect of gender on delayed posttest when controlling for pretest, $F(1,234) = 2.08$, $p = 0.15$. As in Studies 1 and 2, male students gamed the system significantly more than female students within the self-explanation steps, $F(1,235) = 12.58$, $p < 0.001$. As in Studies 1 and 2, there was not a statistically significant difference on the problem-solving step, $F(1,235) = 0.39$, $p = 0.53$. There were also no significant differences in students' decimal self-efficacy, $F(1,235) = 2.45$, $p = 0.12$, or interest in the game, $F(1,235) = 3.28$, $p = 0.072$.

As with Studies 1 and 2, regression models were used to predict the immediate and delayed posttest scores by rates of gaming in the self-explanation steps, controlling for pretest scores. Both pretest and gaming the system in the self-explanation step significantly predicted students' scores on the immediate posttest, with both predictors statistically significant (see Table 10). A model with just pretest scores predicted 64 percent of the variance in posttest scores, and the overall model predicted 66 percent of the variance, indicating that adding the rate of gaming the system on the self-explanation step predicted an additional 2 percent of the variance over the pretest alone. Despite the small amount of

Table 9 Average probabilities of gaming the system for problem solving (PS) and self-explanation (SE) activities, and average self-reported decimal self-efficacy and game interest

	N	Gaming (PS) *M* (*SD*)	Gaming (SE) *M* (*SD*)	Self-efficacy *M* (*SD*)	Interest *M* (*SD*)	Prestest *M* (*SD*)	Immed. Posttest *M* (*SD*)	Delayed posttest
Female	130	0.22 (0.19)	0.28 (0.25)	3.85 (0.87)	3.50 (1.01)	36.84 (11.78)	42.69 (10.06)	43.55 (11.02)
Male	107	0.20 (0.15)	0.41 (0.33)	4.03 (0.88)	3.24 (1.23)	39.73 (12.09)	43.09 (11.33)	44.35 (10.71)

Table 10 Predicting immediate and delayed posttests with pretest and gaming the system

	Immediate posttest	Delayed posttest
Overall model	$R^2 = 0.66$, $F(2,234) = 224.87$, $p < 0.001$	$R^2 = 0.65$, $F(2,234) = 214.59$, $p < 0.001$
Pretest	$\beta = 0.72$, $p < 0.001$	$\beta = 0.70$, $p < 0.001$
Gaming (SE)	$\beta = -0.17$, $p < 0.001$	$\beta = -0.20$, $p < 0.001$

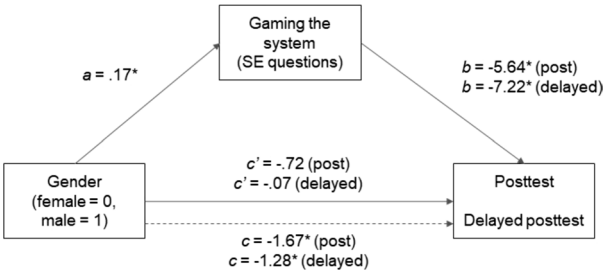

Fig. 6 The mediation model showing path standardized coefficients for a mediation analysis of gender on posttest through gaming the system on self-explanation questions, in Study 3

additional variance explained by gaming the system, this predictor remained statistically significant in a combined model, $t(234) = -3.92$, $p < 0.001$.

In addition, both pretest and gaming the system in the self-explanation step significantly predicted students' scores on the delayed posttest, with both predictors statistically significant (Table 10). A model with just pretest scores predicted 62 percent of the variance in posttest scores, and the overall model predicted 65 percent of the variance, indicating that adding the rate of gaming the system on the self-explanation step predicted an additional 3 percent of the variance over the pretest alone. Despite the small amount of additional variance explained by gaming the system, this predictor remained statistically significant in a combined model, $t(234) = -4.53$, $p < 0.001$.

Since gaming on the self-explanation step was a significant predictor of immediate posttest and delayed posttest, we built mediation models to test the indirect effects of gender (female = 0, male = 1) on both test scores, with gaming the system in the self-explanation step as the mediator. Pretest scores were again included as a covariate.

For the model predicting immediate posttest scores, the indirect effect of gender on immediate posttest through the mediator of gaming the system was significantly different than zero, $ab = -0.94$, 95% CI [−1.63, −0.38], as was the relationship between indirect effect of gender on delayed posttest scores through the mediator of gaming the system, $ab = -1.21$, 95% CI [−20.01, −0.58] (Fig. 6).

Although there were no significant differences in decimal self-efficacy by gender in Study 3, we again ran the mediation model predicting immediate posttest scores including decimal self-efficacy as an additional covariate to account for the possibility that self-efficacy might contribute to gender differences in gaming the system or learning outcomes. The overall results of the mediation model again did not change. Male students were again found to have gamed the system significantly more often than female students in the self-explanation step, $a = 0.17$, $p < 0.001$, and gaming was found to be significantly negatively associated with the immediate posttest scores, $b = -5.38$, $p = 0.001$. The rate of gaming in the self-explanation steps moderated the relationship between gender and the immediate posttest scores. The total effect ($c = -1.78$, $p = 0.034$) was significant but the direct effect

($c' = -0.87$, $p = 0.31$) was not significant; the indirect effect was statistically significantly different than zero, $ab = -0.91$, 95% CI [−1.59, −0.35].

Similar results were found for delayed posttest. Male students gamed the system significantly more often than female students in the self-explanation step, $a = 0.17$, $p < 0.001$, and gaming was negatively associated with the delayed posttest scores, $b = -7.05$, $p < 0.001$. The total effect ($c = -1.36$, $p = 0.12$) and direct effect ($c' = -0.17$, $p = 0.85$) were not significant, and the indirect effect of gender on delayed posttest through the mediator of gaming the system was significantly different than zero, $ab = -1.19$, 95% CI [−2.04, −0.54]. Results from these mediation models suggest that the frequency of gaming in the self-explanation steps explained the impact of gender on both the immediate and delayed posttest scores.

Discussion

As shown in previous research, digital learning games can be particularly effective for female students. In fact, digital learning games have been found to be more effective for female students than for male students in terms of learning and affective outcomes in a number of studies (Arroyo et al., 2014; Hou et al., 2020; McLaren et al., 2017b; Nguyen et al., 2022). However, few studies have tested whether digital learning games influence students' gameplay behaviors differently for female and male students, and whether these differences could account for the better learning outcomes often seen for female students.

Within this paper, we investigate this issue in the context of data obtained from the learning game *Decimal Point.* A number of studies with *Decimal Point*, over a period of more than five years, have shown that playing *Decimal Point* led to greater learning gains for female students than male students (Nguyen et al., 2022). Our current paper considered the hypothesis that this effect may have been due to differences in the frequency of gaming the system, a disengaged behavior. We analyzed three retrospective data sets collected from students playing *Decimal Point*. We found that in the game condition, but not the non-game condition, male students gamed the system significantly more frequently than female students in one key part of the learning experience, the self-explanation step. However, the male students did not game the system more frequently in other activities within *Decimal Point.*

This pattern of results suggests that male students were not generally inclined to game the system more in *Decimal Point*, but rather that one specific element of the digital learning game was associated with differences in gaming the system between female and male students. We then investigated whether this difference in gaming behavior could explain the difference in learning outcomes between female and male students. We found that indeed, the difference in the rates of gaming in the self-explanation step mediated the relation between gender and learning outcomes across the three datasets. This result provides a potential explanation for why female students benefited more from using *Decimal Point* than male students, a finding reported in previous work (Hou et al., 2020; McLaren et al., 2017b; Nguyen et al., 2022). It also provides a broader hypothesis that the differences in learning game effectiveness between female and male students seen in many cases may be due to differences in the engagement produced by specific games when played by different students.

It is worth considering why gaming the system might have played such a significant role in the different levels of learning experienced by female and male students. As discussed

above, gaming the system is generally associated with poorer learning outcomes, but the prevalence of gaming the system in self-explanation activities might have played a particularly important role. Self-explanation activities help students connect their prior knowledge to new content, correct errors in understanding, and develop deeper knowledge that supports more robust learning and transfer (Chi & Wylie, 2014; McNamara, 2017; Richey & Nokes-Malach, 2015; Rittle-Johnson et al., 2017). Gaming the system—and therefore successfully completing the self-explanation activities without actually self-explaining—is likely to eliminate most or all of these benefits. Gaming the system on self-explanation steps might therefore be especially detrimental to students' learning processes, as this choice can disrupt opportunities to connect newly acquired problem-solving skills with existing knowledge and to fill conceptual knowledge gaps related to the content being learned.

Although we had initially hypothesized that gaming behavior could help explain the differences in learning outcomes in *Decimal Point* for female and male students, we did not initially expect that gender differences in gaming the system would emerge only during self-explanation. This finding is surprising because *Decimal Point* generally has less gaming the system than an intelligent tutor covering the same content (Mogessie et al., 2020), but its playful game mechanics are more prominent during the problem-solving steps than the self-explanation steps. Therefore, if the gameplay itself were more engaging for female students than male students, we might have expected to see a greater impact on engagement—and therefore a greater reduction in gaming the system—during problem-solving steps. One possible explanation is that the digital learning game context may have reduced disengagement overall but actually *increased* the likelihood that students would become more disengaged during a specific part of the activity that more closely resembled typical instruction: the self-explanation steps. If this is the case, it may not be that female students found the game more engaging overall, but rather that they were less likely than male students to become disengaged during the less playful components of the game such as the self-explanation steps.

Another possible explanation for this difference comes from the fact that the self-explanation steps were designed in a way that made them easier to game than the problem-solving steps. While a mindless guess-and-check approach to problem-solving steps could include testing a very large number of possibilities (i.e., all possible locations on a number line, a long list of possible values in sequence problems, all order permutations in ordering problems—cf. Paquette et al., 2014), the self-explanation questions were multiple-choice, typically with 3 or 4 options, and therefore could be answered correctly through gaming within a small number of attempts. However, this difference in question design was true of both the game and non-game versions of the content, and the gender differences emerged only in the game. It may be useful for future research to examine whether similar differences in gaming the system between female and male students emerge if self-explanation questions are formatted in a way that can be less easily gamed, such as open-ended responses or drag-and-drop items (McLaren et al., 2022a).

Limitations and future work

This study has several limitations that should be considered in future work. First of all, it would be worthwhile to consider additional behaviors and indicators that represent engagement and disengagement beyond just gaming the system. Positive engagement produced by the game may manifest as experiences of flow (Perttula et al., 2017) or delight (Rodrigo &

Baker, 2011), and may produce positive behaviors such as persistence (Ventura & Shute, 2013). Beyond just gaming the system, disengagement may manifest as careless errors (Hershkovitz et al., 2012), or actions within the game not aimed at completing the learning task (Sabourin et al., 2011). Different engagement measures may capture other cognitive and motivational aspects of student experiences within digital learning games, such as a desire to get the experience over with (carelessness) or general disinterest in the game (game task-unrelated behavior), different in kind than the motivations and attitudes underlying gaming the system. These measures are not yet available for *Decimal Point* but could be developed. Therefore future work should investigate the prevalence of a broader range of behaviors in games such as *Decimal Point*, and whether they play a mediating role in the relationship between gender and learning outcomes within these games. Doing so will help expand understanding of the role that disengagement plays in the different learning gains seen for female and male students within digital learning games. Beyond this, there will be value in repeating this same type of analysis for other learning activities and contexts, towards fully understanding the many proximal variables that play a role in the complex pattern of differences in learning activities between male and female students.

Another area of future work lies in the application of the research paradigm used here to a broader range of differences between students. As Dele-Ajayi et al. (2018) note, there is evidence that many games' effectiveness varies considerably depending on the characteristics of the learners using those games, but there has been insufficient research into why these effects are seen. By applying automated detection of disengaged behaviors and other key processes such as self-regulated learning (Fan et al., 2022), we can obtain a set of measures that can be used as mediators to investigate the differences in the effectiveness of games between groups. It is possible that differences in gaming the system may explain some of the differences in learning game effectiveness between groups—it is quite plausible that some combination of disengaged behavior, affective state, and self-regulated learning will explain many of these differences. Replicating the analytical methods used in the current study, future research can investigate how specific aspects of student identity and individual differences (e.g., race/ethnicity, cultural backgrounds, game preferences) influence how students interact with digital learning games and specific game features differently, and how these differences influence learning outcomes. Results from this line of research will expand current understanding of why digital learning games work and for whom they work, helping to produce digital learning games with more equitable and positive impacts. Additionally, while we focused on gender differences, it may be that other factors such as self-efficacy or game interest—factors that are known to often vary by gender and impact learners' experiences with digital learning games—could also predict disengagement and learning outcomes (Louis & Mistele, 2012; Riconscente, 2013; Sitzmann, 2011).

A third key area of future work involves broadening the conceptualization of gender applied in our current study. When the *Decimal Point* team initially collected the data used in this paper's secondary data analyses, gender was treated as a binary categorization. However, gender is increasingly understood in research as a complex and dynamic set of traits that go beyond the birth-assigned, binary representation (e.g. Hyde et al., 2019). Using the birth-assigned, binary categorization, as our paper's current analysis necessarily did, oversimplifies the complexity of gender and overlooks within-gender heterogeneity and variation in gendered behavior. As shown in Santos et al. (2006), larger differences are often seen when comparing students in terms of self-reported gendered traits (i.e., masculine, feminine, androgynous and undifferentiated traits) instead of binary, birth-assigned genders (i.e., female and male). As such, future work along these lines should leverage a

richer understanding of gender, studying students in terms of a multidimensional gender framework that better captures the complexity of gender. For instance, future work could complement binary measures of gender with categorizations of additional dimensions, such as gender identity (Wood & Eagly, 2015) and gender typicality (Egan & Perry, 2001). In fact, Nguyen et al. (2023) has already taken a step in this direction by presenting a game survey to over 300 elementary and middle school students and analyzing it according to multiple dimensions of gender. Using a multidimensional gender framework for analyses will help to explicate not just the overall relationships between gender, engagement, and learning, but which more nuanced elements and aspects of gender play these roles.

Conclusions

Overall, this paper's results show that gender interacts with student behavior and learning in complex ways within digital learning games. Previously documented effects for the game *Decimal Point* indicating that female students have better learning outcomes were explored, using an automated measure of a disengaged behavior, gaming the system. Prior work did not clearly indicate what aspects of the learning activities these differences manifested in. Our current results indicate that female students are less likely to game the system than male students on self-explanation steps, and that this difference in behavior mediates the difference in learning outcomes previously observed.

This pattern of results highlights the importance of delving into the fine-grained details of student behavior to understand differences in learning, and the role that automated detectors making inference from student log data can play in this type of research. The results also highlight the importance of examining students' interactions with digital learning games in a more comprehensive way that takes users' gender into consideration.

Going forward, our results show that while there are ample studies investigating the features that make a digital learning game effective, it is equally important to understand how games influence students' learning behaviors and how individual differences, such as gender, can predict differences in such behaviors and learning outcomes. Such an approach is critical for building understanding of when and how different game features will benefit specific students. Through understanding how different students interact with digital learning games, our field can work towards designing and developing digital learning games that are more equitable and ultimately more effective.

Acknowledgements Redacted for review.

References

Adamo-Villani, N., Wilbur, R., & Wasburn, M. (2008). Gender differences in usability and enjoyment of VR educational games: A study of SMILE™. In *2008 International conference visualisation* (pp. 114–119). IEEE.

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

Aleksić, V., & Ivanović, M. (2017). Early adolescent gender and multiple intelligences profiles as predictors of digital gameplay preferences. *Croatian Journal of Education: Hrvatski Časopis Za Odgoj i Obrazovanje, 19*(3), 697–727.

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., 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.

Aleven, V., McLaren, B. M., & Sewall, J. (2009). Scaling up programming by demonstration for intelligent tutoring systems development: An open-access website for middle school mathematics learning. *IEEE Transactions on Learning Technologies, 2*(2), 64–78.

Almeda, M. V., & Baker, R. S. (2020). Predicting student participation in STEM careers: The role of affect and engagement during middle school. *Journal of Educational Data Mining, 12*(2), 33–47.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. In *AIED* (Vol. 2007, pp. 195–202).

Arroyo, I., Woolf, B.P., Burleson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal on Artificial Intelligence in Education.* Special Issue on "Landmark AIED Systems for STEM Learning".

Baker, R., & de Carvalho, A. (2008). Labeling student behavior faster and more precisely with text replays. In *Proceedings of educational data mining 2008*.

Baker, R. S., Corbett, A. T., Koedinger, K. R. (2004a). Detecting student misuse of intelligent tutoring systems. In *Proceedings of the 7th international conference on intelligent tutoring systems* (pp. 531–540).

Baker, R. S., Corbett, A. T., Koedinger, K. R., Wagner, A. Z. (2004b). Off-task behavior in the cognitive tutor classroom: When students "game the system". In *Proceedings of ACM CHI 2004: computer-human interaction* (pp. 383–390).

Baker, R. S. J. D., & Gowda, S. M. (2010). An analysis of the differences in the frequency of students' disengagement in urban, rural, and suburban high schools. In *Proceedings of the 3rd international conference on educational data mining* (pp. 11–20).

Baker, R. S. J. D., Corbett, A. T., & Wagner, A. Z. (2006a). Human classification of low-fidelity replays of student actions. In *Proceedings of the educational data mining workshop at the 8th international conference on intelligent tutoring systems* (pp. 29–36).

Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. (2006b). Adapting to when students game an intelligent tutoring system. In *Proceedings of the 8th international conference on intelligent tutoring systems* (pp. 392–401).

Baker, R. S. J. D., de Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T., & Koedinger, K. R. (2009). Educational software features that encourage and discourage "gaming the system". In *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 475–482).

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010a). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241.

Baker, R. S. J. D., Mitrovic, A., & Mathews, M. (2010b). Detecting gaming the system in constraint-based tutors. In *Proceedings of the 18th annual conference on user modeling, adaptation, and personalization* (pp. 267–278).

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729–760.

Bichler, S., Stadler, M., Bühner, M., Greiff, S., & Fischer, F. (2022). Learning to solve ill-defined statistics problems: Does self-explanation quality mediate the worked example effect? *Instructional Science*, 1–25.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725.

Boyle, E. A., & Connolly, T. (2009). Games for learning: Does gender make a difference? *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, 288–303.

Chang, M., Evans, M., Kim, S., Deater-Deckard, K., & Norton, A. (2014). Educational video games and Students' game engagement. In *2014 International conference on information science & applications (ICISA)* (pp. 1–3). IEEE.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).

Cheng, M. T., Chen, J. H., Chu, S. J., & Chen, S. Y. (2015). The use of serious games in science education: A review of selected empirical research from 2002 to 2013. *Journal of Computers in Education, 2*(3), 353–375.

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

Chou, C., & Tsai, M. J. (2007). Gender differences in Taiwan high school students' computer game playing. *Computers in Human Behavior, 23*(1), 812–824.

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

Clark, D. B., Nelson, B. C., Chang, H. Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57*(3), 2178–2195.

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

Cocea, M., Hershkovitz, A., & Baker, R. S. J. D. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? In *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 507–514).

DeFalco, J. A., Baker, R. S., & D'Mello, S. K. (2014). Addressing behavioral disengagement in online learning. In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for intelligent tutoring systems: Volume 2—instructional management* (pp. 49–56). U.S. Army Research Laboratory.

Dele-Ajayi, O. I. (2018). How can digital educational games be used to improve engagement with mathematics in the classroom? *Unpublished doctoral dissertation*, University of Northumbria at Newcastle (United Kingdom).

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

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103–116.

Easterday, M. W., Aleven, V., Scheines, R., & Carver, S. M. (2017). Using tutors to improve educational games: A cognitive game for policy argument. *Journal of the Learning Sciences, 26*(2), 226–276. https://doi.org/10.1080/10508406.2016.126928

Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology, 37*, 451–463.

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*, 1–39.

Fancsali, S. (2014). Causal discovery with models: Behavior, affect, and learning in cognitive tutor algebra. In *Proceedings of educational data mining 2014*.

Ferguson, C. J., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior, 27*(2), 770–775.

Ferguson, C. J., & Olson, C. K. (2013). Friends, fun, frustration and fantasy: Child motivations for video game play. *Motivation and Emotion, 37*(1), 154–164.

Garber, L. L., Hyatt, E. M., & Boya, Ü. Ö. (2017). Gender differences in learning preferences among participants of serious business games. *The International Journal of Management Education, 15*(2), 11–29.

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming, 41*(2), 238–259.

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

Harpstead, E., MacLellan, C. J., Koedinger, K. R., Aleven, V., Dow, S. P., & Myers, B. (2013). Investigating the solution space of an open-ended educational game using conceptual feature extraction. In *Proceedings of the 6th international conference on educational data mining*.

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* (pp. 121–129).

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*(4), 910–931.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Heeter, C., & Winn, B. (2008). Gender identity, play style, and the design of games for classroom learning. *Beyond Barbie and Mortal Kombat: New perspectives on gender and gaming* (pp. 281–300).

Hershkovitz, A., Baker, R. S., Gobert, J., & Nakama, A. (2012). A data-driven path model of student attributes, affect, and engagement in a computer-based science inquiry microworld. In *Proceedings of the 10th international conference of the learning sciences: The future of learning, ICLS 2012* (pp. 167–174).

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127.

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 *Proceedings of the international conference on artificial intelligence in education* (pp. 255–268). Springer.

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-Jones, P. A., & Demetriou, S. (2009). Uncertainty and engagement with learning games. *Instructional Science, 37*(6), 519–536.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35.

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 V. Aleven, J. Kay, J. Mostow (Eds.), *Proceedings of the 10th international conference on intelligent tutoring systems* (ITS-10). Lecture Notes in Computer Science (Vol. 6094, pp. 346–348). Springer.

Jackson, L. A. (2012). The upside of videogame playing. *Games for Health: Research, Development, and Clinical Applications, 1*(6), 452–455.

Jacovina, M. E., Jackson, G. T., Snow, E. L., & McNamara, D. S. (2016). Timing game-based practice in a reading comprehension strategy tutor. In *Proceedings of the international conference on intelligent tutoring systems* (pp. 59–68). Springer

Jenson, J., & de Castell, S. (2005). Her own boss: Gender and the Pursuit of Incompetent Play. In *DiGRA conference*.

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*, 1246–1252.

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

Kao, D., & Harrell, D. F. (2015). Exploring the use of role model avatars in educational games. In *Proceedings of the eleventh artificial intelligence and interactive digital entertainment conference*.
Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development, 64*(2), 219–244.
Khan, A., Ahmad, F. H., & Malik, M. M. (2017a). Use of digital game-based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference. *Education and Information Technologies, 22*(6), 2767–2804.
Khan, J., Wang, J., Wang, X., Zhang, Y., Hammer, J., Stevens, S., & Washington, R. (2017b). Angle Jungle: an educational game about angles. In *Extended abstracts publication of the annual symposium on computer-human interaction in play* (pp. 633–638).
King, D., Delfabbro, P., & Griffiths, M. (2010). Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health and Addiction, 8*(1), 90–106.
Kinzie, M. B., & Joseph, D. R. (2008). Gender differences in game activity preferences of middle school children: Implications for educational game design. *Educational Technology Research and Development, 56*(5–6), 643–663.
Klisch, Y., Miller, L. M., Wang, S., & Epstein, J. (2012). The impact of a science education game on students' learning and perception of inhalants as body pollutants. *Journal of Science Education and Technology, 21*(2), 295–303.
Koedinger, K. R., & Aleven, V. (2016). An interview reflection on "Intelligent Tutoring Goes to School in the Big City." *International Journal of Artificial Intelligence in Education, 26*(1), 13–24.
Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179–188.
Law, E. L. C. (2010). Learning efficacy of digital educational games: The role of gender and culture. *EdMedia+ Innovate Learning*, 3124–3133.
Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W., & Lobene, E. V. (2014). Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences, 264*, 4–18. https://doi.org/10.1016/j.ins.2013.09.005
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).
Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*, 1163–1190.
Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*, 333–369. https://doi.org/10.1207/s15516709cog0504_2
Manero, B., Torrente, J., Fernandez-Vara, C., & Fernandez-Manjon, B. (2016). Investigating the impact of gaming habits, gender, and age on the effectiveness of an educational video game: An exploratory study. *IEEE Transactions on Learning Technologies, 10*(2), 236–246.
Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology., 70*, 531–549.
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., Farzan, R., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017b). Uncovering gender and problem difficulty effects in learning with an educational game. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Proceedings of the 18th international conference on artificial intelligence in education (AIED 2017)*. LNAI 10331 (pp. 540–543). Springer.
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 B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Cognitive Science Society.
McLaren, B. M., Richey, J. E., Nguyen, H. A., & 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., Richey, J. E., Nguyen, H. A., & Mogessie, M. (2022b). 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).
McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes, 54*(7), 479–492.
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: I. Bittencourt,

M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Proceedings of the 21st international conference on artificial intelligence in education. AIED 2020.* Lecture Notes in Computer Science (LNCS, Vol. 12164). Springer. https://doi.org/10.1007/978-3-030-52240-7_38

Murray, R. C., & VanLehn, K. (2005). Effects of dissuading unnecessary help requests while providing proactive help. In *Proceedings of the 2005 conference on artificial intelligence in education: Supporting learning through intelligent and socially informed technology* (pp. 887–889).

Nguyen, H., Harpstead, E., Wang, Y., & McLaren, B. M. (2018). Student agency and game-based learning: A study comparing low and high agency. In *Proceedings of the international conference on artificial intelligence in education* (pp. 338–351). Springer.

Nguyen, H., Else-Quest, N., Richey, J. E., Hammer, J., Di, S., & McLaren, B. M. (2023). 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., Hou, X., Richey, J. E., & McLaren, B. M. (2022). The impact of gender in learning with games: A consistent effect in a math learning game. *International Journal of Game-Based Learning (IJGBL)., 12*(1), 1–29. https://doi.org/10.4018/IJGBL.309128

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.

Osunde, J., Bacon, L., & Mackinnon, L. (2018). Gender differences and digital learning games–one size does not fit all. In *International conference on gender research* (p. 271).

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1–12.

Paquette, L. & Baker, R. S. (2017). Variations of gaming behaviors across populations of students and across learning environments. In *Proceedings of the 18th international conference on artificial intelligence in education* (pp. 274–286).

Paquette, L. & Baker, R. S. (2019). Comparing machine learning to knowledge engineering for student behavior modelling: A case study in gaming the system. *Interactive Learning Environments*, 585–597.

Paquette, L., de Carvalho, A. M. J. A., & Baker, R. S. (2014). Towards understanding expert coding of student disengagement in online learning. In *Proceedings of the 36th annual cognitive science conference* (pp. 1126–1131).

Pardos, Z. A., Baker, R. S., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. *Journal of Learning Analytics, 1*(1), 107–128.

Perttula, A., Kiili, K., Lindstedt, A., & Tuomi, P. (2017). Flow experience in game based learning—A systematic literature review. *International Journal of Serious Games*, *4*(1).

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.

Raney, A. A., Smith, J. K., & Baker, K. (2006). Adolescents and the Appeal of Video Games. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 165–179). Lawrence Erlbaum Associates Publishers.

Renkl, A., & Atkinson, R. K. (2002). Learning from examples: Fostering self-explanations in computer-based learning environments. *Interactive Learning Environments, 10*(2), 105–119.

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)*.

Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust knowledge. *Educational Psychology Review, 27*(1), 181–218.

Riconscente, M. M. (2013). Results from a controlled study of the iPad fractions game Motion Math. *Games and Culture, 8*(4), 186–214.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM—Mathematics Education, 49*(4), 599–611.

Rodrigo, M. M. T., & Baker, R. S. J. D. (2011). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In R. A. Calvo & S. D'Mello (Eds.), *New perspectives on affect and learning technologies* (pp. 183–202). Springer.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In *Proceedings AIED* (Vol. 2007, pp. 203–210).

Romrell, D. (2014). Gender and gaming: A literature review. In *Annual meeting of the AECT international convention*, Hyatt Regency Orange County, Anaheim, CA (pp. 11–22).

Rowe, E., Almeda, M. V., Asbell-Clarke, J., Scruggs, R., Baker, R., Bardar, E., & Gasca, S. (2021). Assessing implicit computational thinking in Zoombinis puzzle gameplay. *Computers in Human Behavior, 120*. Special Issue on "Towards Strengthening Links between Learning Analytics and Assessment: Challenges and Potentials of a Promising New Bond"

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning* (pp. 271–286).

Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In: *Proceedings of the international conference on artificial intelligence in education* (pp. 534–536). Springer.

San Pedro, M. O. Z., Baker, R. S. J. D., Bowers, A. J., & Heffernan, N. T. (2013) Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Proceedings of the 6th international conference on educational data mining* (pp. 177–184).

Santos, D., Ursini, S., Ramirez, M. P., & Sanchez, G. (2006). Mathematics achievement: Sex differences vs. gender differences. In *Proceedings 30th conference of the international group for the psychology of mathematics education* (Vol. 5, pp. 41–48).

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23–52.

Scoresby, J., & Shelton, B. E. (2011). Visual perspectives within educational computer games: Effects on presence and flow within virtual immersive learning environments. *Instructional Science, 39*(3), 227–254.

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology, 43*(4), 540–560.

Shute, V. J., Ventura, M., Kim, Y. J., & Wang, L. (2014). Video games and learning. In W. G. Tierney, Z. Corwin, T. Fullerton, & G. Ragusa (Eds.), *Postsecondary play: The role of games and social media in higher education* (pp. 217–235). John Hopkins University Press.

Shute, V., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M., & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education, 86*, 224–235.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., Kuba, R., Liu, Z., Yang, X., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning, 37*(1), 127–141.

Siew, N. M., Geofrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using Dragonbox 12+ app. *Electronic Journal of Mathematics & Technology, 10*(2).

Singley, M. K., & Anderson, J. R. (1989). *The transfer of a cognitive skill*. Harvard University Press.

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

Slater, S., Ocumpaugh, J., Baker, R., Scupelli, P., Inventado, P.S., & Heffernan, N. (2016) Semantic Features of Math Problems: Relationships to Student Learning and Engagement. *Proceedings of the 9th International Conference on Educational Data Mining*, 223–230.

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.

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 M. v. d. Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 217–224). PME.

Steiner, C. M., Kickmeier-Rust, M. D., & Albert, D. (2009). Little big difference: Gender aspects and gender-based adaptation in educational games. In *Proceedings of the international conference on technologies for e-learning and digital entertainment* (pp. 150–161). Springer.

Tahir, F., Mitrovic, A., & Sotardi, V. (2020). Investigating the effects of gamifying SQL-tutor. In: H. J. So, et al. (Eds.), *Proceedings of the 28th international conference on computers in education* (pp. 416–425). Asia-Pacific Society for Computers in Education. ISBN 978-986-97214-5-5.

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

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior, 29*(6), 2568–2572.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*(3), 229–243.

Walkington, C., & Maull, K. (2011). Exploring the assistance dilemma: The case of context personalization. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33, No. 33).

Walonoski, J. A., & Heffernan, N. T. (2006a). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *International conference on intelligent tutoring systems* (pp. 382–391).

Walonoski, J. A., & Heffernan, N. T. (2006b). Prevention of off-task gaming behavior in intelligent tutoring systems. In *International conference on intelligent tutoring systems* (pp. 722–724). Springer.

Wang, L., Kim, Y. J., & Shute, V. (2013). "Gaming the system" in Newton's playground. In *AIED 2013 workshops proceedings volume 2 scaffolding in open-ended learning environments* (p. 85).

Wolters, C. A., & Pintrich, P. R. (1998). Contextual differences in student motivation and self-regulated learning in mathematics, English, and social studies classrooms. *Instructional Science, 26*(1), 27–47.

Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles, 73*(11), 461–473.

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

Xia, M., Asano, Y., Williams, J. J., Qu, H., & Ma, X. (2020). Using information visualization to promote students' reflection on "gaming the system" in online learning. In *Proceedings of the seventh ACM conference on learning@ scale* (pp. 37–49).

Authors and Affiliations

Ryan S. Baker[1] · J. Elizabeth Richey[2] · Jiayi Zhang[1] · Shamya Karumbaiah[3] · Juan Miguel Andres-Bray[1] · Huy Anh Nguyen[3] · Juliana Maria Alexandra L. Andres[1,2,3] · Bruce M. McLaren[3]

✉ Ryan S. Baker
rybaker@upenn.edu

J. Elizabeth Richey
jelizabethrichey@gmail.com

Jiayi Zhang
joycez@upenn.edu

Shamya Karumbaiah
schoduma@andrew.cmu.edu

Juan Miguel Andres-Bray
miglimjapandres@gmail.com

Huy Anh Nguyen
hn1@cs.cmu.edu

Juliana Maria Alexandra L. Andres
alexandralandres@gmail.com

Bruce M. McLaren
bmclaren@andrew.cmu.edu

[1] University of Pennsylvania, Philadelphia, PA, USA

[2] University of Pittsburgh, Pittsburgh, PA, USA

[3] Carnegie Mellon University, Pittsburgh, PA, USA