Computers & Education 178 (2022) 104366

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

How instructional context can impact learning with educational technology: Lessons from a study with a digital learning game

Bruce M. McLaren [a,*], J. Elizabeth Richey [b], Huy Nguyen [a], Xinying Hou [c]

[a] Carnegie Mellon University, USA
[b] University of Pittsburgh, USA
[c] University of Michigan, USA

ARTICLE INFO

Keywords:
Distance education and online learning
Games
Gender studies
Mathematics learning
Hints

ABSTRACT

There is a tendency to think of the impact of educational technology in a vacuum. However, it is likely that the *instructional context* in which educational technology is used affects student learning. For instance, outcomes may differ when using educational technology in a classroom versus at home, in a quiet versus noisy environment, or in a context where support is readily available versus not available. The COVID-19 pandemic provided an unexpected opportunity to explore this issue. Intending to explore how providing hints and feedback within a digital learning game (*Decimal Point*) impacts mathematics learning, we instead found ourselves exploring a new question: How did learning with the game differ between classrooms and at home? After two of five middle schools had participated in our classroom experiment, we switched to at-home use of the Internet-based game for the final three schools due to the pandemic. The different instructional settings led to significantly different completion rates, likely due to students in the classroom (N = 151) being monitored by experimenters and teachers (completion rate of 88.8%), while students at home (N = 126) were not monitored nor strictly required to finish (completion rate of 56.5%). In addition, the two versions of the game, one that provided students with on-request hints and error feedback (*Hint* condition) and one that did not (*No-Hint* condition), led to different classroom versus at-home results. On the delayed posttest, students in the *No-Hint* condition did significantly better in the classroom, while there was no significant difference between conditions at home. In addition, students in the *Hint* condition used significantly more hints in the classroom than they did at home. There was also a significant effect of gender in the classroom, with female students out-performing male students on the immediate posttest, but with no effect of gender remotely. We performed post-hoc analyses to better understand students' learning processes and gameplay behaviors. In summary, our study clearly illustrates how educational technology can be sensitive to instructional context, yet just cracks open the door to much more research on this topic.

1. Introduction

The instructional context in which students learn can make a huge difference to their learning (Ferdig et al., 2020; Sadeghi, 2019; Tessmer & Richey, 1997; Wang & Huang, 2018). Yet, the world of educational technology, both research and application, has been

* Corresponding author. Carnegie Mellon University, 2602I Newell-Simon Hall, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA.
E-mail address: bmclaren@cs.cmu.edu (B.M. McLaren).

https://doi.org/10.1016/j.compedu.2021.104366
Received 4 May 2021; Received in revised form 12 August 2021; Accepted 2 November 2021
Available online 5 November 2021
0360-1315/

largely founded on and driven by cognitivist principles (Anderson et al., 1996), which posit that much of learning involves absorbing "rules of the mind" (Anderson, 1993) that are *free* of context and can transfer to other contexts. Given this history and perspective on learning, how context impacts learning with educational technology is not well understood and has been understudied (Wen et al., 2021). In this work, we take a step toward correcting this shortcoming by explicitly investigating how context impacts student learning with educational technology.

This research direction has recently taken on practical urgency and significance in education. Given the COVID-19 pandemic of 2020–2021 and the necessary shift from classroom to remote learning, there has been a huge spike in home use of educational technology for learning (BrightBytes, 2020; Reich et al., 2020). Yet, the rapid shift to remote learning with technology has presented clear challenges for students and teachers (Lockee, 2021; Means et al., 2021). For instance, up to a third of students report that they sometimes or never have technical support for learning remotely and parents overestimate how often their children have a quiet place to work on a computer at home (BrightBytes, 2020).

The pandemic provided us with a unique opportunity to explore this issue, comparing the use of educational technology in the classroom to its use at home. We were conducting an experiment, and had already been in two K-12 schools, when the pandemic forced students across the U.S. to continue learning from home. Thus, we conducted the study at the final three K-12 schools with students working online at home. While unfortunate, this change in the study context provided us with the opportunity to contrast our primary research question – "Do students who learn with a digital learning game benefit from on-demand hints and feedback versus no such support?" – in a classroom versus at home.

The differences between classroom and at-home learning with educational technology are in many ways stark and clear. In the typical K-12 classroom, a teacher acts as a guide and facilitator of learning when students use software to learn. Students are subject to the expectations of the teacher. They are also under time pressure; classes are typically 40–90 min in length and occur at a specific time of day. Teachers typically maintain a quiet environment in the classroom, and rules govern what students can do in a classroom, such as when they can take breaks, whom they can sit next to, and whether there is any background noise. In contrast, at-home learning leaves students much more on their own, with varied support from others, such as parents and teachers, for learning and technical support. Also, the noise level in learning from home can vary dramatically, wholly dependent on each student's particular home environment. Students often can work when they want to at home, day or night, and for as long or as short as they would like. They are also much less subject to rules, except perhaps those imposed by their parents, and those rules could be quite different from those of teachers.

As mentioned, our original intent in this study was not to explore instructional context; rather, it was to investigate how adding on-demand hints and feedback to a digital learning game, *Decimal Point* (Forlizzi et al., 2014; McLaren et al., 2017a), would impact learning and engagement with the game. Note that it is not clear that adding scaffolding in the form of hints and feedback will improve either learning or engagement. On one hand, the additional instructional support provided by hints and feedback might be expected to help students better learn the domain content with a digital learning game, as has been shown in many intelligent tutoring systems, across a variety of domains (see e.g., Ma et al., 2014; VanLehn, 2006, 2011; Xu et al., 2019), as well as in multi-media learning more generally (Johnson & Priest, 2014, pp. 449–463). On the other hand, providing such support during gameplay might have the adverse effect of dampening enthusiasm and engagement, reducing opportunities for self-explanation and other constructive knowledge building processes, and thereby decreasing learning. While adding hints and feedback to digital learning games has been explored in some past work (e.g., Burgers et al., 2015; Conati et al., 2013; Drey et al., 2020; Easterday et al., 2017; Lee & Chiou, 2020; Melero et al., 2012; O'Rourke et al., 2014), the results have been mixed, not always leading to learning and engagement benefits and often focusing more on how hints are perceived and used than whether they lead to learning.

Between our original study plan to explore on-demand hints and feedback in the context of the digital learning game and the emergent issue of instructional context, our research questions in this study were:

RQ1. *How does the instructional context (i.e., classroom versus remote learning) impact playing of the game and learning with respect to all of the following research questions?* As this research question emerged during the course of the study, we did not have any predictions for how instructional context would impact RQ2 through RQ4. Instead, we viewed this as an exploratory research question that ultimately became our most important question to answer.

RQ2. *Do students who learn with a digital learning game benefit from on-demand hints and feedback (Hint condition) versus no such support (No-Hint condition)?* Due to the additional support students receive for learning in the Hint condition, as well as the evidence that such support has helped students to learn with intelligent tutoring systems (Ma et al., 2014; VanLehn, 2006, 2011; Xu et al., 2019), we predicted that students in the Hint condition would learn more than students in the No-Hint condition. On the other hand, this outcome is far from assured, as adding hints to digital learning games has not always resulted in learning benefits (e.g., Moyer-Packenham et al., 2019; O'Rourke et al., 2014).

RQ3. *Do female or male students benefit more from the hints?* Because we had previously identified a gender effect in our studies with *Decimal Point* (Hou et al., 2020; McLaren et al., 2017b; Nguyen et al., under review), in which female students learned more from the game than male students, we were interested in the question of what, if any, effect adding on-demand hints and feedback might have on that difference. We expected that female and male students would equally benefit from the addition of hints and feedback, so we predicted that there would be no learning differences between female and male students in their use of hints. In other words, we predicted that female students would still benefit more from the game than male students, as in previous studies, and that hints would help both genders, but that the difference between genders would remain the same.

RQ4. *In the Hint condition, how do hint behaviors relate to prior knowledge and learning outcomes?* Since higher achieving students also tend to be better self-regulated learners (SRL) than students with lower prior knowledge (Winne, 2015; Zimmerman, 1990), we predicted that the students with higher prior knowledge would use hints more productively. Note that more productive hint behaviors from students with higher prior knowledge is not guaranteed, given that students at any knowledge level may still abuse hints or struggle to recognize when they need hints and when they are better off attempting a response without a hint.

2. Relevant past research

2.1. Instructional context: classroom versus at home

In our work, we define “instructional context” as the physical location and surroundings in which learning takes place, such as classrooms, museums, workplaces, or homes, along with the resources that are typically available in such locations (e.g., people, artifacts, books, online information). Definition is important, because “context” can also mean, for instance, historical period, social relationships, cultural environment and activity (Dohn et al., 2018). Some have even defined it as the goals a student might have, such as whether they intend to major in a particular topic or undertake a particular profession (Artino, 2009).

Perhaps the most extensive analysis of context comes from Tessmer and Richey (1997) who look at both the spatial (e.g., classroom or workplace) and temporal (pretraining, training, posttraining) qualities of context in which learning occurs. They argue that “context has a complex and powerful influence upon successful performance-based learning, and yet is largely ignored (or at the least deemphasized) in most current instructional design models” (Tessmer & Richey, 1997, p. 85). While this paper is now over two decades old, its premise still stands: instructional context is important to learning but has been understudied. Undertaking such study is especially valuable in the modern era of online learning, which takes place in a multitude of places: at home, in a coffee shop, on a bus or even in a forest, in addition to traditional environments, such as the classroom.

Nevertheless, instructional context has not always factored into the cognitive models that inform the design of educational technology (e.g., Anderson et al., 1996). In fact, there has been limited research regarding the efficacy of learning with educational technology in the classroom versus at home (Wen et al., 2021). For instance, a meta-analysis of intelligent tutoring systems (ITSs) by Steenbergen-Hu and Cooper (2013) identified studies of classroom ITS use and homework ITS use, with slightly larger effect sizes for homework ITSs. However, only two of these studies looked at ITSs used at home to complete homework, and these focused exclusively on homework use, meaning they didn’t test context as a moderator or statistically compare effect sizes for classroom versus home use. Thus, while we may be able to conclude that ITSs can promote learning when used for assignments at home, the relative *differences* between classroom and home use are not at all clear from these results. The few efforts in the direction of comparing classroom and home use of educational technology include the work of Makransky et al. (2019), who directly compared classroom and home use of a virtual reality science simulation and found no difference in students’ learning or motivation based on context, and Wen et al. (2021) who reviewed prior studies of educational technology research outside the classroom in an effort to define the prerequisites for home-based learning (contrasted with classroom learning).

2.2. Instructional context: the impact of the COVID-19 pandemic

More practically and more recently, there has been some examination of remote learning, due to the COVID-19 pandemic (BrightBytes, 2020; Reich et al., 2020). In a survey of nearly 50,000 students, 11,889 teachers, 33,182 parents and 580 school principals done by the International Society for Technology in Education (ISTE) and Brightbytes (BrightBytes, 2020) between April and June of 2020 – during the first months of the pandemic – some clear differences were identified between students studying remotely at home and in the classroom. For instance, only a very slim majority of students said they consistently receive the help they need (52%), while 21% students said they “sometimes” or “never” receive teacher help. Despite students’ needs for guidance and support typically varying with age, this is a significant contextual difference, particularly for students who struggle and need encouragement to work regardless of whether they are using learning technology. Furthermore, 30% of students said they “sometimes” or “never” had a quiet space to work, yet only 13% of parents said the same about their students. Although these results reflect families’ experiences with emergency remote education, rather than planned and well-structured remote teaching, they are still informative regarding the types of challenges students commonly experience when completing work at home. Consistent with these findings, a pre-pandemic comparison of students learning through in-person lectures versus attending the same lectures remotely online, indicated that students learning remotely experienced significantly greater cognitive load, which can ultimately reduce learning outcomes (Andersen & Makransky, 2021).

Another key issue that arose with at-home learning during COVID-19 was equitable access to technology. Nearly all U.S. state agencies emphasized issues of equity with respect to access to online options and encouraged asynchronous over synchronous learning, due to technology access constraints at home (Reich et al., 2020). While asynchronous learning certainly makes practical sense, given the constraints, it would also make it difficult for teachers to monitor their students’ activity and provide the real-time guidance so readily available in the classroom.

2.3. Use of hints in learning: with and without educational technology

Regarding our original research questions of hints versus no hints in the context of *Decimal Point*, there is substantial research regarding feedback and support in learning more generally; that is, without technology. Hattie and Timperley (2007), in a classic and thorough article about the value of feedback in learning without technology, determined that the type of feedback and the way it is given to learners can make a significant difference to learning. Their research, which cited a wide range of studies (e.g., Kluger & DeNisi, 1996; Kulik & Kulik, 1988), revealed that immediate and delayed feedback vary on their beneficial effects based on the task level; immediate feedback tends to lead to more benefits in task acquisition, while delayed feedback tends to better support strategy learning. Both positive and negative feedback can have beneficial effects on learning, depending on whether that feedback is about the task, processing the task, or self-regulation.

With respect to learning *with* technology, Koedinger and Aleven (2007, p. 239) raised and analyzed the important "assistance dilemma" question: "How should learning environments balance information or assistance *giving* and *withholding* to achieve optimal student learning?" They expanded on this question by pointing out that providing too much assistance could lead to the thrill of (supported) success, but at the potential cost of shallow processing and not updating long-term memory, while too little assistance could lead to independent success and engaging long-term memory, but at the potential cost of floundering and wasted time. They reported on attempts to address this question in a variety of studies with cognitive tutors. This work was inspired by earlier research with a geometry cognitive tutor in which they observed that students quite often don't know when they need help, seeking help when they don't need it or not seeking help when they do need it (Aleven & Koedinger, 2000). Subsequently, this led to a line of research with the *Help Tutor*, a computer-based meta-cognitive tutor designed to help students learn when to seek help (Aleven et al., 2016). The *Help Tutor* gave in-context, real-time feedback on students' help-seeking behavior, as they were learning with an intelligent tutoring system. While this research resulted in students demonstrating better help-seeking behaviors, it did not lead to better learning outcomes in the domain of geometry. In summation, while we have learned much about feedback in instruction, we still have not been entirely successful in building learning technology that optimizes when and how to provide feedback to learners.

2.4. Use of hints in digital learning games

In the context of digital learning games, the picture is even murkier, with much less research on providing help to learners while game playing and results that point in different directions. Digital learning games are a relatively nascent learning technology (Clark et al., 2016; Mayer, 2014, 2019), with substantially less proven learning impacts than, for instance, intelligent tutoring systems (Ma et al., 2014; VanLehn, 2006, 2011; Xu et al., 2019). Providing help in the context of games raises the question of whether the hoped for flow (Czikszentmihalyi, 1990) and engagement (Bouvier et al., 2013) of learning games might be disrupted if help is inserted within the games. Flow and engagement are, in fact, key reasons why digital learning games have been hyped as one of the most promising computer-based methods to support learning (Gee, 2003; Prensky, 2006; Squire & Jenkins, 2003).

Yet, there has been some empirical research in adding hints to the context of digital learning games. For instance, O'Rourke et al. (2014), in an experiment involving over 50,000 students with *Refraction*, a digital learning game about fractions, explored different hint content (concrete versus abstract) and hint presentation (by level versus reward). In a 2 × 2 comparison of hint content and hint presentation, plus a condition with no hints at all, they found that students in the no-hint condition learned more than students in any of the other conditions. While their huge sample size is impressive, it is offset by shortcomings of the study: It had no pre or posttest, relying solely on progress in the game as a learning metric, since students had to correctly solve problems to advance. Another potential flaw to the approach of *Refraction* is that students would receive hints as rewards for success in the game, which seems counter to a sound instructional approach of providing hints to the students who are less successful and more in need of help. Other studies of hints used with digital learning games have focused more on whether students pay attention to and are engaged with hints than whether they actually learn from the support. For instance, Conati et al. (2013) used eye tracking technology to capture student attention to a digital learning game's adaptive hints, which were both game generated (similar to the error messages of *Decimal Point*) and on demand, and found that as the game proceeded, students paid less and less attention to the hints. Across three studies comparing hints and no hints in puzzle-solving digital learning games, Melero et al. (2012) found that students who received hints said the hints were helpful, and that many students in the non-hint conditions said they could have used help. In more recent work, Lee and Chiou (2020) compared different types of hints – static, video, and multiple levels of static – in the context of *Gidget*, a digital learning game for learning programming. In a study involving 150 students, they found that video hints both engaged students the most and were most preferred. Yet, there was no evidence they actually used the hints to support their learning. In sum, these results suggest that students often desire hints and perceive them as useful, but they may not pay careful attention to hints when they are available, which could limit the potential learning benefits from hints.

Another investigation of help provided to students while learning with digital games is the Moyer-Packenham et al. (2019) study of 12 digital math learning games. They examined different game design features (e.g., hints, progressive levels, game efficiency) and reported that games with hints were not more likely to lead to learning gains. However, this study was not an experimental study; it did not compare an experimental hint condition with a no-hint control for any of the 12 games. Drey et al. (2020) is promising recent work in which adaptive hints were inserted in the context of a virtual reality game, but also with no empirical results as of yet, either in studying students' affinity for hints or learning benefits.

The only experimental study of a digital learning game that, to our knowledge, actually demonstrated learning benefits with hints and feedback was that done by Easterday et al. (2017). In a study involving 105 university students who played *Policy World*, a learning game for learning policy argumentation, either with or without an underlying tutoring system that provided on-demand hints and feedback, the students who played the game with the underlying tutor learned more. A path analysis supported the authors' claim that the feedback of the tutor led to better learning, which in turn increased students' feelings of competence and interest. This is an important finding, since it showed that a digital learning game supported with hints and feedback did not have to sacrifice engagement to lead to learning.

Given the paucity of well-founded evidence about whether hints and feedback in the context of a digital learning game can lead to learning – essentially one experimental study in which students without hints learned more than those provided with hints (O'Rourke et al., 2014) and one study in which students without hints learned less than those provided with hints (Easterday et al., 2017) – we decided to explore this issue in the context of *Decimal Point*, a digital learning game developed in our lab and for which we have run a series of experiments. We now turn to a description of this game and how we enhanced it with hints.

3. *Decimal Point*

Decimal Point, depicted in Fig. 1, is a single-player game with an amusement park metaphor targeted at middle-school students learning about decimal numbers (McLaren et al., 2017). The game has been used as a platform to explore a variety of key issues of learning with digital learning games, such as student agency (Harpstead et al., 2019; Nguyen et al., 2018), order of gameplay (Wang et al., 2019) and the tension between enjoyment and learning (Hou et al., 2020). *Decimal Point* runs on the Internet, within a standard browser, and was developed using HTML/JavaScript and the Cognitive Tutor Authoring Tools (CTAT) (Aleven et al., 2016). The materials have been deployed on a web-based learning management system, TutorShop (Aleven et al., 2009), which presents the materials in a given sequence and logs all student actions. *Decimal Point* is composed of a series of "mini-games" (e.g., Enter if You Dare, Space Raider, Castle Attack) within the larger amusement park game, as can be seen in Fig. 1.

Decimal Point presents students with 5 types of decimal problems, each corresponding to a different mini-game type: (1) ordering decimals; (2) correctly placing decimals on a number line; (3) completing decimal sequences; (4) placing decimals in less-than and greater-than "buckets" in comparison to a given decimal; and (5) adding decimals. Forty-eight decimal problems (two problems per 24 mini-games) have been implemented for the game. Figs. 2 and 3 show, respectively, two examples of *Decimal Point* mini-games: Castle Attack, an ordering mini-game, and Capture the Ghost, a bucket mini-game.

The game problem types were developed through a review of the mathematics education literature on common decimal misconceptions (Isotani et al., 2010; Stacey et al., 2001), as well as an analysis of the types of problems students currently encounter in popular math curricula (e.g., Bell et al., 2012). For instance, the Castle Attack mini-game of Fig. 2 probes the misconception many students have that "longer decimals are larger." The student playing Castle Attack in Fig. 2 is, in fact, displaying that misconception in their on-going solution. This type of problem is found in *Everyday Mathematics*, as well, i.e., "Order the following numbers from smallest to greatest: 0.12, 0.05, 0.2, 0.78, 0.6, 0.043, 0.1" (Study Link 5.7).

Each item in *Decimal Point* has only one correct answer, and the player cannot advance until they have provided the correct answer. Players also encounter problems in a pre-set order. For example, the student who is playing *Decimal Point* in Fig. 1 is about to play Castle Attack, indicated by a pulsating circle around that game and the fact that all the circles indicating games before Castle Attack on

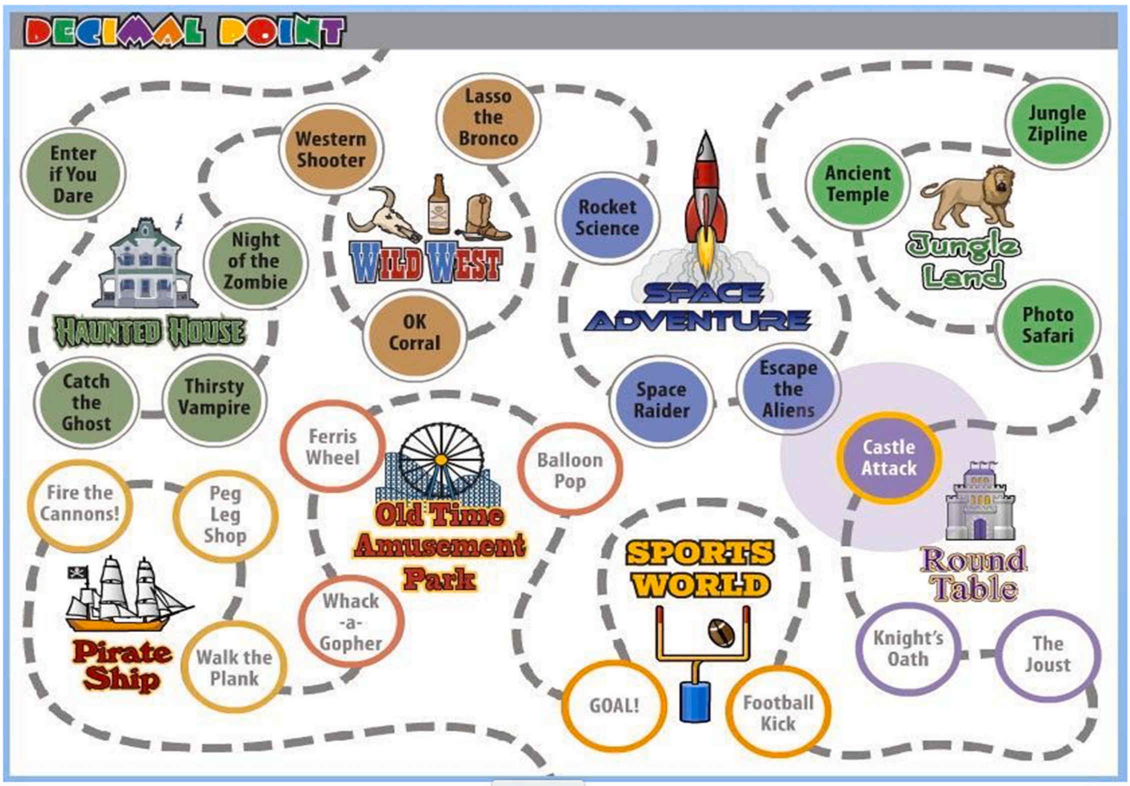

Fig. 1. The main game map of *Decimal Point*.

Fig. 2. "Castle Attack" one example of a mini-game in *Decimal Point*.

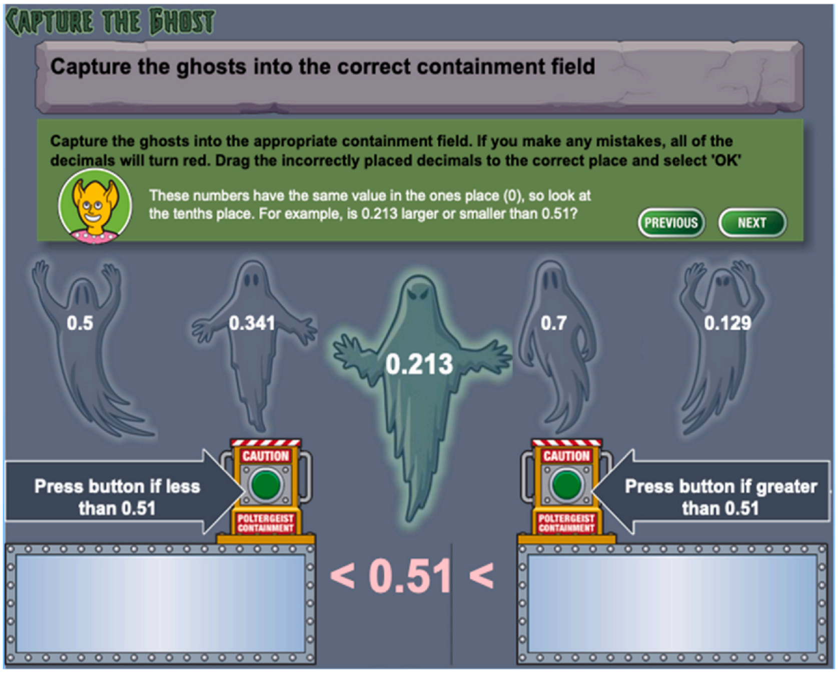

Fig. 3. "Capture the Ghost," a second example of a mini-game in *Decimal Point*.

the map are filled in.

Decimal Point has six characters that serve as guides and cheerleaders for the player throughout the game. These game elements provide fantasy (Malone & Lepper, 1987), as well as giving the player a narrative context for why they are prompted to perform various activities and problem solving. The interface and feedback design presents students with problem-solving activities embedded playfully in the mini-game context. Students are prompted by the characters to correct mistakes after an initial attempt within the game.

Decimal Point also has many of the basic game features outlined by Costello and Edmonds (2007), such as being low action, non-competitive, and without a reward system. In fact, the *Decimal Point* mini-games are intentionally designed to be relatively simple. The Clark et al. (2016) meta-analysis suggests that more complex game designs, with a wide range of solution paths, do not necessarily equate to better learning outcomes, as both simple and complex game mechanics have led to positive learning effects. There may actually be an important instructional design advantage to simplicity: avoiding distracting students with unnecessary decision-making and seductive details (Harp & Mayer, 1998).

The study presented in this paper was designed to explore whether on-demand hints and error messages in the context of *Decimal Point* would lead to better learning and engagement for students than the original version of the game that did not provide such support. Thus, there were two study conditions:

1) The *No-Hint* condition in which students played the original version of the game that provided no hints and only correctness feedback (i.e., turning correct answers green and incorrect answers red) within the individual mini-games, and
2) The *Hint* condition in which students played a version of the game that, in addition to correctness feedback, provided on-demand hints and error messages on common student errors (i.e., when students entered a common error, based on a data analysis of past game use, they received a message specifically addressing the error immediately after entering the incorrect response).

Fig. 2 shows an example of how an on-demand hint is requested (see the button in the middle, on the right). A hint is displayed after the student selects the "Hint" button; see the hint in Fig. 3 (in the middle). After the "Hint" button is selected, the student can go forward and backward through three levels of hints with "Previous" and "Next" buttons (see the buttons in the middle right of Fig. 3). Fig. 3 also shows an example hint (i.e., "These numbers have the same value in the ones place (0), so look at the tenths place. For example, is 0.213 larger or smaller than 0.51?"). There are three levels of hints for each problem: Level 1 hints are very general, typically reminding the student of basic decimal knowledge (e.g., For the problem shown in Fig. 3 - "Compare digits in the same place values of the decimal numbers, moving from the leftmost digit to the rightmost."), Level 2 hints are somewhat more detailed

Table 1
Example Hint and Error Messages. There are three levels of hints per problem-solving step; we provide an example of all three levels for an Ordering mini-game and examples of one level for the others.

Mini-game problem type	Hint examples	Error message example
Ordering	*Level 1*: Compare digits in the same place values of the decimal numbers, moving from the leftmost digit to the rightmost. *Level 2*: Since these numbers all have the same ones place (0), compare the tenths place. Which has the smallest tenths place? *Level 3*: 0.0234 has the smallest tenths place, followed by 0.14, 0.323, 0.4. *(These are the three hint levels provided when the student is given an ordering problem with the decimal numbers 0.14, 0.4, 0.0234, 0.323)*	Start by comparing the first digit to the right of the decimal point, even if the digit is 0. (*If the student orders the decimal numbers 0.14, 0.0234, 0.323, 0.4*)
Number line	*Level 2*: If you divide the space between 0 and 1 into two pieces, 0.5 is at the end of the first piece. Is 0.456 smaller or larger than 0.5? *(Level 2 hint when the student is given a number line problem to place 0.456 on a numberline running -1.0 to 1.0)*	0.456 is greater than 0, so it goes to the right of 0. (*If student clicks to the left of 0, to where the decimal number would be negative*)
Complete Sequence	*Level 2*: What's the change from 0.3 to 0.6? From 0.6 to 0.9? (*Level 2 hint when the student is asked to complete the sequence: 0.3, 0.6, 0.9, __, __*)	Add the same place values and carry to the next place value to the left. (*If student types the next value of the sequence as follows: 0.3, 0.6, 0.9, _ 0.12_, __*)
Buckets (less than/ greater than comparisons)	*Level 1*: Compare digits in the same place values of the decimal numbers, moving from the leftmost digit to the rightmost. *(Level 1 hint when the student is given a middle decimal number of 0.29 and decimal numbers to place to the left and right of that number, smaller and larger: 0.4, 0.101, 0.384, 0.299)*	The number of digits in a decimal number doesn't tell you how big it is. You should compare place value, moving left to right. (*If the student puts all shorter decimals in the left bucket* (e.g., *0.4, and all longer decimals in the right bucket, 0.101, 0.384, 0.299*)
Adding decimals	*Level 1*: Add the digits after the decimal place first, starting with the digits on the far-right side and moving left. Add 0 + 0 first, 5 + 9 s, and so on. *(Level 1 hint when the student is given the addition problem: 7.50 + 3.90)*	Please fill in the boxes to indicate where you are carrying (*If the student does not indicate the carry of "1" across the decimal point.*)

suggestions on how to think about and solve the given problem (e.g., For the problem in Fig. 3, what is shown - “These numbers have the same value in the ones place (0), so look at the tenths place. For example, is 0.213 larger or smaller than 0.51?“), and Level 3 hints, also called “bottom-out hints,” essentially provide the student with the answer (e.g., For the problem in Fig. 3 the bottom-out hint is “0.213, 0.341, 0.5 and 0.129 are less than 0.51, and 0.7 is greater than 0.51.“). Examples of hint and error messages for each of the five problem types are shown in Table 1. The hints and error messages were developed by the authors, together with a mathematics education specialist.

4. Method

4.1. Participants and design

For the pre-COVID, classroom study period (February 3, 2020 through March 5, 2020), we assigned students to condition, *Hint* or *No-Hint*, by class, due to concerns about students observing differences in the game within a classroom (i.e., students without hints might notice their classmates receiving hints, and students with hints might share them with classmates not receiving hints). To roughly balance prior knowledge between conditions, we asked teachers to characterize classes as low, medium, and high performers and then did quasi-random condition assignments so that we had (close to) the same number of classes of each level within each condition. Two schools with a total of 170 5th and 6th grade students participated in the study in classrooms. Five classes were assigned to the *Hint* condition and five classes were assigned to the *No-Hint* condition. Of the 170 students, 18 (8 in the *Hint* condition and 10 in the *No-Hint* condition) were excluded for failing to complete the materials (pretest, game, immediate posttest, or delayed posttest) within the allotted time. One additional student in the *Hint* condition was excluded for performing more than three standard deviations below the mean on the posttest and delayed posttest. Among the remaining 151 students (mean age 11.04, completion rate of 88.8%), sixty-seven (67) students (31 female, 36 male) were assigned to the *Hint* condition and eighty-four (84) students (41 female, 43 male) were assigned to the *No-Hint* condition. For the remote portion of the study, when students were working from home due to the COVID pandemic (April 13, 2020 through May 8, 2020), we randomly assigned students to condition, since there was no longer a concern about students seeing one another’s work. Three schools with a total of 223 6th grade students participated in the study remotely. Ninety-seven (97) students (51 in the *Hint* condition and 46 in the *No-Hint* condition) were excluded from analyses for failing to complete the materials in the allotted time. Among the remaining 126 students (mean age 11.81, completion rate of 56.5%), sixty-four (64) students (33 female, 31 male) were assigned to the *Hint* condition and sixty-two (62) students (31 female, 31 male) were assigned to the *No-Hint* condition.

Table 2 shows the type of school, the number of participating students, and the completion rates of each school.

Students took a pretest, completed the intervention according to condition (i.e., played all of the mini-games in the order shown in Fig. 1, with two problems per mini-game), took an immediate posttest, and, one week after completing the posttest, completed a delayed posttest. Examples of the same mini-game from the *Hint* and *No-Hint* condition are shown in Figs. 4 and 5. The only differences between the games were (1) students in the *Hint* condition were provided with a “Hint” button (such as shown in Fig. 2) that allowed them to request on-demand hints and (2) common errors made by students elicited an error message, such as what appears in Fig. 6.

In the classroom version of the study, students worked during 45- to 55-min class periods for up to a full week on the pretest, game, and immediate posttest. Students worked at their own pace and took between three and five days to complete the materials (see Results for time on task by condition). Students logged out at the end of each class and, when logging back in the next time, were placed in the materials wherever they stopped previously. A week after the immediate posttest, students took the delayed posttest within one class period.

While playing the game in the classroom condition, students in the *No-Hint* condition were given the answer by experimenters if they asked for help after they (a) made at least 5 mistakes and (b) worked for at least 5 min (a subtle indicator on the screen would alert the experimenters that the student had met the (a) and (b) conditions and could thereby know whether they could provide help). This was done to prevent students from getting stuck on a single problem. Students rarely met these criteria for assistance (10 or fewer

Table 2
Participating schools.

	School	Type of School	N	Dropped	Completion Rate
Classroom	School 1 (8 classes)	Rural	116	10 (7 H, 3 NH; includes 1 outlier)	92.1%
	School 2 (2 classes)	Suburban	35	9 (2 H, 7 NH)	79.5%
	Totals		151	19 (9 H, 10 NH)	88.8%
Remote	School 3 (3 classes)	Suburban	39	11 (2 H, 9 NH)	78.0%
	School 4 (4 classes)	Suburban	55	24 (14 H, 10 NH)	69.6%
	School 5 (7 classes)	Urban	32	62 (35 H, 27 NH)	34.0%
	Total (24 classes)		126	97 (51 H, 46 NH)	56.5%

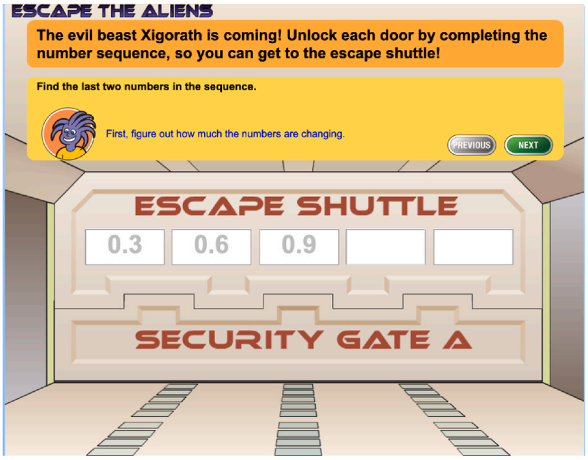

Fig. 4. The “Escape the Aliens” mini-game in the *Hint* condition. Note the hint text shown in the yellow box and the “Previous” and “Next” buttons used to navigate through hints.

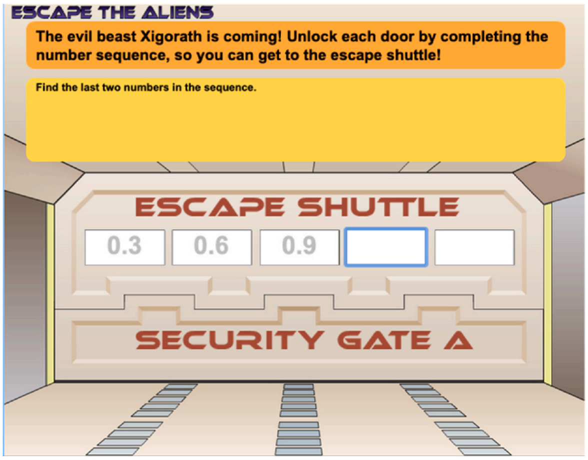

Fig. 5. The “Escape the Aliens” mini-game in the *No-Hint* condition, without any hint buttons.

instances total across the study).

In the remote version of the study, students had a lot more freedom regarding when to work and for how long. Students were given a full week to complete the materials, completing as much, or as little, as they wanted to each day (see Results for time on task by condition). As in the classroom version, when they logged out, their place would be saved, and they would begin from that spot in the materials the next time they logged in. At the end of the week, the initial materials (i.e., pretest, game, and immediate posttest) were locked and a week later the students were given a day to complete the online delayed posttest. Students who used the materials remotely were asked not to use any other resources to help them (e.g., asking their parents, looking things up on the Internet, etc.), but of course there was no way to monitor this activity, as there was in the classroom version of the study.

As a final note, during the pandemic, and in particular during the course of the at-home version of this study, people were in close to a full lockdown, expected to wear masks in public, and some of our students likely contracted COVID-19. It was also the case that in the spring of 2020, when this study was conducted, all schools were operating fully remotely.

4.2. Materials

A web-based learning environment was used to deploy all of the materials in this study, as well as to collect detailed data from the use of the materials (Aleven et al., 2009). As mentioned above, the materials included three tests and the two game conditions.

Pretest, Immediate Posttest, and Delayed Posttest: Each test, which was developed by math education experts and vetted over many prior studies, consisted of 43 items. Most items were worth one point each, while some multi-part items were worth several points, for a total of 52 points per test. The items were designed to probe for specific decimal misconceptions, and involved either the five decimal skills targeted by the game (e.g., “Choose the largest of the following three numbers.__0.22, _0.31, _0.9”; “Write down the next item in the following sequence. 0.3, 0.5, 0.7, 0.9, __”) or conceptual questions (e.g., “is a decimal number that starts with 0 smaller than 0?“; “Is a longer decimal larger than a shorter decimal?”). Three test forms (A, B and C) that were isomorphic and positionally

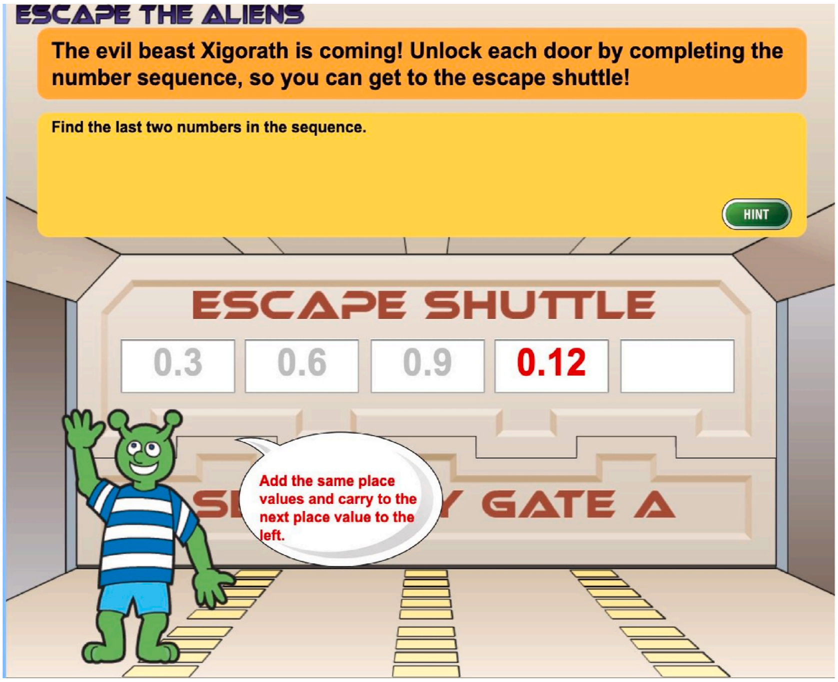

Fig. 6. *Hint* condition - The “Escape the Aliens” mini-game with an error message. Here a student makes a very common mistake.

counterbalanced across conditions were used.

5. Results

We report results according to each of our research questions. Since RQ1 examines the impact of context on all of the other research questions, we report all results as a 2 × 2 (Classroom/Remote x Hint/No-Hint) analysis. We report effect sizes (η^2_p and φ) for all main effects, interactions, and planned comparisons, and we interpret effects as small when $\eta^2_p < .06$ or φ = 0.1; medium when.06 $< \eta^2_p <$ .14 or φ = 0.3; and large when $\eta^2_p > .14$ or φ = 0.5 (Cohen, 1988; Olejnik & Algina, 2000).

First, a chi-square analysis assessing completion rates, Classroom (88.8%) versus Remote (56.5%), revealed a significant difference, X^2 (2, N = 393) = 51.3, $p < .001$, φ = 0.36, indicating that more students completed the materials in the classroom context. A 2 × 2 analysis of covariance (ANCOVA) comparing pretest scores by condition and Classroom/Remote context while controlling for school indicated that students in the *Hint* condition performed significantly worse than students in the *No-Hint* condition, F (4, 272) = 7.50, p = .007, η^2_p = .027. Students also performed significantly better on the pretest during the remote learning period compared to the classroom learning period, F (4, 272) = 104.39, $p < .001$, η^2_p = .28. There was no significant interaction, F (4, 273) = 0.94, p = .33, η^2_p = .003. Given the differences in pretest scores across conditions, we control for pretest as a covariate or assess learning gains in all analyses reported below.

Overall, students appeared to learn from the materials. A repeated-measures analysis of variance (ANOVA) indicated a significant, large effect of test comparing students’ pretest and posttest scores, F (1, 276) = 159.00, $p < .001$, η^2_p = .37. There was also a large effect comparing students pretest and delayed posttest scores, F (1, 276) = 149.09, $p < .001$, η^2_p = .35. There was no significant difference between posttest and delayed posttest scores, F (1, 276) = 0.05, p = .82, $\eta^2_p < .001$.

As a manipulation check, we compared combined error rates across the first opportunities of each of the five problem types across conditions. If the hints were helpful for getting the correct answer, students in the *Hint* conditions would be expected to demonstrate fewer errors on these first opportunities, regardless of whether they ultimately learned more in the *Hint* or *No-Hint* conditions. We conducted a 2 × 2 ANCOVA testing the effect of hint condition and context on the sum of errors students made on their first encounters with all five problem types, controlling for pretest and school. Results indicated a significant effect of hint condition, F (5, 271) = 23.25, $p < .001$, η^2_p = .079, with students who received hints (M = 16.71, SD = 15.12) making roughly half as many errors as students in the No-Hint condition (M = 32.27, SD = 43.46). There was no significant effect of context, F (5, 271) = 1.48, p = .23, η^2_p = .005, but there was a significant interaction, F (5, 271) = 10.03, p = .002, η^2_p = .036. Pairwise comparisons indicated a large effect among students in the classroom context, F (3, 147) = 23.02, $p < .001$, η^2_p = .14, and a medium effect among students in the remote context, F (3, 122) = 5.77, p = .018, η^2_p = .045. In both cases, students in the *Hint* conditions made significantly fewer errors than students in the *No-Hint* conditions.

RQ1. *How does the instructional context (i.e., classroom versus remote learning) impact playing of the game and learning with respect to all of the following research questions?*

Table 3
Mean scores on the pretest, immediate posttest, and delayed posttest, out of a possible 52 points, by condition and location.

	N	Condition	Pretest Mean (*SD*)	Posttest Mean (*SD*)	Delayed Mean (*SD*)	Pre-Post Gain (SD)	Pre-Delayed Gain (SD)	Time on game (minutes)
Classroom	67	*Hint*	21.40 (10.05)	26.73 (10.24)	26.12 (10.29)	5.33 (6.38)	4.72 (7.30)	61.60 (21.26)
	84	*No-Hint*	25.07 (10.32)	30.82 (9.54)	31.17 (9.84)	5.75 (6.74)	6.10 (6.50)	75.20 (44.46)
	151	Total	23.44 (10.33)	29.01 (10.02)	28.93 (10.32)	5.56 (6.57)	5.48 (6.88)	69.16 (36.59)
Remote	64	*Hint*	32.69 (11.25)	37.09 (9.31)	38.00 (8.91)	4.41 (6.81)	5.31 (7.16)	53.21 (28.06)
	62	*No-Hint*	34.02 (9.52)	38.39 (6.62)	37.94 (8.17)	4.37 (6.62)	3.92 (6.86)	58.61 (30.72)
	126	Total	33.34 (10.42)	37.73 (8.09)	37.97 (8.52)	4.39 (6.69)	4.63 (7.02)	55.87 (29.40)

Given that this research question crosses with all of the other questions, we report our results regarding this question within each of the research question subsections below.

RQ2. *Do students in the Hint condition learn more than students in the No-Hint condition?*

The mean scores on the pretest, immediate posttest, and delayed posttest by condition are shown in Table 3, along with gain scores and time spent playing the game in minutes. A 2 × 2 analysis of covariance (ANCOVA) assessing effects of the two learning conditions and Classroom versus Remote context on posttest performance and controlling for pretest and school indicated a marginally significant effect of *Hint* condition, $F\ (5, 271) = 3.17, p = .076, \eta^2_p = .012$, with students performing better in the *No-Hint* condition. Even when controlling for the pretest and school, students performed significantly better during the remote learning period compared with the classroom learning period, $F\ (5, 271) = 13.04, p < .001, \eta^2_p = .046$. There was no significant interaction, $F\ (5, 271) = 0.83, p = .36, \eta^2_p = .003$.

A 2 × 2 ANCOVA assessing effects of condition and Classroom/Remote context on delayed posttest performance and controlling for pretest and school indicated that there was no effect of learning condition, $F(5, 271) = 2.26, p = .13, \eta^2_p = .008$. Even when controlling for pretest and school, students performed significantly better during the Remote learning period compared with the Classroom learning period, $F(5, 271) = 18.40, p < .001, \eta^2_p = .0264$. There was also a significant interaction, $F(5, 271) = 5.96, p = .015, \eta^2_p = .022$.

We conducted a pairwise comparison controlling for pretest and school to interpret the interaction (Fig. 7). Within the Classroom group, there was an effect of hints, with students in the *No-Hint* condition performing significantly better on the delayed posttest than students in the *Hint* condition, $F(32, 147) = 6.46, p = .03612, \eta^2_p = .042$. Within the Remote group, there was no effect of the Hint condition on delayed posttest performance, $F(3, 122) = 0.32, p = .57, \eta^2_p = .003$.

A 2 × 2 ANCOVA comparing times spent playing the game and controlling for school and pretest showed a medium, significant effect of learning condition, $F(5, 271) = 16.17, p < .001, \eta^2_p = .056$, with students in the *Hint* condition taking less time than students in the *No-Hint* condition. There was no effect of context, $F(5, 271) = 0.56, p = .45, \eta^2_p = .002$. There was a marginally significant interaction, $F(5, 271) = 3.24, p = .073, \eta^2_p = .012$.

To better understand how hint usage might have led to these results, we conducted analyses of hint use among students in the *Hint* condition. Hint use was coded for each problem in the game on a scale from 0 to 3, with 0 meaning the student did not view any hints for a particular problem and 3 meaning the student viewed all 3 hints. A one-way ANOVA comparing students' total hint use in Classroom versus Remote contexts indicated a significant effect, $F(1, 130) = 19.08, p < .001, \eta^2_p = .13$, with students in the Classroom ($M = 51.09$, $SD = 42.86$) using significantly more hints than students in the Remote group ($M = 21.94$, $SD = 32.80$). In other words, students in the Classroom *Hint* condition used significantly more hints yet had worse outcomes compared to the Classroom *No-Hint* condition.

We also looked at learning curves for the various mini-game types (e.g., Buckets, Number Line). Learning curves provide an indication of how quickly students reduce their errors given "opportunities" (i.e., mini-games) to solve different problem types. The Y-axis of the curves in Figs. 8 and 9 show the number of students that made errors, at opportunity 1, 2, 3, and so forth, for two of the mini-

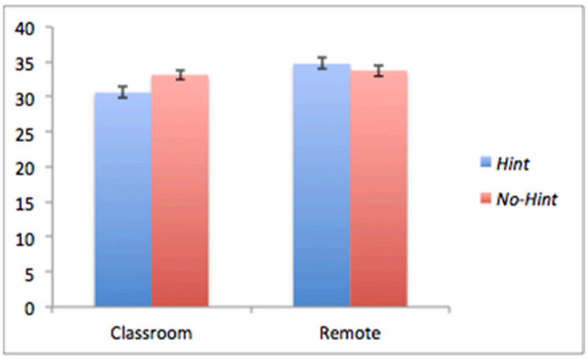

Fig. 7. Interaction effect at delayed posttest showing estimated marginal means controlling for pretest: *No-Hint* students performed significantly better than the *Hint* students in the Classroom but not Remotely.

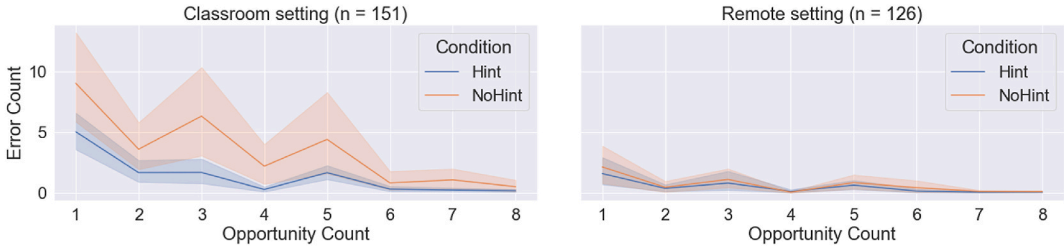

Fig. 8. Error count by opportunity in the *Buckets* mini-games.

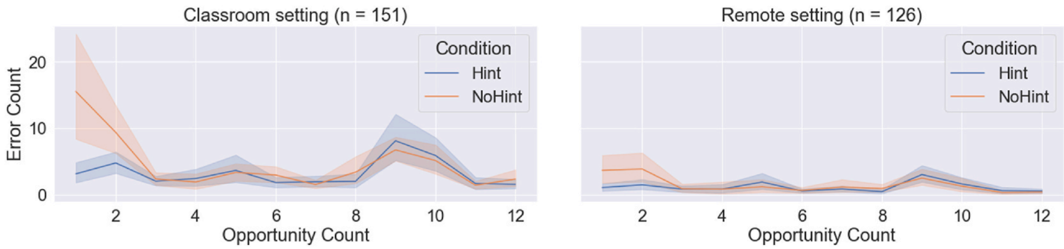

Fig. 9. Error count by opportunity in the *Number Line* mini-games.

games. Students in the *No-Hint* condition made more errors at early opportunities but the errors drop and the curves converge very soon after the initial opportunities, suggesting that error rates between the two hint conditions became similar after a few opportunities on each problem type. This pattern is consistent for all the mini-game types (see Appendix A).

In summary, our hypothesis that students who had access to hints and received feedback would learn more was not confirmed; in the classroom, receiving hints and error messages was associated with worse performance. Learning curve analyses suggested that while hints and error messages might reduce errors the first few times students encountered problem types, the accuracy benefits appeared to wear off with increased opportunities.

RQ3. Do female or male students benefit more from the hints?

Consistent with prior work (Hou et al., 2020; McLaren et al., 2017b; Nguyen et al., under review), an ANCOVA assessing the effects of gender on immediate posttest and controlling for pretest indicated a significant effect, $F(2, 274) = 10.35$, $p = .001$, $\eta^2_p = .036$, with female students performing better than male students, as can be seen in Table 4. Similarly, there was a significant effect on delayed posttest score controlling for pretest score, $F(2, 274) = 10.25$, $p = .002$, $\eta^2_p = .036$, with female students performing better than male students. Female students also spent significantly more time playing the game than male students, $F(1, 275) = 5.95$, $p = .015$, $\eta^2_p = .021$.

We examined potential interactions between gender and *Hint* condition using a series of 2 × 2 ANCOVAs controlling for pretest. Results predicting immediate posttest indicated that there was a main effect of gender, $F(4, 272) = 9.81$, $p = .002$, $\eta^2_p = .035$. There was no main effect of *Hint* condition, $F(4, 272) = 1.46$, $p = .23$, $\eta^2_p = .005$, and no interaction, $F(4, 272) = 2.49$, $p = .12$, $\eta^2_p = .009$. Similarly, when predicting delayed posttest, there was a main effect of gender, $F(4, 272) = 9.75$, $p = .002$, $\eta^2_p = .035$. There was no main effect of *Hint* condition, $F(4, 272) = 0.84$, $p = .36$, $\eta^2_p = .003$, and no interaction, $F(4, 272) = 1.72$, $p = .19$, $\eta^2_p = .006$.

We also examined potential interactions between gender and Classroom/Remote context using a series of 2 × 2 ANCOVAs controlling for pretest. Results predicting immediate posttest indicated that there was a main effect of gender, $F(4, 272) = 8.83$, $p = .003$, $\eta^2_p = .031$, and a significant main effect of Classroom/Remote context, $F(4, 272) = 5.36$, $p = .021$, $\eta^2_p = .019$. There was also a significant interaction, $F(4, 272) = 4.42$, $p = .036$, $\eta^2_p = .016$. We conducted a planned, pairwise comparison to interpret the

Table 4
Mean scores on the pretest, posttest, and delayed posttest by gender.

	N	Pretest Mean (*SD*)	Posttest Mean (*SD*)	Delayed Mean (*SD*)	Pre-Post Gain (SD)	Pre-Delayed Gain (SD)	Time on game (minutes)
Female	136	27.84 (11.35)	34.03 (9.18)	34.17 (9.81)	6.19 (6.91)	6.33 (6.90)	68.16 (38.12)
Male	141	28.05 (11.62)	31.96 (10.96)	31.95 (11.13)	3.91 (6.18)	3.90 (6.80)	58.25 (29.03)

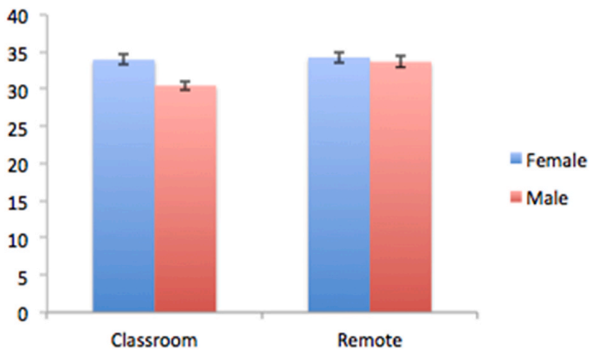

Fig. 10. Interaction effect at immediate posttest showing estimated marginal means controlling for pretest: Female students performed significantly better than male students in the Classroom but not Remotely.

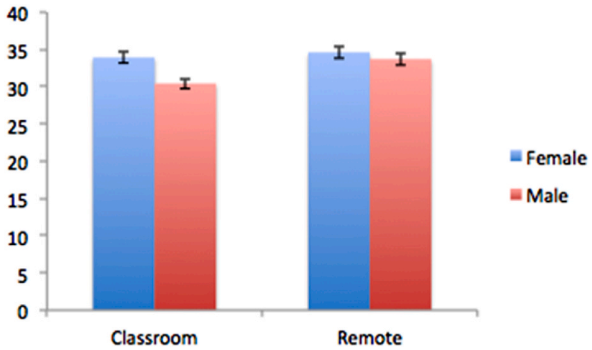

Fig. 11. Interaction at delayed posttest showing estimated marginal means controlling for pretest: Female students performed marginally better than male students in the Classroom but not Remotely.

interaction. Within the classroom group, there was a significant effect of gender, with female students performing better on the immediate posttest than male students, $F(2, 148) = 13.96$, $p < .001$, $\eta^2_p = .086$ (Fig. 10). However, within the remote group, there was no effect of gender on posttest performance, $F(2, 123) = 0.48$, $p = .49$, $\eta^2_p = .004$.

When predicting delayed posttest, there was a main effect of gender, $F(4, 272) = 8.86$, $p = .003$, $\eta^2_p = .032$, and a main effect of Classroom/Remote context, $F(4, 272) = 5.87$, $p = .016$, $\eta^2_p = .021$. There was a marginally significant interaction, $F(4, 272) = 3.07$, $p = .081$, $\eta^2_p = .011$, again, with female students performing better than male students in the classroom (See Fig. 11).

In summary, our hypothesis about gender was confirmed. As in prior studies, female students overall benefited more from the game than male students. In addition, access to hints and feedback equally impacted female and male students, as the gender effect remained the same regardless of whether students were in the *Hint* or *No-Hint* condition. Yet we uncovered an interesting difference between the classroom and remote versions of the study: In the classroom, female students benefited more from the game than males, as demonstrated on the immediate posttest and in previous findings; yet, in the remote study female students did not perform better than male students on the immediate posttest.

RQ4. In the Hint condition, how do hint behaviors relate to prior knowledge and learning outcomes?

In order to test our hypothesis that higher prior knowledge students would use hints more productively in the *Decimal Point* than lower prior knowledge students and thereby have greater learning gains, we first needed to define what we mean by unproductive hint use. Based on prior work regarding productive help-seeking behaviors (Roll et al., 2011), we broke it down into the following metrics of hint use:

Metric 1: *Using hints before making an attempt.* While a student who is unsure of what to do should probably request help, it is also the case that making an attempt at solving problems before requesting explanatory hints is likely to lead to better learning, even if the student is unsure of how to proceed (Pressley et al., 1992).
Metric 2: *Gaming the system by reading hints quickly.* We assessed how often students hurriedly moved to the bottom-out hint, without carefully reading each hint. This is hypothesized to be the more unproductive of the two hint behaviors we assessed, and one that can reflect a host of negative affective states (e.g., confusion, frustration, cognitive disengagement; Baker, 2011; Baker et al., 2008).

To assess Metric 1 (“*Using hints before making an attempt*”), we analyzed how often students requested a hint before trying to solve

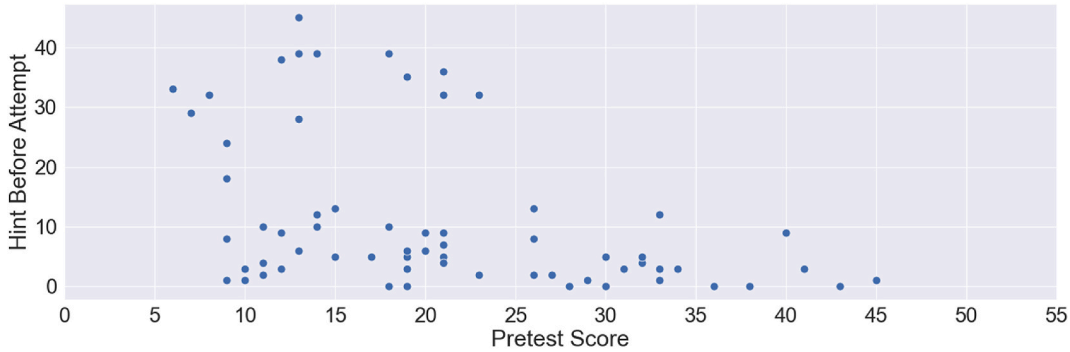

Fig. 12. Plotting of prior knowledge (performance on pretest) versus hint usage before first attempt for the Classroom students.

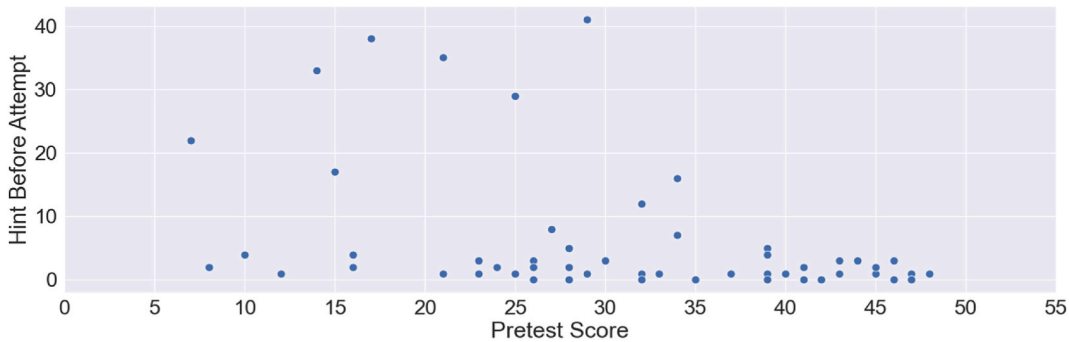

Fig. 13. Plotting of prior knowledge (performance on pretest) versus hint usage before first attempt for the Remote students.

mini-game problems. Specifically, we checked the number of mini-game instances in which students in the *Hint* condition asked for a hint before their first attempt at solving a problem. Figs. 12 and 13 show the Classroom versus Remote context plotting of "hint before first attempt" against pretest scores for all of the students who had access to hints. As can be seen, the students who did better on the pretest, in both the Classroom and Remote situations, tended to ask for fewer hints before making an attempt. This result is significant for both the Classroom ($r = -0.441, p < .001$) and Remote ($r = -0.436, p < .001$) groups. A one-way ANCOVA comparing hint usage before attempt in the classroom and at home and controlling for pretest was not significant, $F(1, 128) = 0.04, p = .84$. This suggests that students did not tend to ask for more hints before making an attempt in one setting compared to the other (Classroom vs. Remote), when controlling for prior knowledge.

To analyze Metric 2 ("*Gaming the system by reading hints quickly*"), we used reading time research to analyze the time it should take students to read hints before moving to the next hint or trying to solve a problem. Using an average reading time of 4 words per second (Just & Carpenter, 1987), we analyzed all hints requested by students, checking whether the time gap between one hint request and the next was smaller than (hint word length/4) seconds (The average length of hints in the game is 18.64 words). Such an analysis uncovers the number of times a student doesn't read hints deliberately. Figs. 14 and 15 show the Classroom versus Remote context plotting of "reading hints too quickly" against pretest scores for all of the students in the *Hint* condition. As can be seen, the students who did better on the pretest, in both the Classroom and Remote situations, tended to less frequently read hints too quickly. This result is significant for both the Classroom ($r = -0.573, p < .001$) and Remote ($r = -0.538, p < .001$) groups. A one-way ANCOVA comparing

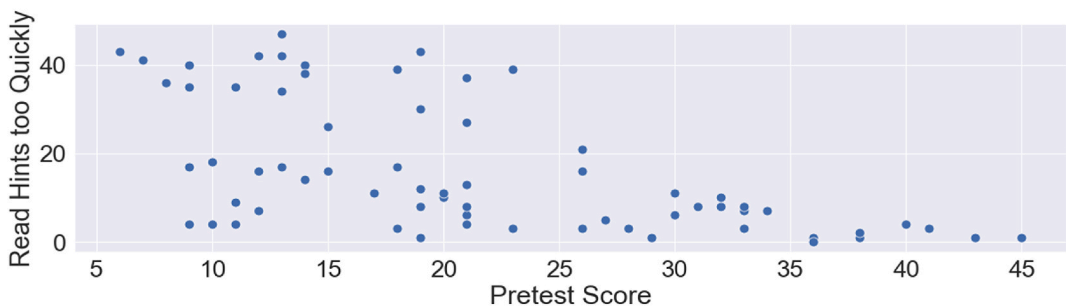

Fig. 14. Plotting of prior knowledge (performance on pretest) versus Read Hints too Quickly by the Classroom students.

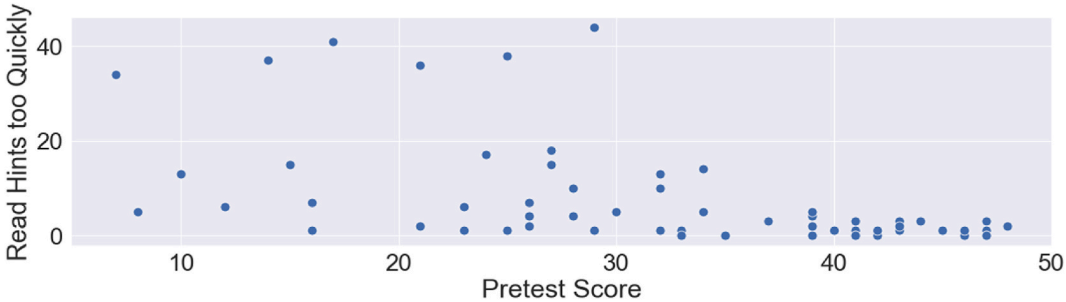

Fig. 15. Plotting of prior knowledge (performance on pretest) versus Read Hints too Quickly by the Remote students.

Table 5
Correlations between hint metrics and learning gains.

	Pretest-to-Posttest Gain	Pretest-to-Delayed Posttest Gain
Classroom context		
Metric 1: Hint Before Attempt	−0.37[a]	−0.30[a]
Metric 2: Read Hints Too Quickly	−0.35[a]	−0.27[a]
Remote context		
Metric 1: Hint Before Attempt	−0.04	0.11
Metric 2: Read Hints Too Quickly	−0.00	0.12

[a] Indicates *p* significant after applying Benjamini-Hochberg correction for multiple comparisons where false discovery rate $Q = .05$ (Benjamini & Hochberg, 1995).

rates of reading hints too quickly in the classroom and at home, controlling for pretest, was not significant, $F(1, 128) = 0.38$, $p = .54$. This suggests that students did not tend to read hints more quickly in one setting compared to the other (Classroom vs. Remote), when controlling for prior knowledge.

To assess the relations between these hint use metrics and learning outcomes, we examined correlations between each hint behavior and learning outcomes on the posttest and delayed posttest (Table 5). We used learning gains to account for prior knowledge and assessed correlations separately for classroom and remote contexts. While both hint behaviors were significantly, negatively correlated with learning gains in the classroom context, neither were significantly correlated with learning gains in the remote context.

In summary, our prediction that students with higher prior knowledge would use hints more productively (Metrics 1 and 2) was confirmed. Additionally, our prediction that Metrics 1 and 2 represented unproductive learning behaviors (i.e., associated with worse learning outcomes) was supported by results indicating significant, negative correlations with learning gains in the Classroom context. In contrast, neither of the hint behaviors were associated with learning gains in the Remote context, which is consistent with the lack of effect of *Hint* condition on learning in RQ2.

6. Discussion

While our original intent was to explore how on-demand hints and error messages help (or hurt) students in the context of a digital learning *Decimal Point*, we ended up additionally exploring the critical question of how instructional context, in the form of working with online materials in a classroom versus working with those same materials remotely at home, impacts students' behavior and learning. In the following we discuss these results and put forward possible explanations and suggestions for further investigation.

First, why did we see significantly different completion rates between classroom use (88.8%) and home use (56.5%)? This is likely explained by the difference in the level of guidance between the classroom and home. Students in the classroom were monitored closely by experimenters and teachers, including being prompted to continue working when they appeared disengaged and given encouragement to finish their work. Conversely, students at home had no such guidance and encouragement, not only because their teachers were not physically present with them as they worked, but also because the stresses of the COVID-19 situation necessitated leniency in completing work at all three of the schools where students worked at home. Furthermore, the lack of expectation of completing their work likely explains another result: Students who worked remotely learned significantly more than those who worked in the classroom, even taking into account the pretest as a covariate. Since students were not monitored as closely in the Remote group and also not strictly required to complete the materials, many dropped out. As might be expected, the higher performing students were more likely to persevere and complete all of the materials, as shown by the significantly higher pretest values in the Remote vs. Classroom context. Thus, it is perhaps unsurprising that in the Remote study the students who stuck it out and completed everything learned significantly more than students in the Classroom study.

Second, why did we see students in the *No-Hint* condition do better, in the classroom on the delayed posttest, than students in the *Hint* condition? From a cognitive engagement (Chi & Wylie, 2014) and constructivist perspective (Applefield et al., 2001) this result makes sense. We note that when we look at learning curves (Figs. 8 and 9, also Appendix A), students in the *No-Hint* condition initially do a little worse than students in the *Hint* condition, i.e., they make more errors, but eventually the *No-Hint* condition students reduce their errors to approximately the level of the *Hint* condition students and consequently perform better on the delayed posttest. In other words, the *No-Hint* condition students' knowledge appears to be more robust, as indicated by knowledge that transfers to the delayed posttest (Barnett & Ceci, 2002), likely because constructing their own understanding involves deeper learning processes such as knowledge retrieval, self-explanation, and knowledge revision that might be disrupted by overuse of hints in the *Hint* condition. The idea that overusing hints might hurt robust learning is also supported by the result showing that students in the Remote *Hint* condition used significantly fewer hints than students in the Classroom *Hint* condition and did not perform worse compared to the Remote *No-Hint* condition. While it might be that students in the Remote context had less need for hints, given their higher pretest scores, they nevertheless scored well below the maximum possible score on the pretest. Remote students in the *Hint* condition made significantly fewer initial errors than students in the *No-Hint* condition, suggesting that hints were still useful to them. Our result is also in line with the results of O'Rourke et al. (2014), in which gameplaying students in a no-hint condition learned more than gameplaying students in four other hint conditions. This result also provides another data point for the assistance dilemma (Koedinger & Aleven, 2007): Students in our particular *Hint* condition appear not to have struggled enough to learn.

Third, why did we get the result in which female students in the classroom performed significantly better on the immediate posttest than male students in the classroom, yet there was no effect of gender remotely on immediate posttest performance? This effect was primarily driven by the fact that girls' performance appears to be approximately the same at home and in the classroom, while boys' performance was much better at home compared to the classroom. These results in the classroom are consistent with girls' general tendency to outperform boys in classroom settings (Dwyer & Johnson, 1997; Entwisle et al., 1997), which has been attributed at least in part to girls' greater likelihood of holding mastery goals and less frequent behavioral disruptions (Kenney-Benson et al., 2006). It may be that these effects are diminished at home, where students have more autonomy and can work outside of typical classroom structures. Future work should further explore how girls' and boys' behaviors and affective experiences differ in the classroom context.

With respect to our research question about prior knowledge (RQ4: "*In the Hint condition, how does students' prior knowledge relate to how they use the hints?*"), we found that students with lower prior knowledge engaged in a number of negative hint use behaviors, including asking for hints before making an attempt and moving to the next hint without taking enough time to read the previous hint. In looking more closely at the students with lower prior knowledge, we saw that the students who took their time to read each hint and attempted to solve the problem along the way also learned more in the Classroom context, as shown by the negative correlations between unproductive hint behaviors and learning gains. In the Remote context, unproductive hint behaviors were not associated with learning gains. Although students with lower prior knowledge appeared to engage in unproductive hint use as much in the Remote context as in the Classroom context, as indicated by the lack of significant differences in behaviors across contexts when controlling for pretest, there tended to be fewer students with lower prior knowledge in the Remote context. As a result, there may not have been enough students engaging in high rates of unproductive behaviors to see a significant relation with learning outcomes. Since we didn't explicitly manipulate hint behaviors in our study, we can't say whether these hint behaviors *caused* the better learning outcomes in the classroom or whether it's a reflection of something different about those students (i.e., that they are more engaged, more conscientious, etc.). A future study could be designed to force students to follow positive hint behaviors (e.g., students can't access the next hint until 10 seconds and 1 attempt after the previous hint) to see whether we can lift all learners up by encouraging fewer unproductive hint behaviors. Such a study would shift the research question more to what happens with productive versus unproductive hint behaviors rather than how prior knowledge relates to hint usage. Of course, a danger of implementing such an approach could be to frustrate non-gaming students who don't need such a time delay, so it may not necessarily promote positive hint behaviors. Attempts to implement such an approach in other educational technology, for instance intelligent tutoring systems, have only been done with small sample sizes and with indeterminate results (Murray & VanLehn, 2005).

Interestingly, as illustrated by the learning curves of Figs. 8 and 9, we found that while hints seem to be beneficial for the first attempt made by students on individual problems, the *No-Hint* and *Hint* students eventually achieved similar error rates after the initial few problems. Thus, contrary to our predictions, it doesn't appear that hints had a positive effect. It turns out, at least in the case of students in this study, that the additional support students received for learning in the *Hint* condition was not helpful; in fact, students in the *No-Hint* condition in the classroom actually learned more than their fellow students who received hints. Instead, we may have provided a too easy way to get answers and thus undermined their learning. It may also be that the utility of hints depends on students' prior knowledge levels. If too little prior knowledge leads to unproductive hint use and a high level of prior knowledge makes hints unnecessary, then students might benefit most from hints when they have a moderate level of prior knowledge.

In closing, we mention the limitations to this study and our findings. First, because we lost so many of the students who worked from home – likely losing more students with less motivation, prior knowledge, or family resources to support them through the challenges of working remotely during the pandemic – we also lost the opportunity to see how the game and its hints might have helped a more representative sample of students at home. Second, the low remote student participation, disruption of our classroom data collection, and introduction of a second factor to the study reduced the statistical power to detect effects. Third, because the "instructional context" aspect of the study was a reaction to the pandemic, rather than a planned situation, we were unable randomize assignment to in-school and at-home use of the game. In particular, differences in the schools and students who completed the work in the classroom vs. remotely could have impacted the results, though we have attempted to address these differences by controlling for pretest and school. Relatedly, since the remote learning situation during the pandemic, and particularly the early months of full lockdown, clearly differed from intentional, planned remote learning, we must be cautious about how these results would generalize to intentional remote learning contexts.

7. Conclusions

Although we started this research as an exploration into an important, understudied topic – do hints and error feedback help in the context of the digital learning game? – we fortuitously stumbled upon perhaps a more important topic during the course of the study: How does the instructional context – classroom versus remote learning – impact learning with a digital learning game? The COVID-19 pandemic of 2020–2021 pushed this question on us. We were able to uncover some interesting differences between students working in school versus at home, as well as with and without hints. First, we found that the different instructional settings led to significantly different completion rates, likely due to the stricter control and supervision students received in the classroom. Second, the *Hint* and *No-Hint* versions of the game led to different results in the classroom versus at-home. Students in the *No-Hint* condition performed better in the classroom, while there was no difference between conditions at home. We speculate that this may have occurred because the students who didn't have hints struggled in a productive way, in line with cognitive engagement and constructivist theories, while students who had hints in the classroom tended to use them in unproductive ways that may have disrupted engagement and knowledge construction. In the remote condition, where students tended not to use hints unproductively, hints did not have the same detrimental effect. Also, classroom students in the *Hint* condition used significantly more hints than the students at home in the *Hint* condition. This indicates that a greater use of hints does not benefit the students. Finally, as we have seen in prior *Decimal Point* studies, we observed a gender effect, specifically with female students outperforming male students in the classroom but not remotely. This may have been due, at least in part, to prior research showing that female students often outperform male students in classroom contexts. These findings both illuminate differences between learning in the classroom and at home and point to potential future work.

Credit author statement

Bruce M. McLaren – Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **J. Elizabeth Richey** – Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Huy Nguyen** – Methodology, Software, Formal analysis, Data curation. **Xinying Hou** – Methodology, Formal analysis, Data curation.

Acknowledgments

This work was supported by the National Science Foundation Award #DRL-1661121. The opinions expressed are those of the authors and do not represent the views of NSF. Thanks to Jodi Forlizzi, Jon Star, Michael Mogessie Ashenafi, Scott Herbst, Craig Ganoe, Darlan Santana Farias, Rick Henkel, Patrick Bruce Gonçalves McLaren, Grace Kihumba, Kim Lister, Kevin Dhou, John Choi, and Jimit Bhalani, all of whom made important contributions to the design, development, and early experimentation with the Decimal Point game. Special thanks to Jon Star who provided input on the hint and error messages used in this study. Special thanks also to Rosta Farzan who first had the idea to do gender analyses of Decimal Point use.

Appendix A

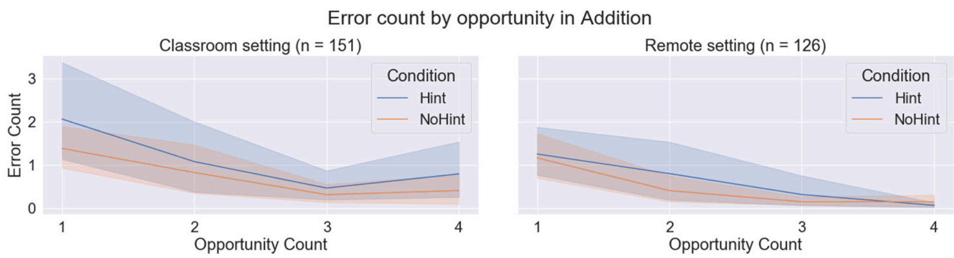

Fig. 16.

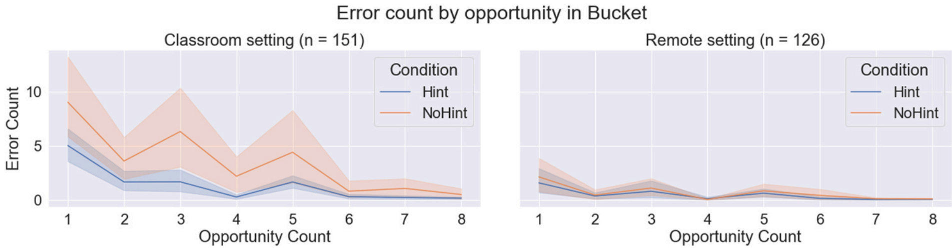

Fig. 17.

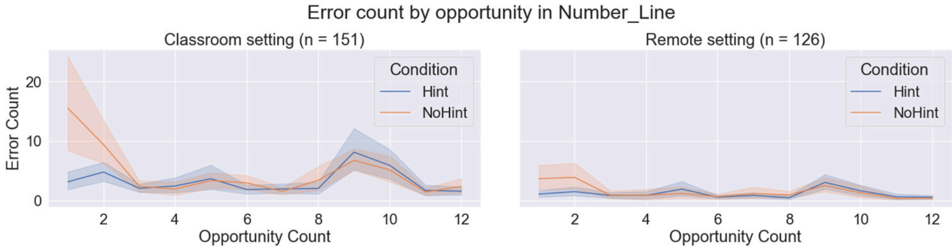

Fig. 18.

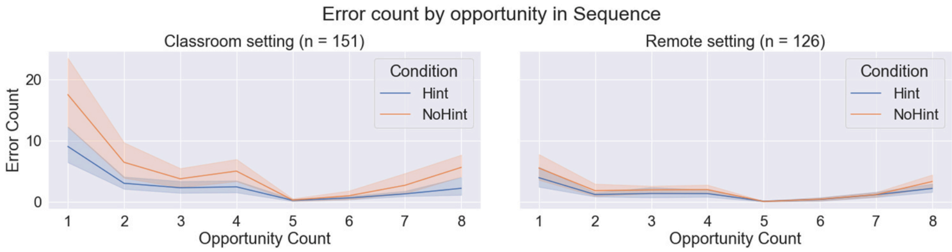

Fig. 19.

Fig. 20.

References

Aleven, V., & Koedinger, K. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *The Proceedings of the 5th international Conference on intelligent tutoring systems, ITS 2000* (pp. 292–303). Berlin: Springer Verlag, 2000. Best Paper Award ITS 2000.

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.

Aleven, V., McLaren, B. M., Sewall, J., Van Velsen, M., Popescu, O., Demi, S., & Koedinger, K. R. (2016a). Example-tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education, 26*(1), 224–269.

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016b). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 26*(1), 205–223. https://doi.org/10.1007/s40593-015-0089-1

Andersen, M. S., & Makransky, G. (2021). The validation and further development of the multidimensional cognitive load scale for physical and online lectures (MCLS-POL). *Frontiers in Psychology, 12*, 713.

Anderson, J. R. (1993). *Rules of the mind*. Psychology Press, 1993.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

Applefield, J. M., Huber, R. L., & Moallem, M. (2001). Constructivism in theory and practice: Toward a better understanding. *High School Journal*. https://doi.org/10.2307/40364404

Artino, A. R. (2009). Online learning: Are subjective perceptions of instructional context related to academic success? *Internet and Higher Education, 12*, 117–125, 2009.

Baker, R. S. J.d. (2011). Gaming the system: A retrospective look. *Philippine Computing Journal, 6*(2), 9–13.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612.

Bell, M, Bretzlauf, J, Dillard, A, et al. (2012). *Everyday Mathematics at Home Fifth Grade Student Resources*. The University of Chicago School Mathematics Project (UCSMP) Chicago: Everyday Learning Corporation.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*, 289–300.

Bouvier, P., Lavoué, E., Sehaba, K., & George, S. (2013). Identifying learner's engagement in learning games: A qualitative approach based on learner's traces of interaction. In *5th international conference on computer supported education (CSEDU 2013), may 2013, aachen, Germany* (pp. 339–350).

BrightBytes. (2020). *2020 remote learning survey research results*. https://www.brightbytes.net/rls-research.

Burgers, C., Eden, A., van Englenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94–103. https://doi.org/10.1016/j.chb.2015.01.038, 2015.

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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NK: Lawrence Erlbaum Associates.

Conati, C., Jaques, N., & Muir, M. (2013). Understanding attention to adaptive hints in educational games: An eye-tracking study. *International Journal of Artificial Intelligence in Education, 23*, 131–161. https://doi.org/10.1007/s40593-013-0002-8

Costello, B., & Edmonds, E. (2007). A study in play, pleasure and interaction design. In *Proceedings of the 2007 conference on designing pleasurable products and interfaces* (pp. 22–25). ACM.

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

Dohn, N. B., Hansen, S. B., & Klausen, S. H. (2018). On the concept of context. *Education Sciences, 8*, 111. https://doi.org/10.3390/educsci803011, 2018.

Drey, T., Jansen, P., Fishbach, F., Frommel, J., & Rukzio, E. (2020). Towards progress assessment for adaptive hints in educational virtual reality games. In *Late breaking works of CHI 2020 (SIGCHI conference on human factors in computing systems), apr. 2020*. ACM.

Dwyer, C., & Johnson, L. (1997). Grades, accomplishments, and correlates. In W. Willingham, & N. Cole (Eds.), *Gender and fair assessment* (pp. 127–156). Mahwah, NJ: Erlbaum.

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

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview Press.

Ferdig, R. E., Baumgartner, E., Hartshorne, R., Kaplan-Rakowski, R., & Mouza, C. (2020). *Teaching, technology, and teacher education during the COVID-19 pandemic: Stories from the field*. Association for the Advancement of Computing in Education (AACE). Retrieved April 26, 2021 from https://www.learntechlib.org/p/216903/.

Forlizzi, J., McLaren, B. M., Ganoe, C. H., McLaren, P. B., Kihumba, G., & Lister, K. (2014). *Decimal Point: Designing and developing an educational game to teach decimals to middle school students, (ECGBL-2014)*. pp. 128–135). Reading, U.K: Academic Conferences and Publishing International Limited.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (1st ed.). New York: Palgrave Macmillan.

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 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. 12163)*. Cham: Springer.

Harp, S., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434.

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

Hattie, & Timperley. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

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.), *Lecture Notes in Computer Science, 6094* pp. 346–348). Berlin: Springer.

Johnson, C. I., & Priest, H. A. (2014). *The feedback principle in multimedia learning. The Cambridge Handbook of Multimedia Learning.* Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.023

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Allyn and Bacon, Inc.

Kenney-Benson, G., Pomerantz, E., Ryan, A., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*, 11–26.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review, 19*(3), 239–264.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79–97.

Lee, M. J., & Chiou, J. (2020). Animated hints help novices complete more levels in an educational programming game. *Journal of Computing Sciences in Colleges, 35*(8). https://dl.acm.org/doi/10.5555/3417639.3417652.

Lockee, B. B. (2021). Shifting digital, shifting context:(re) considering teacher professional development for online and blended learning in the COVID-19 era. *Educational Technology Research & Development, 69*(1), 17–20.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

Makransky, G., Mayer, R. E., Veitch, N., Hood, M., Christensen, K. B., & Gadegaard, H. (2019). Equivalence of using a desktop virtual reality science simulation at home and in class. *PLoS One, 14*(4), Article e0214944.

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

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

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)*. Berlin: Springer.

Means, B., Peters, V., Neisler, J., Wiley, K., & Grifths, R. (2021). *Lessons from remote learning during COVID-19*. https://doi.org/10.51388/20.500.12265/116. Digital Promise.

Melero, J., Hernández-Leo, D., & Blat, J. (2012). Considerations for the design of mini-games integrating hints for puzzle solving ICT-related concepts. In *The Proceedings of the IEEE international Conference on advanced learning technologies* (pp. 138–140). https://doi.org/10.1109/ICALT.2012.60

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

Murray, C., & VanLehn, K. (2005). Effects of dissuading unnecessary help requests while providing proactive help. In C.-K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Proceedings of the 12th international conference on artificial intelligence in education, AIED 2005* (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 C. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018). LNAI 10947 (pp. 338-351)*. Berlin: Springer.

Nguyen, H.A, Hou, X., Richey, J.E., & McLaren, B.M. (under review). Uncovering the gender effect across multiple studies of a digital learning game. Submitted to the International Journal of Game-Based Learning.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25* (3), 241–286.

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

Prensky, M. (2006). *Don't bother me, Mom, I'm learning! How computer and video games are preparing your kids for 21st century success and how you can help*. St. Paul: Paragon House.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*(1), 91–109.

Reich, J., Buttimer, C. J., Fang, A., Hillaire, G., Hirsch, K., Larke, L., Littenberg-Tobias, J., Moussapour, R., Napier, A., Thompson, M., & Slama, R. (2020). *Remote learning guidance from state education agencies during the COVID-19 pandemic: A first look*. Retrieved from osf.io/k6zxy/.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and instruction, 21*(2), 267–280.

Sadeghi, M. (2019). A shift from classroom to distance learning: Advantages and limitations. *International Journal of Research in English Education, 4*(1), 80–88.

Squire, K., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*(7), 5–33.

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). Utrecht: PME.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987.

Tessmer, M., & Richey, R. C. (1997). The role of context in learning and instructional design. *ETR&D, 45*, 85–115. https://doi.org/10.1007/BF02299526, 1997.

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

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

Wang, Q., & Huang, C. (2018). Pedagogical, social and technical designs of a blended synchronous learning environment. *British Journal of Educational Technology, 49* (3), 451–462. https://doi.org/10.1111/bjet.12558

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

Wen, Y., Gwendoline, C. L. Q., & Lau, S. Y. (2021). ICT-supported home-based learning in K-12: A systematic review of research and implementation. *TechTrends*, 1–8.

Winne, P. H. (2015). Self-regulated learning. In *The international Encyclopedia of the social & behavioral sciences* (2nd ed.).

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

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review, 2*, 173–201. https://doi.org/10.1007/BF01322178