

Computers & Education

How and for Whom does Educational Technology Support Learning across Instructional Context?

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Distance education and online learning, Games, Gender studies, Pedagogical issues, Hints
Corresponding Author:	Bruce McLaren Carnegie Mellon University Pittsburgh, UNITED STATES
First Author:	Bruce McLaren
Order of Authors:	Bruce McLaren J. Elizabeth Richey, PhD Huy Nguyen Xinying Hou
Abstract:	<p>There is a tendency to think of the impact of educational technology in a vacuum, i.e., that the results of experiments are somehow absolute. However, it is clear that the instructional context in which educational technology is used can make a huge difference to 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 or not. 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 learning, we instead found ourselves exploring a new question: How did the game impact student learning in classrooms at school versus remotely at home? After two of five schools had participated in our classroom experiment, we had to switch to at-home use of the Internet-based game for the final three schools due to the pandemic. This led to some new and interesting findings. The different 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 remote 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 remotely. In addition, students in the Hint condition used significantly more hints in the classroom than they did remotely. There was also a significant effect of gender in the classroom, with female students performing better on the immediate posttest than male students, but with no effect of gender remotely on immediate posttest performance. We explored these findings by performing a variety of post-hoc analyses to better understand students' learning processes and gameplay behaviors. In summary, our study clearly illustrates how sensitive educational technology can be to instructional context, yet just cracks open the door to much more research on this topic.</p>
Suggested Reviewers:	<p>James Lester, PhD Professor, North Carolina State University lester@ncsu.edu James has done related digital learning games research</p> <p>Matthew Easterday, PhD Northwestern University easterday@northwestern.edu Matt has done related digital learning games research.</p>

April 6, 2021

Dear Drs. Bennett, van Braak, Heller, Nussbaum, and Tsai:

We would like to submit the attached manuscript, titled, "How and for Whom does Educational Technology Support Learning across Instructional Contexts?," for consideration to be published in Computers & Education. The manuscript investigates and compares the use of a digital learning game used both in the classroom and at home. It further investigates the inclusion of hints and error messages in the context of the game. Results have implications for online learning, not only with games but with other forms of educational technology, as well.

Thank you very much for your time and consideration.

Sincerely,

The authors

How and for Whom does Educational Technology Support Learning across Instructional Contexts?

By Bruce M. McLaren, J. Elizabeth Richey, Huy Nguyen, and Xinying Hou
Carnegie Mellon University

Corresponding Author:
Bruce M. McLaren
2602I Newell-Simon Hall
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
bmclaren@cs.cmu.edu

Indicate clearly if color should be used for any figures in print:
Color should be used in all figures

Abstract

There is a tendency to think of the impact of educational technology in a vacuum, i.e., that the results of experiments are somehow absolute. However, it is clear that the *instructional context* in which educational technology is used can make a huge difference to 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 or not. 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 learning, we instead found ourselves exploring a new question: How did the game impact student learning in classrooms at school versus remotely at home? After two of five schools had participated in our classroom experiment, we had to switch to at-home use of the Internet-based game for the final three schools due to the pandemic. This led to some new and interesting findings. The different 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 remote 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 remotely. In addition, students in the *Hint* condition used significantly more hints in the classroom than they did remotely. There was also a significant effect of gender in the classroom, with female students performing better on the immediate posttest than male students, but with no effect of gender remotely on immediate posttest performance. We explored these findings by performing a variety of post-hoc analyses to better understand students' learning processes and gameplay behaviors. In summary, our study clearly illustrates how sensitive educational technology can be to instructional context, yet just cracks open the door to much more research on this topic.

Keywords: Distance education and online learning, Games, Gender studies, Pedagogical issues, Hints

1. Introduction

The instructional context in which students learn can make a huge difference to their learning (Tessmer & Richey, 1997, Greeno, 2011). Yet, the world of educational technology, both research and application, has been 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 & Bellezza, 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. In this work, we take a step toward correcting this shortcoming of educational technology research 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 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. 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 schools, when the pandemic forced students across the U.S. to continue learning

from home. Thus, we conducted the study at the final three 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 learning and at-home learning with educational technology are in many ways stark and clear. In the 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 to 90 minutes 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, 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 and other parents.

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, Adams, et al., 2017), 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 within 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). 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 a 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., 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* (McLaren, Farzan, et al., 2017; Hou et al., 2020; 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 (i.e., 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 far from 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

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. This is especially true 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.

In this article we don’t intend to engage in the now somewhat dated debate between cognitivists and constructivists regarding context. Nevertheless, it is worth noting that not all researchers would accept context as a critical component to learning (e.g., Anderson et al., 1996). The cognitivist view, which has informed much of educational technology research, emphasizes the cognitive processes related to the content being learned and wouldn’t expect students to learn any differently from, for instance, solving algebra problems in a classroom to solving and learning those same problems at home. Conversely, constructivism proposes that learners construct interpretations of their *experiences*, through a meaning-making search and through social negotiation (Applefield et al., 2001). The “meaning-making search” can occur in a variety of circumstances and social situations, according to constructivists. Without weighing in on the old debate, we simply say that we are empirically exploring context -- the classroom versus at home -- to see what we find when the same materials are used in each of these contexts.

In fact, there has been limited research regarding the efficacy of learning with educational technology in the classroom versus at home. For instance, a meta-analysis of intelligent tutoring systems (ITSs) by

1
2
3
4 Steenbergen-Hu and Cooper (2013) identified studies of classroom ITS use and homework ITS use, with
5 slightly larger effect sizes for homework ITSs. However, only two of these studies looked at ITSs used at
6 home to complete homework, and these focused exclusively on homework use, meaning they didn't test
7 context as a moderator or statistically compare effect sizes for classroom versus home use. Thus, while
8 we may be able to conclude that ITSs can promote learning when used for assignments at home, the
9 relative *differences* between classroom and home use are not at all clear from these results.

11 More practically and more recently, there has been examination of remote learning, due to the COVID-
12 19 pandemic (Brightbytes, 2020; Reich et al., 2020). In a survey of nearly 50,000 students, 11,889
13 teachers, 33,182 parents and 580 school principals done by the International Society for Technology in
14 Education (ISTE) and Brightbytes (Brightbytes, 2020) between April and June of 2020 -- during the first
15 months of the pandemic -- some clear differences were identified between students studying remotely at
16 home and in the classroom. For instance, only a very slim majority of students said they consistently
17 receive the help they need (52%), while 21% students said they "sometimes" or "never" receive teacher
18 help. Particularly for students who struggle and need encouragement to work, whether in using learning
19 technology or not, this is a significant contextual difference. Furthermore, 30% of students said they
20 "sometimes" or "never" had a quiet space to work, yet only 13% of parents said the same about their
21 students. Additionally, nearly all U.S. state agencies emphasized issues of equity with respect to access
22 to online options and encouraged asynchronous over synchronous learning, due to technology access
23 constraints at home (Reich et al., 2020). While asynchronous learning certainly made practical sense,
24 given the constraints, it would also make it difficult for teachers to monitor their students' activity and
25 provide the real-time guidance so readily available in the classroom.

28 Regarding our original research questions of hints versus no hints in the context of the *Decimal Point*
29 game, there is substantial research regarding feedback and support in learning more generally and,
30 specifically, regarding feedback and support in the context of digital learning games. Hattie and Timperley
31 (2007), in a classic and thorough article about the value of feedback in learning (without technology),
32 determined that the type of feedback and the way it is given to learners can make a significant difference
33 to learning. Their research, which cited a wide range of studies (e.g., Kluger & DeNisi, 1996; Kulik &
34 Kulik, 1988), revealed that immediate and delayed feedback vary on their beneficial effects based on the
35 task level; immediate feedback tends to lead to more benefits in task acquisition, while delayed feedback
36 tends to better support strategy learning. Both positive and negative feedback can have beneficial effects
37 on learning, depending on whether that feedback is about the task, processing the task, or self-regulation.

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

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 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 x 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 any of the other conditions. While their huge N 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 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.

The only digital learning game study that, to our knowledge, actually demonstrated learning benefits with hints and feedback was that done by Easterday and colleagues (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 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 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 Figure 1, is a single-player game with an amusement park metaphor targeted at middle-school students learning about decimal numbers (Forlizzi et al., 2014; McLaren, Adams, et al., 2017). The game has been used as a platform to experiment with 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, McLaren et al., 2016). The materials have been deployed on the web-based learning management system, TutorShop (<https://adapterrex.cs.cmu.edu/> - 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 Figure 1.

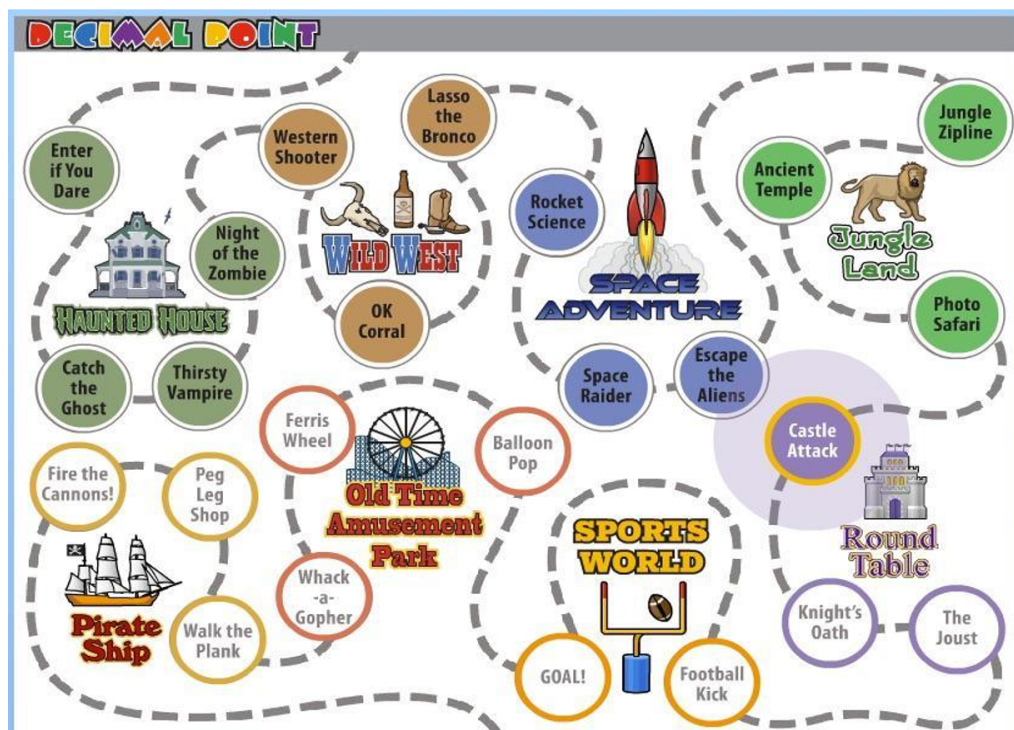


Figure 1: The *Decimal Point* game map

Decimal Point presents students with 5 types of decimal problems, each corresponding to a different mini-game type: (1) ordering decimals; (2) correctly placing a point 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. Figures 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., *Everyday Mathematics* by The University of Chicago School Mathematics Project). For instance, the Castle Attack mini-game of Figure 2 probes the misconception many students have that “longer decimals are larger.” The student playing Castle Attack 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 Figure 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 the map are filled in.

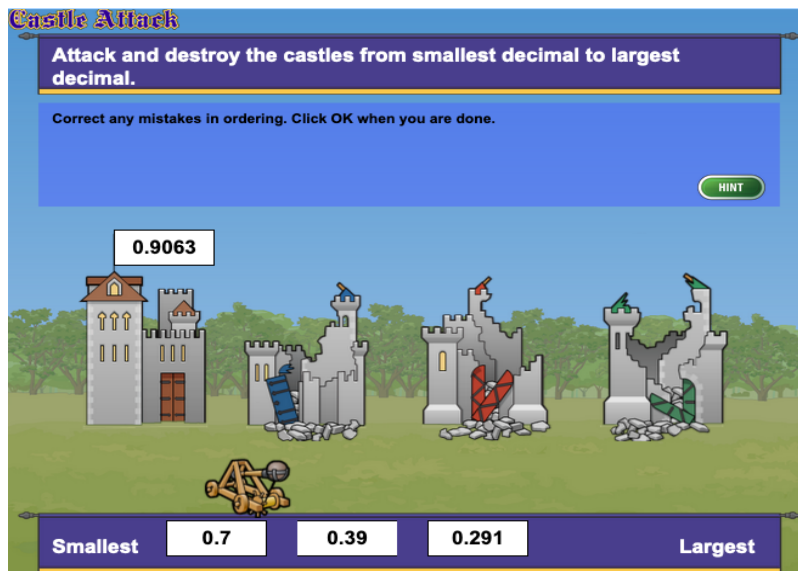
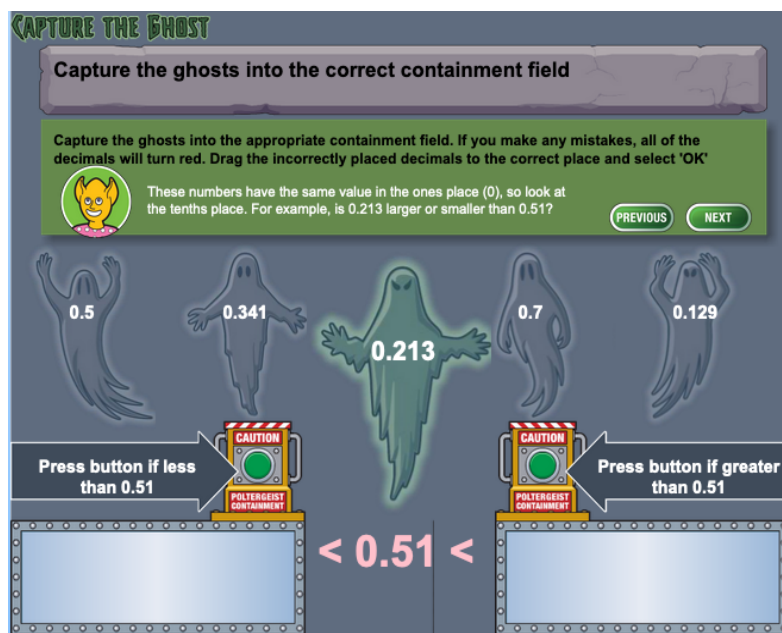


Figure 2: "Castle Attack" one example of a *Decimal Point* mini-game.



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

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 the *Decimal Point* game 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 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.

Figure 2 shows an example of how an on-demand hint is requested, with a “Hint” button on the game display (in the middle, to the right). Once the “Hint” button is selected, the student can go forward and backward through the sequence of hints with “Previous” and “Next” buttons, as shown in Figure 3 (again, see the middle, to the right). Figure 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 Figure 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 suggestions on how to think about and solve the given problem (e.g., For the problem in Figure 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., Figure 3 - “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, Jon Star.

Table 1: Example Hint and Error Messages

Mini-game problem type	Hint example	Error message example
Ordering	Since these numbers all have the same ones place (0), compare the tenths place. Which has the smallest tenths place? <i>(Level 2 hint when the student is given an ordering problem with the decimal numbers 0.14, 0.4, 0.0234, 0.323)</i>	Start by comparing the first digit to the right of the decimal point, even if the digit is 0. <i>(If the student orders the decimal numbers 0.14, 0.0234, 0.323, 0.4,)</i>
Number line	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? <i>(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)</i>	0.456 is greater than 0, so it goes to the right of 0. <i>(If student clicks to the left of 0, to where the decimal number would be negative)</i>

Complete Sequence	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, <u>0.12</u> , ____)
Buckets (less than/greater than comparisons)	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	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 second, 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.)

4. Method

Participants and Design

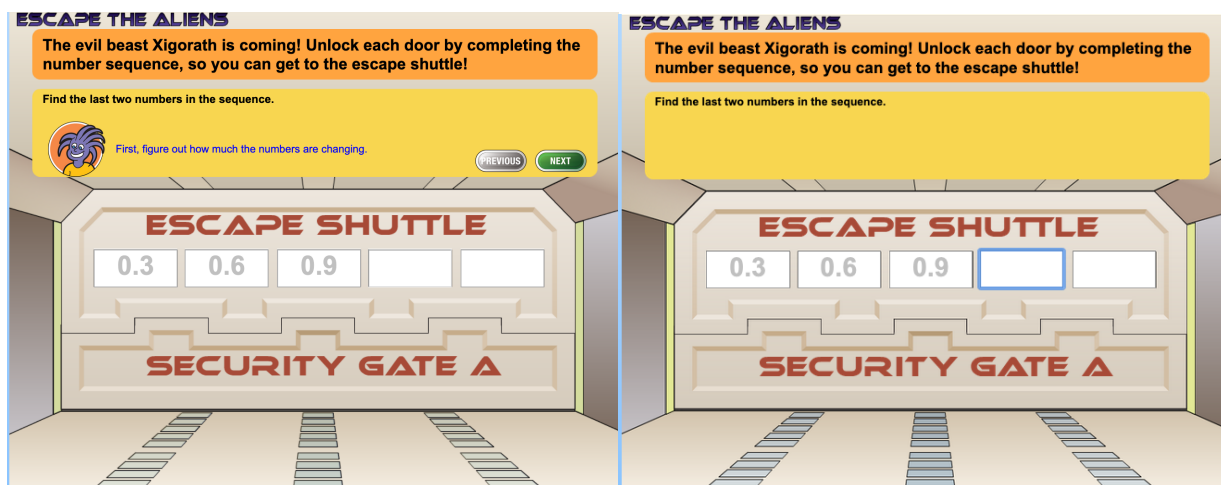
For the pre-COVID, classroom study period, we assigned students to condition, *Hint* or *No-Hint*, by class, due to concerns about students observing differences in the game within a classroom or sharing support with classmates (i.e., students without hints might notice their classmates receiving hints, 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 students participated in the study in classrooms. 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 (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, 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 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 (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.

Table 2: Participating Schools

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

Students took a pretest, completed the intervention according to condition (i.e., played all of the mini-games in the order shown in Figure 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 Figures 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 Figure 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 Figure 6.



Figures 4 & 5: The “Escape the Aliens” mini-game in the *Hint* condition (with “Previous” and “Next” buttons, shown on left) and the *No-Hint* condition, without any hint buttons (right). (See the yellow box, at the bottom, for the hint difference)

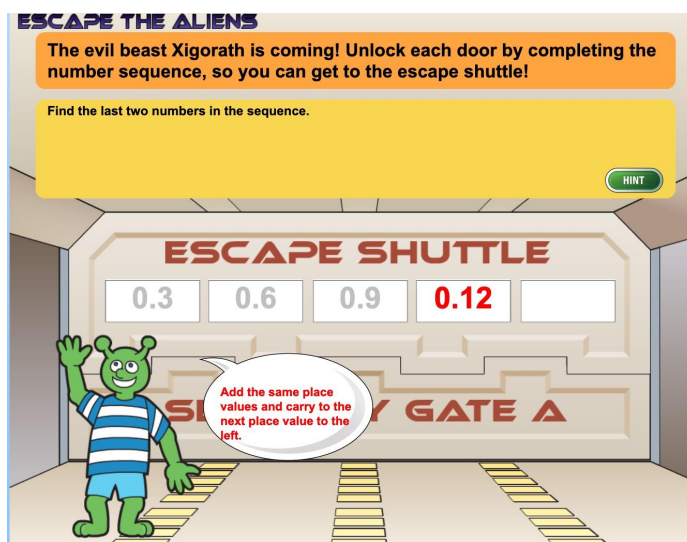


Figure 6: *Hint* condition - The “Escape the Aliens” mini-game with an error message. Here a student makes a very common mistake.

In the classroom version of the study, students worked during 45 to 55 minute class periods for up to a full week on the pretest, game, and immediate posttest. While playing the game, students in the *No-Hint* condition were given the answer by experimenters in the classroom if they asked for help after they (a) made at least 5 mistakes and (b) worked for at least 5 minutes (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 worked at their own pace, logging out at the end of each class and, when logging back in the next day (or whenever they returned to class), were placed in the materials wherever they stopped the previous day (or previous login). A week after the immediate posttest, students took the delayed posttest within one class period.

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. 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.

Materials

A web-based learning environment was used to deploy all of the materials in this study (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 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 or conceptual questions (e.g., “is a decimal number that starts with 0 smaller than 0?”). Three test forms (A, B and C) that were isomorphic and positionally counterbalanced across conditions were used.

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 x 2 (Classroom/Remote x Hint/No-Hint) analysis. We report effect sizes (partial eta squared, η^2_p) for all main effects, interactions, and planned comparisons, and we interpret effects as small when $\eta^2_p < .06$, medium when $.06 < \eta^2_p < .14$, and large when $\eta^2_p > .14$ (see Olejnik & Algina, 2000).

First, a chi-square analysis assessing completion rates, Classroom (88.8%) versus Remote (56.5%), revealed a significant difference, $\chi^2(2, N = 393) = 51.3, p < .001, \phi = .36$, indicating that more students completed the materials in the classroom context. A 2 x 2 ANOVA comparing pretest scores by condition and Classroom/Remote context indicated that students in the *Hint* condition performed significantly worse than students in the *No-Hint* condition, $F(3, 273) = 4.01, p = .046, \eta^2_p = .014$. Students also performed significantly better on the pretest during the remote learning period compared to the classroom learning period, $F(3, 273) = 65.73, p < .001, \eta^2_p = .19$. There was no significant interaction, $F(3, 273) = 0.88, p = .35, \eta^2_p = .003$. Given the differences in pretest scores across conditions, we control for pretest as a covariate or assess learning gains in all analyses reported below.

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?*

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. A 2 x 2 analysis of covariance (ANCOVA) assessing effects of the two learning conditions and Classroom versus Remote context on posttest performance and controlling for pretest indicated that there was no significant effect of *Hint* condition, $F(4, 272) = 1.96, p = .16, \eta^2_p = .007$. Even when controlling for the pretest, students performed significantly better during the remote learning period compared with the classroom learning period, $F(4, 272) = 7.03, p = .008, \eta^2_p = .025$. There was no significant interaction, $F(4, 272) = 0.73, p = .39, \eta^2_p = .003$.

A 2 x 2 ANCOVA assessing effects of condition and Classroom/Remote context on delayed posttest performance and controlling for pretest indicated that there was no significant effect of learning condition, $F(4, 272) = 1.05, p = .31, \eta^2_p = .004$. Even when controlling for pretest, students performed significantly better during the Remote learning period compared with the Classroom learning period, $F(4, 272) = 8.04, p = .005, \eta^2_p = .029$. There was also a significant interaction, $F(4, 272) = 5.49, p = .020, \eta^2_p = .020$.

We conducted a pairwise comparison to interpret the interaction (Figure 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, $F(2, 148) = 4.49, p = .036, \eta^2_p = .029$. Within the Remote group, there was no effect of the Hint condition on delayed posttest performance, $F(2, 123) = 0.73, p = .39, \eta^2_p = .006$.

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)
Classroom	67	<i>Hint</i>	21.40 (10.05)	26.73 (10.24)	26.12 (10.29)
	84	<i>No-Hint</i>	25.07 (10.32)	30.82 (9.54)	31.17 (9.84)
	151	Total	23.44 (10.33)	29.01 (10.02)	28.93 (10.32)
Remote	64	<i>Hint</i>	32.69 (11.25)	37.09 (9.31)	38.00 (8.91)
	62	<i>No-Hint</i>	34.02 (9.52)	38.39 (6.62)	37.94 (8.17)
	126	Total	33.34 (10.42)	37.73 (8.09)	37.97 (8.52)

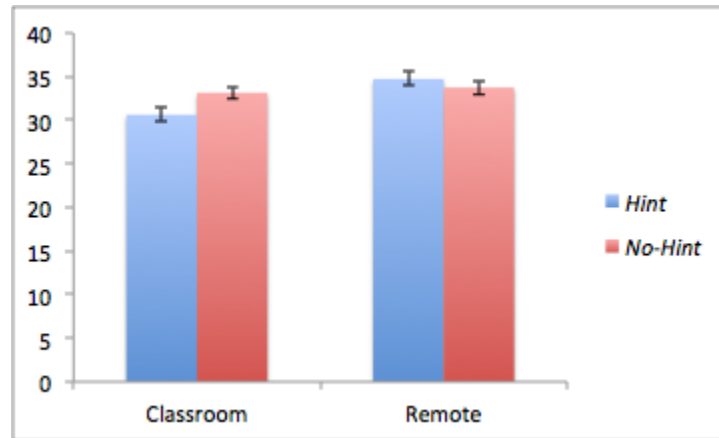


Figure 7: Interaction effect at delayed posttest showing estimated marginal means controlling for pretest: *No-Hint* students performed significantly better than the *Hint* students in the Classroom but not Remotely.

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 Figures 8 and 9 show the number of students that made errors, at opportunity 1, 2, 3, and so forth, for two of the Buckets and Number Line mini-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).

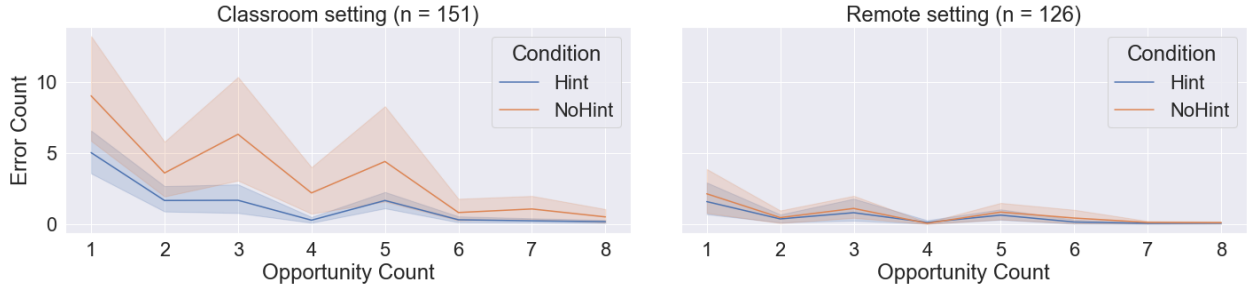


Figure 8: Error count by opportunity in the *Buckets* mini-games.

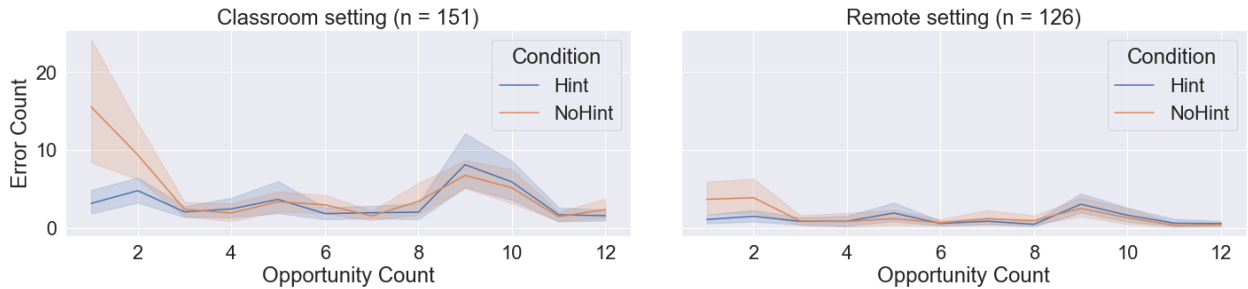


Figure 9: Error count by opportunity in the *Number Line* mini-games.

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 (McLaren, Farzan, et al., 2017; Hou et al., 2020; 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.

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, $F(4, 272) = 1.72, p = .19, \eta^2_p = .006$.

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

	N	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Mean (SD)
Female	136	27.84 (11.35)	34.03 (9.18)	34.17 (9.81)
Male	141	28.05 (11.62)	31.96 (10.96)	31.95 (11.13)

We also examined potential interactions between gender and Classroom/Remote context using a series of 2 x 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 interaction. Within the classroom group, there was a significant effect of gender, with female students performing better on the immediate posttest than male students, $F(2, 148) = 13.96, p < .001, \eta^2_p = .086$ (Figure 10). However, within the remote group, there was no effect of gender on posttest performance, $F(2, 123) = 0.48, p = .49, \eta^2_p = .004$.

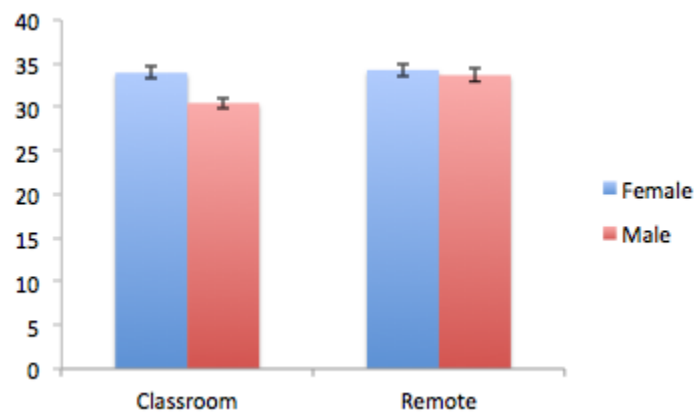


Figure 10: Interaction effect at **immediate posttest** showing estimated marginal means controlling for pretest: Female students performed significantly better than male students in the Classroom but not Remotely.

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 Figure 11).

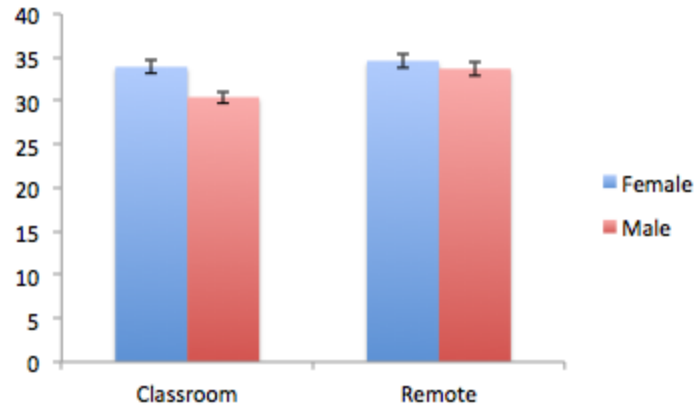


Figure 11: Interaction at **delayed posttest** showing estimated marginal means controlling for pretest: Female students performed marginally better than male students in the Classroom but not Remotely.

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 *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 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. Figures 12 and 13 show the Classroom versus Remote context plotting of “hint before 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.

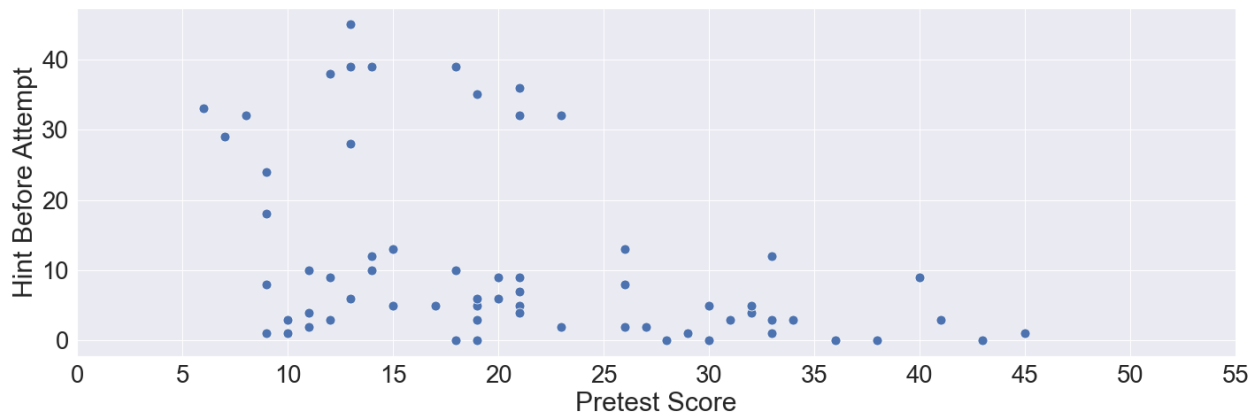


Figure 12: Plotting of prior knowledge (performance on pretest) versus hint usage before attempt for the **Classroom** students.

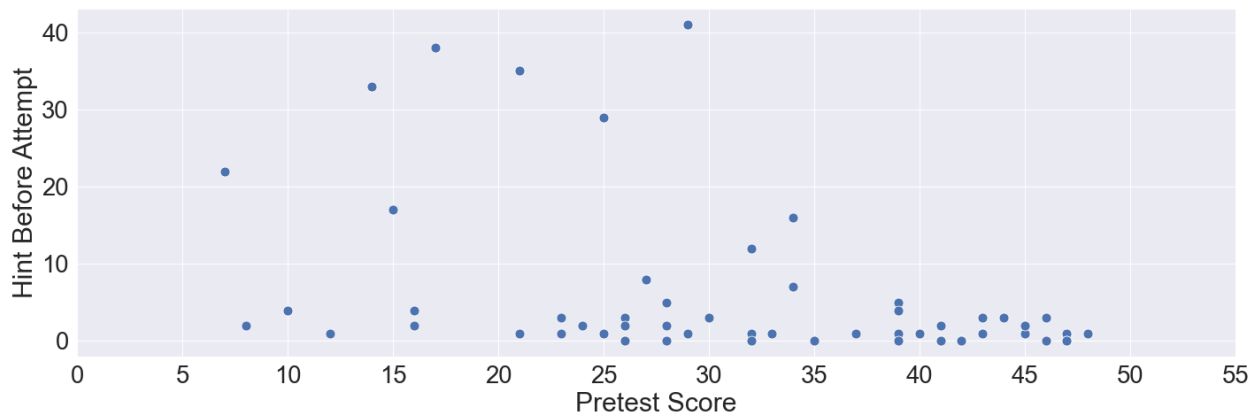


Figure 13: Plotting of prior knowledge (performance on pretest) versus hint usage before attempt for the **Remote** students.

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 *Decimal Point* hints is 18.64 words). Such an analysis uncovers the number of times a student doesn’t read hints deliberately. Figures 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 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 too quickly in one setting compared to the other (Classroom vs. Remote), when controlling for prior knowledge.

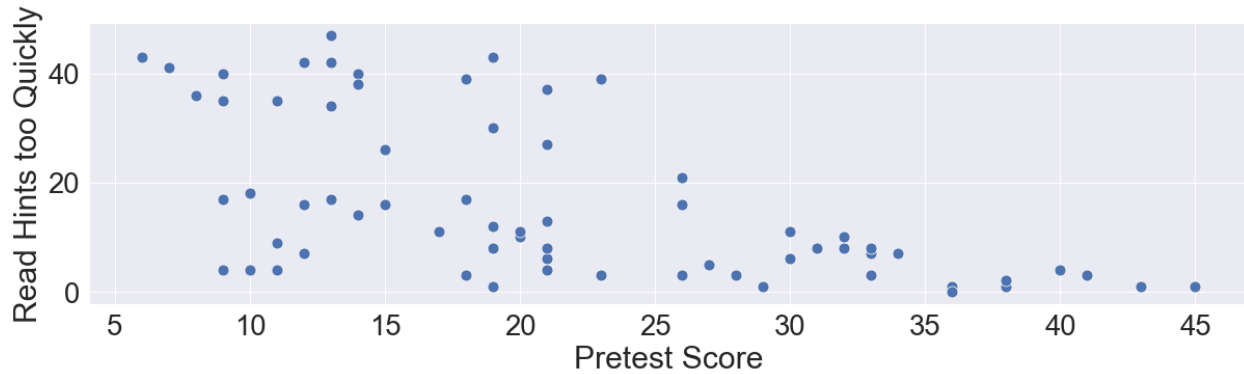


Figure 14: Plotting of prior knowledge (performance on pretest) versus Read Hints too Quickly by the **Classroom** students.

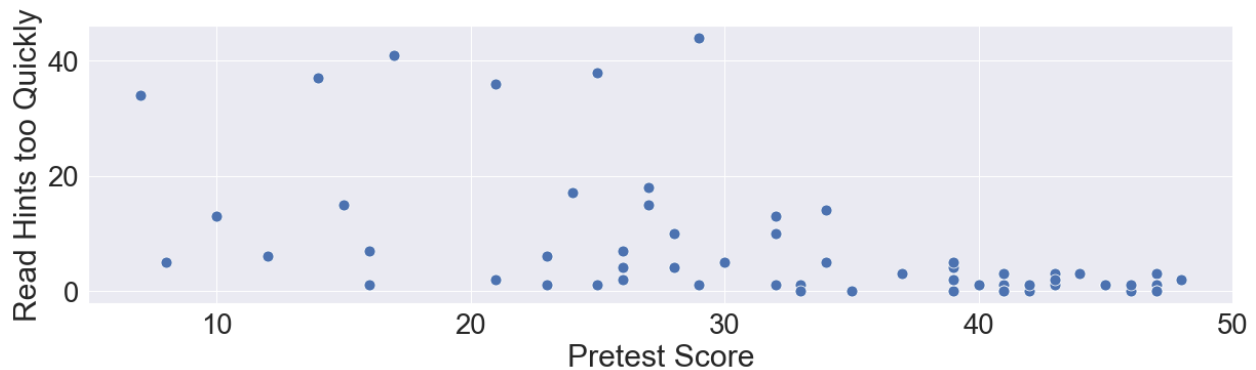


Figure 15: Plotting of prior knowledge (performance on pretest) versus Read Hints too Quickly by the **Remote** students.

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.

Table 5: Correlations between hint metrics and learning gains

	Pretest-to-Posttest Gain	Pretest-to-Delayed Posttest Gain
<i>Classroom context</i>		
Metric 1: Hint Before Attempt	-0.37*	-0.30*
Metric 2: Read Hints Too Quickly	-0.35*	-0.27*
<i>Remote context</i>		
Metric 1: Hint Before Attempt	-0.04	0.11
Metric 2: Read Hints Too Quickly	-0.00	0.12

*indicates $p < .05$.

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 game, *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, this 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 (Figures 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 is harder and involves making more errors than requesting support and answers 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. Our result is in line with the results of O'Rourke et al. (2014), in which students in a no-hint condition learned more than 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).

In closing, as illustrated by the learning curves of Figures 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.

7. Conclusions

Although we were exploring an important, understudied topic -- do hints and error feedback help in the context of a digital learning game? -- we fortuitously stumbled upon a perhaps more important topic during the course of this study: How does the instructional context -- classroom versus remote learning -- impact learning with a digital learning game? We were able to uncover some interesting differences between the two environments that both illuminate differences, as well as point to potential future work.

References

- Aleven, V., McLaren, B. M., Sewall, J., Van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., & Koedinger, K. R. (2016). Example-tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education*, 26(1), 224-269.
- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). 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. doi: 10.1007/s40593-015-0089-1.
- 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. & Koedinger, K. (2000). Limitations of student control: Do students know when they need help? In the *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000*, edited by G. Gauthier, C. Frasson, and K. VanLehn, (pp. 292-303). Berlin: Springer Verlag, 2000. *Best Paper Award ITS 2000*.
- Anderson, J.R., Reder, L.M. & Simon, H.A. (1996). Situated Learning and Education. *Educational Researcher* 25 (4), 5-11.
- Anderson, J.R. & Bellezza, F.S. (1993). *Rules of the Mind*. Psychology Press, 1993.
- Applefield, J.M., Huber, R.L. & Moallem, M. (2001). Constructivism in Theory and Practice: Toward a Better Understanding. *The High School Journal*. DOI:[10.2307/40364404](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 (2009) 117–125.
- 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.
- 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. *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 (2015) 94–103. doi: <https://doi.org/10.1016/j.chb.2015.01.038>.
- 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. DOI: 10.3102/0034654315582065.
- 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. doi: 10.1007/s40593-013-0002-8.
- Costello, B. & Edmonds, E. (2007) A study in play, pleasure and interaction design. *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces*, 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* **2018** 8, 111. doi: 10.3390/educsci803011.
- 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., Alevan, V., Scheines, R. & Carver, S.M. (2017). Using tutors to improve educational games: A cognitive game for policy argument, *Journal of the Learning Sciences*, 26:2, 226-276, DOI: 10.1080/10508406.2016.126928.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview Press.
- Forlizzi, J., McLaren, B.M., Ganoë, C.H., McLaren, P.B., Kihumba, G., & Lister, K. (2014). Decimal Point: Designing and developing an educational game to teach decimals to middle school students. In C. Busch

(Ed.) *Proceedings of the 8th European Conference on Games Based Learning (ECGBL-2014)*. (pp. 128-135) Academic Conferences and Publishing International Limited, Reading, U.K.

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

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. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK'19)*. (pp. 121-129). ACM. DOI: <https://doi.org/10.1145/3303772.3303797>.

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

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). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_21.

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

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.

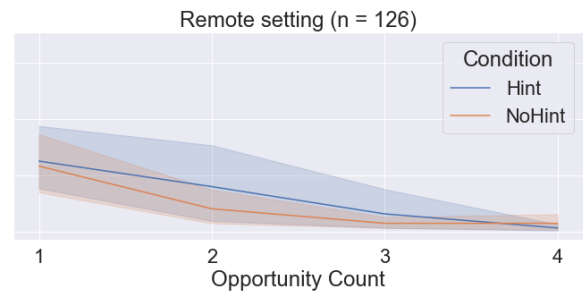
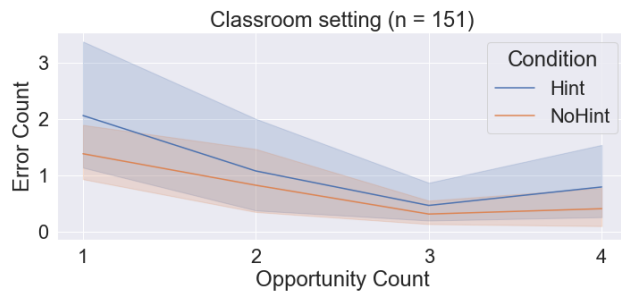
Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79–97.

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) doi: <https://dl.acm.org/doi/10.5555/3417639.3417652>.

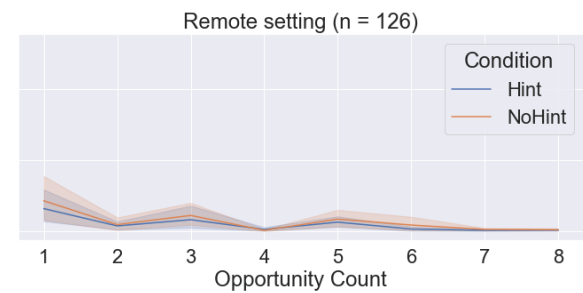
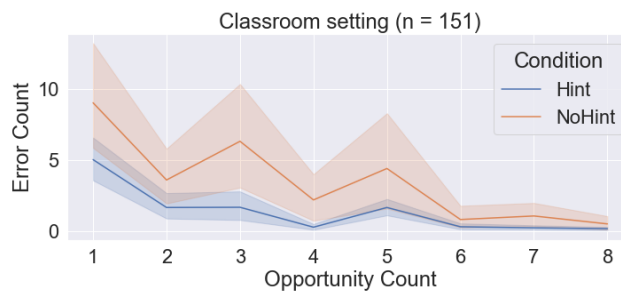
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918.
<http://dx.doi.org/10.1037/a0037123>.
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, 3(1987), 223-253.
- Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*. 70: 531–49,
- Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. Cambridge, MA: MIT Press.
- McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017). A computer-based game that promotes mathematics learning more than a conventional approach. *International Journal of Game-Based Learning (IJGBL)*, 7(1), 36-56. doi:10.4018/IJGBL.2017010103.
- McLaren, B.M., Farzan, R., Adams, D.M., Mayer, R.E., & Forlizzi, J. (2017). Uncovering gender and problem difficulty effects in learning with an educational game. In E. André, R. Baker, X. Hu, M.M.T. Rodrigo, and B. du Boulay (Eds.). *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017)*. LNAI 10331 (pp. 540-543). Springer: Berlin.
- 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). doi: 10.1109/ICALT.2012.60.
- 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.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 *Computers In Human Behavior*.
- Nguyen, H., Harpstead, E., Wang, Y. & McLaren, B.M. (2018). Student agency and game-based learning: A study comparing low and high agency. In C. Rosé, R. Martínez-Maldonado, H.U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren and B. du Boulay (Eds.). *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018)*. LNAI 10947 (pp. 338-351). Springer: Berlin.
- Olejník, 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)*. (pp. 51-60). dos: <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., Buttner, 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.
- Squire, K., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight (American Society of Ophthalmic Registered Nurses)*, 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 (1997). <https://doi.org/10.1007/BF02299526>.
- 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. doi: <http://dx.doi.org/10.1080/00461520.2011.611369>.
- 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: Isotani S., Millán E., Ogan A., Hastings P., McLaren B., Luckin R. (eds). *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED 2019)*. LNAI 11625 (pp. 518-531). Springer, Cham.
- Winne, P. H. (2015). Self-Regulated Learning. In the *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition).
- 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>.

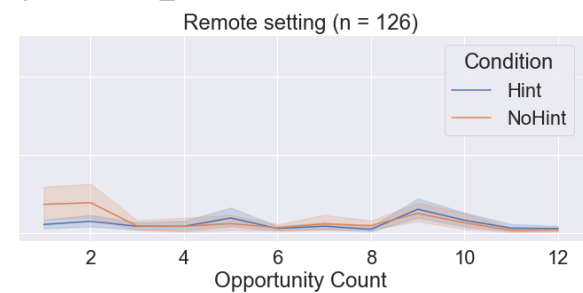
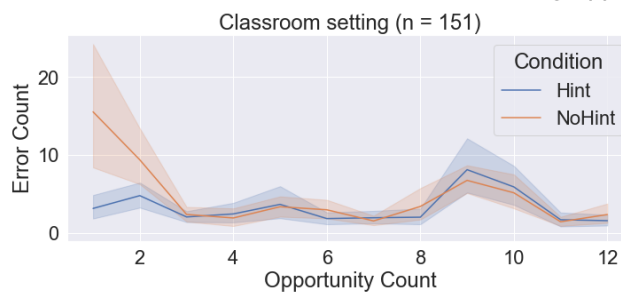
Appendix A



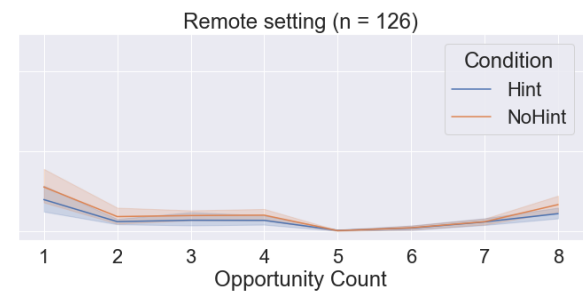
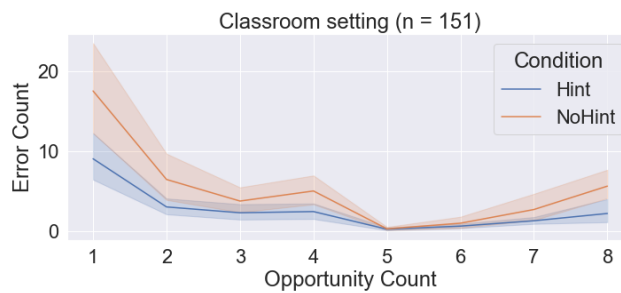
Error count by opportunity in Addition



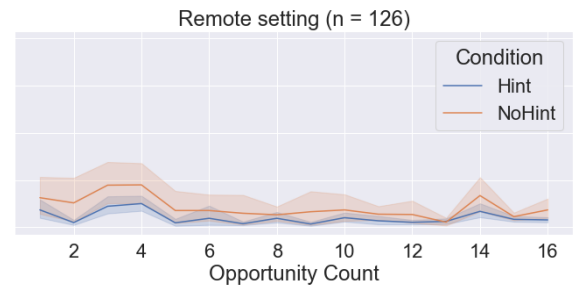
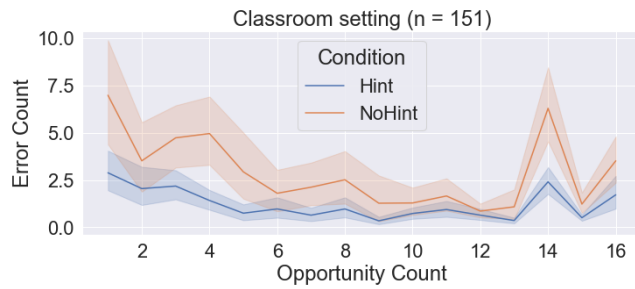
Error count by opportunity in Bucket



Error count by opportunity in Number Line



Error count by opportunity in Sequence



Error count by opportunity in Sorting

Acknowledgements

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 Ganoë, 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 development and early experimentation with the Decimal Point game. Thanks to Jon Star who provided input on the hint and error messages used in this study. Special thanks to Rosta Farzan who first had the idea to do gender analyses of Decimal Point use."

Highlights

- Using technology in class versus at home makes a difference to behavior and learning
- Students complete online materials at a much higher rate in class than at home
- In class, no hints in a game led to significantly better learning than hints in a game
- Students used significantly more hints in class than at home
- Female students performed significantly better than male students in class