Gaming and confrustion explain learning advantages for a math digital learning game

J. Elizabeth Richey[1[0000-0002-0045-6855]], Jiayi Zhang[2], Rohini Das[2], Juan Miguel Andres-Bray[2], Richard Scruggs[2[0000-0002-0366-5644]], Michael Mogessie[1[0000-0001-6769-5941]], Ryan S. Baker[2[0000-0002-3051-3232]], Bruce M. McLaren[1[0000-0002-1196-5284]]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
[2] University of Pennsylvania, Philadelphia PA 19104 USA
jelizabethrichey@cmu.edu

Abstract. Digital learning games are thought to support learning by increasing enjoyment and promoting deeper engagement with the content, but few studies have empirically tested hypothesized pathways between digital learning games and learning outcomes. *Decimal Point*, a digital learning game that teaches decimal operations and concepts to middle school students, has been shown in previous studies to support better learning outcomes than a non-game, computer-based instructional system covering the same content. To investigate the underlying causes for *Decimal Point*'s learning benefits, we developed log-based detectors using labels from text replay coding of the data from an earlier study. We focused on gaming the system, a form of behavioral disengagement that is frequently associated with worse learning outcomes, and confrustion, an affective state that combines confusion and frustration that has shown mixed results related to learning outcomes. Results indicated that students in the non-game condition engaged in gaming the system at nearly twice the level of students in the game condition, and gaming the system fully mediated the relation between learning condition and learning outcomes. Students in the game condition demonstrated higher levels of confrustion during the self-explanation phase of the game, and while confrustion was not related to learning outcomes in the game condition, it was associated with better learning outcomes in the non-game condition. These results provide evidence that digital learning games may support learning by reducing behavioral disengagement, and that the effects of confusion and frustration may vary depending on digital learning context.

Keywords: Digital learning games, Affect detector, Ed. Data Mining.

1 Introduction

1.1 Digital learning games and learning outcomes

Most American children play digital games. The Common Sense Census [18] found that 66% of U.S. tweens and 56% of teens report playing digital games on any given day, with an average time of two or more hours per day among those who play. Recognizing this enthusiasm for games, more than half of U.S. teachers ask their students to use digital learning games in class at least once a week [22, 25]. Although data are still emerging on how digital learning game use has changed during the COVID-19 pandemic, Internet search intensity for online learning resources doubled in the early months of the pandemic [4] and interactive learning environment usage has increased [10]. The increased reliance on digital learning tools is not likely to abate even when face-to-face instruction can consistently be resumed, and the importance of digital learning games in educational settings seems likely to continue to grow in the future.

A number of studies have found improved learning outcomes for digital learning games compared to non-game learning conditions [16, 58]. Several meta-analyses have also revealed motivational benefits of digital learning games, including benefits to self-efficacy and attitudinal outcomes compared to more traditional instruction [54, 59]. Prior research has shown learning and engagement benefits from digital learning games in a variety of academic domains, including mathematics [27, 47, 53], science [13, 14], and language learning [57, 62]. However, designing games that teach academic topics is still a challenging task that is not always successful, and the educational effectiveness of digital learning games varies depending on a number of circumstances [19, 33, 37, 60]. For instance, educational benefits are more likely to occur when games are specifically designed based on cognitive theories of learning [44].

In particular, there has been limited empirical evidence about what is effective for mathematics games, with a recent review finding only six methodologically sound experiments that compared learning mathematical material in a game versus more conventional media [37]. Of those six experiments, four produced positive results favoring game playing. In this paper, we focus on one of those games, *Decimal Point* [23, 38], which was designed in consultation with a mathematics education expert and based on theory and evidence about common student misconceptions regarding decimal mathematics [26, 31, 56]. Like many digital learning games, *Decimal Point* was designed to support students' learning after initial instruction on the relevant topics by providing engaging opportunities for additional practice. In a study involving more than 150 5th and 6th grade students, *Decimal Point* led to significantly more learning and was rated by students as significantly more engaging than a more conventional but still effective computer-based tutoring approach [38].

Experimental comparisons between digital learning games and conventional learning technology can establish digital learning games as effective (or not) at producing desired learning outcomes, but these methods do not get at the underlying *reasons* for the effects. Very few studies have tested specific cognitive or affective processes as potential mediators of learning from games compared to non-games. There is a general lack of understanding about how digital learning games support learning, and digital game designers often must work without empirical guidance for how to make learning games more effective. In some cases, this results in uninformed adoption of extrinsic rewards such as points, badges, and competition, which often do not foster productive learning processes [40, 41, 51]. Understanding how digital learning games support learning is essential for informing better digital learning game design. Additionally, teachers have limited class time available, and greater evidence of when and how students learn from digital learning games—and especially how they might learn *differently* from games compared to non-games—will help inform teachers' choices about which digital learning games to incorporate into their teaching and how to enhance their students' learning.

This suggests a need to take a more detailed look at the underlying cognitive and affective mechanisms that lead to learning with games. The field of learning analytics provides tools to help in identifying the cognitive and affective processes that educational technology supports [6, 24, 28, 52, 55]. In the current study, we use behavioral data and learning analytics to examine the cognitive and affective pathways through which digital learning games operate to support learning outcomes. Specifically, we reanalyze an existing dataset [38] to assess two potential paths—gaming the system and confrustion—that might explain differences in learning processes and outcomes.

1.2 Gaming the system and confrustion

The last few decades have seen a surge in scholarship around student behaviors and emotions or affect while learning [6, 11, 42, 61]. Gaming the system—attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material—has been a behavior of particular interest within computer-based game and tutoring contexts due to its negative relation with learning outcomes [7, 17, 39]. Gaming the system has both an immediate and long-term impact on learning and academic performance. One study investigating the effects of gaming using log data from a middle-school Cognitive Tutor mathematics curriculum found that gaming the system was associated with immediate poorer learning *and* an aggregate negative impact on learning [7]. In addition, students who game the system in middle school mathematics are less likely to enroll in higher education [49] or to take a STEM job after college [3].

Several studies have also found evidence that differences in learner emotions or affect are associated with learning outcomes in both the short term [46] and long term [49]. Two affective states that have been of interest in affective computing research are confusion and frustration, which have both been found to be associated with student learning. Some studies have found strong positive correlations between confusion or frustration and learning [20, 35], while others have found strong negative correlations to learning [48, 50]. Whether confusion and frustration support or hinder learning may be related to whether the student has support or metacognitive skills to resolve their confusion and frustration [21, 36]. Learning context may also affect the relation between confusion or frustration and learning outcomes. Previous research that identified positive relations between confusion or frustration and learning was conducted in non-game digital learning environments [20, 35]. Fewer studies have examined the relation between affect and learning in the context of digital learning games, where confusion and frustration may be more disruptive to game play, but at least one recent study using *Decimal Point* found a negative relation [39]. Confusion and frustration are often difficult to distinguish when judging only based on students' interactions with educational technology. Due to this and to their similar relation with learning, a number of previous studies have investigated a combination of the two states instead, called "confrustion" [36, 39, 45].

2 Method

We obtained interaction and outcome data collected through *Decimal Point* in an experiment first reported in [38]. We developed log-based detectors using labels from text replay coding of the data [8, 34]. We briefly describe the methods of the previous study; for a more detailed description of both the game and study, see [23] and [38].

2.1 Participants

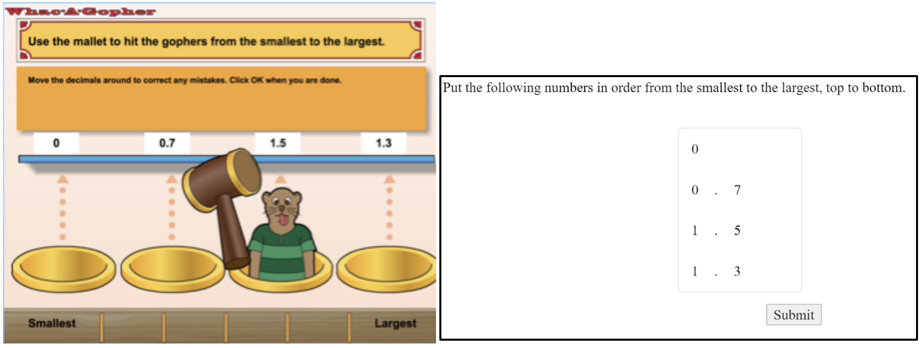
Students participated in the study as part of their normal math instruction at two middle schools in a northeastern major metropolitan area. A total of 213 students participated in the study, but 39 students (19 in the game condition and 20 in the non-game condition) were dropped from analyses for failing to complete the pretest, posttest, or delayed posttest. Of the remaining 174 students (97 female students, 76 male students, and 1 missing gender information), 81 students were assigned to play *Decimal Point*, while 93 students completed a non-game, computer-based instructional system covering the same content.

2.2 Materials and Procedure

Decimal Point is a single-player game with an amusement park metaphor targeting 5^{th} and 6^{th} grade students learning about decimal numbers. *Decimal Point* runs on the web, within a standard browser, and was developed using HTML/JavaScript and the Cognitive Tutor Authoring Tools, or CTAT [2]. The materials are deployed on the web-based learning management system TutorShop [1], which logs all student actions. *Decimal Point* is composed of a series of 24 "mini-games" within a larger amusement park map. Forty-eight decimal problems (two problems for each of the 24 mini-games) were implemented for the game.

Decimal Point presents students with five types of mini-game problems: (1) ordering decimals; (2) number line placement; (3) decimal sequences; (4) sorting decimals into less-than and greater-than "buckets"; and (5) adding decimals (Fig. 1). After solving each problem, students are prompted to self-explain their answer by selecting from a multiple-choice list of possible explanations. For example, after an ordering problem, the student might see the following: "To order these decimals from smallest to largest, start by finding: a) the longest decimal; b) the decimal with the smallest tenths place value; c) the shortest decimal; or d) the decimal with the smallest hundredths place value." This employs a well-established learning science principle that can promote deeper learning [15, 32]. To develop the game problem types, the developers surveyed problems students currently encounter in popular math curricula and designed mini-games and tests to probe for decimal misconceptions [56].

Decimal Point has six characters that serve as guides and cheerleaders for the player throughout the game. These game elements provide fantasy [9], as well as giving the player a narrative context for why they are performing various problem-solving activities. 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.



Figs. 1 & 2. *Decimal Point* "Whac-A-Gopher" (left), an example of an ordering mini-game, and the non-game equivalent (right).

The non-game control condition (Fig. 2) presented the same mathematical content, including both problem-solving and self-explanation elements, without the game features or narrative. Problems were presented on a plain background in a manner consistent with many intelligent tutoring systems. As with *Decimal Point*, students had to complete all problems in a predetermined sequence. In both the game and non-game versions, students were told immediately if their answers on the problem-solving and self-explanation questions were incorrect, and they could not advance to the next problem until they correctly answered the current problem.

Students completed three isomorphic versions of a test on decimal number operations and concepts. Tests were administered before students completed the materials (pretest), immediately after completion (posttest), and a week after completion (delayed posttest). Versions of the test were counterbalanced across time points to control for any unintended variations in the tests. Each test contained 24 problems, including some problems with similar decimal number content to what was presented in the game and non-game systems and other problems that targeted underlying concepts related to decimal number operations but not explicitly taught within the game and non-game. Students could earn multiple points on some problems, with a total of 61 points possible for correctly answering all questions on the test.

2.3 Detector construction

Text replay coding has been used to identify learner behaviors and affect [8, 34]. In this method, coders base their affect coding on log data gathered on the students' interaction with the learning environment. Text replay coding involves breaking down the existing data set into text replays, or clips, each either spanning a specific amount of time, a specific number of transactions, or delineated by start or end events.

Whereas our previous detectors were built using problem-level labels, the current study broke each problem or game level down into their two steps during the labeling process: problem solving and self-explanation. As such, text replay coding had to be conducted in four iterations: once each for gaming and confrustion in the problem-solving step; and once each again for gaming and confrustion in the self-explanation

Table 1. Detector performance for gaming and confrustion detectors in the problem-solving and self-explanation steps.

	Problem Solving	Self-Explanation
Gaming	*AUC*=0.889, *k*=0.504	*AUC*=0.999, *k*=0.952
Confrustion	*AUC*=0.915, *k*=0.565	*AUC*=0.956, *k*=0.645

step. In each iteration, text replay coding was conducted in three phases. In phase 1, two human coders coded a set of clips together in order to establish a labeling rubric. In phase 2, both coders coded another set of clips separately, in order to assess inter-rater reliability. If the coders attained acceptable reliability, the coders moved on to phase 3. If not, the coders discussed the differences in their labeling, and then did another round of phase 2 coding, repeating this process until they attained acceptable reliability. Two rounds of phase 2 coding were conducted for confrustion in the problem-solving clips, and one round of phase 2 coding was conducted for the other three detectors. For the problem-solving clips, the inter-rater reliability (IRR) kappa was 0.74 for both confrustion and gaming. Kappa was 0.62 and 0.88 for confrustion and gaming, respectively, in the self-explanation clips. Once in phase 3, the two coders divided the remaining clips and coded them separately. Since less confrustion was observed in the self-explanation clips, almost twice as many self-explanation clips as problem-solving clips needed to be coded to have enough data to build the model. In total, 800 problem-solving clips and 1,500 self-explanation clips were coded and used to construct the automated affect detectors. Furthermore, clips were stratified to equally represent schools, problem type, and experiment condition.

The labeled data were input into machine learning algorithms to emulate the coders' judgments, based on prior studies that showed it was feasible to detect gaming [43] and confrustion [34] using this approach. The gaming and confrustion detectors were all built using the Extreme Gradient Boosting (XGBoost) classifier [12]. The classifier uses an ensemble technique that trains an initial, weak decision tree and calculates its prediction errors. It then iteratively trains subsequent decision trees to predict the error of the previous decision tree, with the final prediction representing the sum of the predictions of all the trees in the set. Four automated detectors were built in total, i.e., gaming in the problem-solving step, confrustion in the problem-solving step, gaming in the self-explanation step, and confrustion in the self-explanation step. Based on 10-fold student-level cross-validation, we determined that the models could reliably predict the two constructs in both the problem-solving and self-explanation steps. Detector performance can be found in Table 1. The detectors were then applied to predict gaming and confrustion in the rest of the data set.

3 Results

Results were previously reported regarding the effect of the game compared to the non-game on posttest and delayed posttest performance [38]. Specifically, analyses of covariance (ANCOVAs) revealed that students in the game condition outperformed

Table 2. Average probabilities of gaming the system and confrustion by condition for problem-solving (PS) and self-explanation (SE) activities.

	Gaming (PS) *M (SD)*	Gaming (SE) *M (SD)*	Confrustion (PS) *M (SD)*	Confrustion (SE) *M (SD)*
Game	.14 (.099)	.22 (.11)	.18 (.086)	.041 (.035)
Non-game	.27 (.12)	.30 (.14)	.15 (.056)	.066 (.055)

students in the non-game condition on posttest, $F(1,172) = 11.50$, $p = .001$, $\eta_p^2 = .063$, and delayed posttest performance, $F(1, 172) = 11.86$, $p = .001$, $\eta_p^2 = .065$.

To understand the effect of the game on students' cognitive and affective processes, we compared predicted rates of gaming the system and confrustion among students playing the game against those completing the non-game version. We examined rates during problem solving and rates while completing the self-explanation questions separately (Table 2). Students using the non-game demonstrated almost double the levels of gaming the system while problem solving as students playing the game, and this difference was significant, $F(1, 173) = 57.64$, $p < .001$, $\eta_p^2 = .25$. On self-explanation questions, students in the non-game also showed significantly higher levels of gaming the system, $F(1, 173) = 17.87$, $p < .001$, $\eta_p^2 = .09$, and confrustion, $F(1, 173) = 12.40$, $p = .001$, $\eta_p^2 = .07$. In contrast, students using the non-game condition show significantly *lower* levels of confrustion during the problem-solving portion, $F(1, 173) = 5.77$, $p = .017$, $\eta_p^2 = .03$.

To understand how these cognitive and affective processes related to posttest performance, we assessed a regression model predicting posttest scores with pretest scores, gaming probabilities for problem solving and self-explanation, and confrustion probabilities for problem solving and self-explanation (Table 3). The resulting model predicted 68.9 percent of the variance. Within the model, pretest scores, gaming the system for problem-solving questions, and gaming the system for self-explanation questions were all significant predictors of posttest scores. We assessed the same model predicting delayed posttest scores. The resulting model predicted 66.1 percent of the variance and, within the model, pretest scores and gaming the system on problem-solving were again significant predictors of delayed posttest scores; additionally, confrustion on self-explanation emerged as a significant predictor.

Table 3. Regression models predicting posttest and delayed posttest scores with pretest scores, gaming probabilities, and confrustion probabilities.

	Posttest	Delayed posttest
Overall model	$R^2 = .70$, $F(5,168) = 77.60$, $p < .001$	$R^2 = .67$, $F(5,168) = 68.54$, $p < .001$
Pretest	$\beta = .48$, $p < .001$	$\beta = .45$, $p < .001$
Gaming (PS)	$\beta = -.30$, $p < .001$	$\beta = -.42$, $p < .001$
Gaming (SE)	$\beta = -.16$, $p = .005$	$\beta = -.077$, $p = .19$
Confrustion (PS)	$\beta = -.017$, $p = .77$	$\beta = .062$, $p = .20$
Confrustion (SE)	$\beta = .042$, $p = .36$	$\beta = .12$, $p = .012$

Finally, we wanted to understand whether differences in cognitive or affective processes explained the effect of the game on learning outcomes. Given that gaming the system on problem-solving questions predicted learning outcomes at posttest and delayed posttest and that levels of gaming differed across conditions, we examined gaming the system on problem-solving questions as a mediator between condition and each test (posttest and delayed posttest; Fig. 3). We used the PROCESS macro for SPSS statistical software [30], which applies 5000 bootstrap estimates to create confidence intervals, to test the indirect effect of condition (game = 0, non-game = 1) on posttest and delayed posttest with gaming the system on problem-solving questions as the mediator. Pretest scores were included as a covariate. Results indicated that students in the non-game condition had significantly greater probabilities of gaming the system, $a = .70$, $p < .001$. Gaming the system was negatively associated with performance on the posttest regardless of condition, $b = -.37$, $p < .001$, and there was no direct effect of condition on posttest performance when controlling for gaming the system, $c = -.07$, $p = .48$. Consistent with our mediation prediction, the indirect effect of condition on posttest through gaming the system was significantly different than zero, $ab = -.26$, 95% CI [-.12, -.064]. Similar results were found for the delayed posttest: gaming the system was negatively associated with performance on the delayed posttest, $b = -.42$, $p < .001$, and there was no direct effect of condition on delayed posttest performance when controlling for gaming the system, $c = -.062$, $p = .56$. Again, the indirect effect of condition on delayed posttest through gaming the system was significantly different than zero, $ab = -.29$, 95% CI [-.44, -.18].

Given the mixed results regarding confrustion in prior literature and in our findings, we examined whether the relation between confrustion and learning might differ between the game and non-game contexts. To do this, we tested game condition as a moderator of the relation between confrustion and each test (posttest and delayed posttest) while controlling for pretest. Moderation analyses in PROCESS showed no significant interaction between confrustion on problem-solving questions and condition when predicting posttest, $b = -15.70$, $p = .26$, 95% CI [-43.24, 11.84], or delayed posttest, $b = -22.29$, $p = .13$, 95% CI [-51.44, 6.85]. However, there was a significant interaction between confrustion on self-explanation questions and condition when predicting posttest, $b = 69.66$, $p = .003$, 95% CI [23.73, 115.58], and inclusion of the

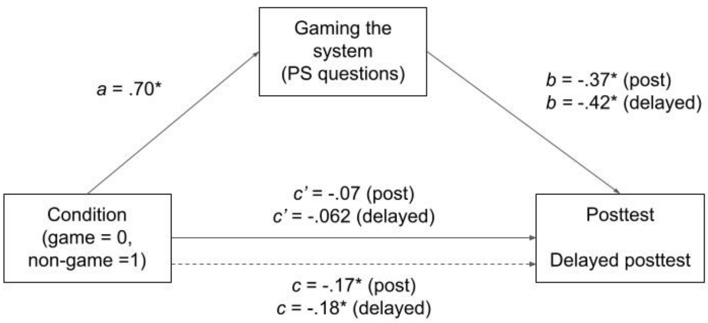
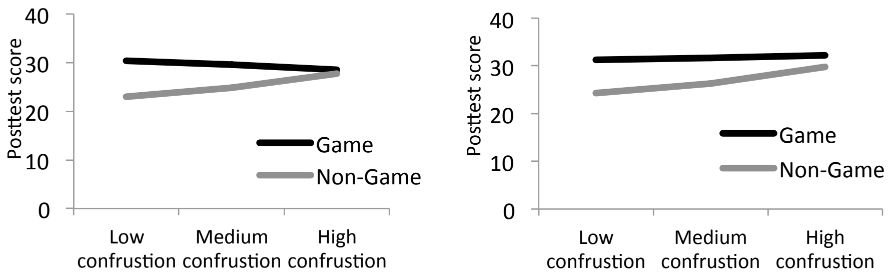


Fig. 3. The mediation model showing path standardized coefficients for a mediation analysis of learning condition on posttest through gaming the system on problem-solving questions.

interaction term explained significantly more variance in the model, $\Delta R^2 = .018$, $F(1, 169) = 8.97$, $p = .003$. While confrustion was not related to posttest performance in the game condition ($b = -18.90$, $p = .34$), it was positively related to posttest performance in the non-game condition ($b = 50.76$, $p < .001$; Fig. 4). There was a similar interaction predicting delayed posttest, $b = 47.63$, $p = .049$, 95% CI [.32, 94.94], and inclusion of the interaction term again explained significantly more variance in the model, $\Delta R^2 = .009$, $F(1, 169) = 3.95$, $p = .049$. As with the posttest, confrustion was not related to delayed posttest performance in the game condition ($b = 10.16$, $p = .62$), but it was positively related to delayed posttest performance in the non-game condition ($b = 57.79$, $p < .001$; Fig. 5).



Figs. 4 & 5. Interaction of confrustion (SE) items and condition predicting posttest (left) and delayed posttest score (right). Scores were calculated using the regression equation for low (16th percentile), medium (50th percentile), and high (84th percentile) values of confrustion.

4 Discussion and Conclusion

Although digital learning games continue to grow in use, relatively few studies have empirically assessed differences in cognitive and affective processes between games and non-game, computer-based systems covering the same content. This paper presents a promising approach using educational data mining to build log-based detectors that can capture such differences. Results showed that the positive effect of learning with the game was fully mediated by students' lower levels of gaming the system when playing the game. Gaming the system has been consistently associated with negative short-term and long-term outcomes, ranging from lower achievement in the task where gaming is measured to reduced likelihood of enrolling in college or choosing a STEM-related job [3, 7, 49]. While it is not surprising that gaming the system was associated with worse performance in *Decimal Point,* it is an important and novel finding that the game reduced students' tendencies to game the system compared to the non-game version and that this reduction in gaming explained differences in learning outcomes. Gaming the system is considered a form of behavioral disengagement, and digital learning games are thought to increase students' engagement through game features such as fantasy and narrative context. Results appear to support the idea that introducing engaging features can reduce students' disengaged behaviors and thereby enhance learning, though causality cannot be inferred from these data.

Confrustion did not consistently predict learning outcomes, but these results are similar to prior research finding conflicting relations between confusion or frustration and learning. We found that confrustion on self-explanation questions played a different role in learning depending on whether students were working in the game or non-game context. In the game, confrustion did not predict learning outcomes, while in the non-game, greater levels of confrustion on self-explanation questions were associated with better learning outcomes. When students experience confusion or frustration while learning, it can trigger productive cognitive and metacognitive processes such as trying a different strategy and monitoring progress [21]. Students experiencing confrustion in the non-game may have engaged in these productive strategies to resolve their confrustion and ultimately gain more from the self-explanation process. On the other hand, confrustion may be less beneficial in a game setting because it feels disruptive to the engaging, playful interactions students expect from a game.

This work suggests several fruitful avenues for further advancing researchers' and developers' understanding of how digital learning games support learning. While our results suggest that differences in gaming the system could explain many of the benefits of games, there are a variety of other cognitive and affective processes that might also play a role. Developing additional detectors for constructs such as boredom, delight, engaged concentration, and carelessness could identify additional pathways that mediate the effect of digital learning games on learning. These detectors should also be applied to log data from other digital learning games and, ideally, non-game, computer-based controls. Given the large number of game features present across the diversity of digital learning games [9], it is important to explore whether gaming the system is reduced by a variety of games or if this mechanism is related to specific game features present in *Decimal Point.* Future research could explore how manipulating other game features, such as agency, might influence students' behavioral interactions and affective states [29]. Ultimately, understanding the connection between specific game features, cognitive and affective learning processes, and learning outcomes will provide digital learning game designers and teachers with a much more robust set of tools for determining when and how to implement digital learning games to best support students' learning. For example, if particular game features are especially effective at reducing problematic behaviors and affect (e.g., gaming, anxiety), a game with those features could be deployed when the context or content is likely to elicit those behaviors and affective states.

5 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 Jimit Bhalani, John Choi, Kevin Dhou, Darlan Santana Farias, Rosta Farzan, Jodi Forlizzi, Craig Ganoe, Rick Henkel, Scott Herbst, Grace Kihumba, Kim Lister, Patrick Bruce Gonçalves McLaren, and Jon Star for important contributions to the development and early experimentation with *Decimal Point*.

6 References

1. Aleven, V., McLaren, B.M., & Sewall, J.: 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 (2009).
2. Aleven, V., McLaren, B. M., Sewall, J., Van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., & Koedinger, K. R.: Example-tracing tutors: Intelligent tutor development for non-programmers. International Journal of Artificial Intelligence in Education, 26(1), 224-269 (2016).
3. Almeda, M. V., & Baker, R. S.: Predicting student participation in STEM careers: The role of affect and engagement during middle dchool. Journal of Educational Data Mining, 12(2), 33-47 (2020).
4. Bacher-Hicks, A., Goodman, J., & Mulhern, C.: Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. National Bureau of Economic Research, No. w27555 (2020).
5. Baker, R. S.: Gaming the system: A retrospective look. Philippine Computing Journal, 6(2), 9-13 (2011).
6. Baker, R. S., De Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T., & Koedinger, K. R.: Educational software features that encourage and discourage “gaming the system”. In Proceedings of the 14th International Conference on Artificial Intelligence in Education, pp. 475-482 (2009).
7. Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z.: Off-task behavior in the cognitive tutor classroom: when students "game the system." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 383-390 (2004).
8. Baker, R. S., Corbett, A. T., & Wagner, A. Z.: Human classification of low-fidelity replays of student actions. In Proceedings of the Educational Data Mining Workshop at the 8th International Conference on Intelligent Tutoring Systems, pp. 29-36 (2006).
9. Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E.: Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming*, *43*(6), 729-760 (2012).
10. BrightBytes, Inc.: 2020 Remote learning survey research results, https://www.brightbytes.net/rls-research, last accessed 2021/2/12.
11. Calvo, R. A., & D'Mello, S.: Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing, 1(1), 18–37 (2010).
12. Chen, T., & Guestrin, C.: Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794 (2016).
13. Cheng, M. T., Chen, J. H., Chu, S. J., & Chen, S. Y.: The use of serious games in science education: A review of selected empirical research from 2002 to 2013. Journal of Computers in Education, 2(3), 353–375 (2015).
14. Cheng, M. T., Rosenheck, L., Lin, C. Y., & Klopfer, E.: Analyzing gameplay data to inform feedback loops in The Radix Endeavor. Computers & Education, 111, 60–73 (2017).
15. Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C.: Eliciting self-explanations improves understanding. Cognitive Science, 18(3), 439-477 (1994).
16. Clark, D. B., Tanner-Smith, E., & Killingsworth, S.: Digital games, design, and learning: A systematic review and meta-analysis. Review of Educational Research, 86(1), 79-122 (2016).

17. Cocea, M., Hershkovitz, A., & Baker, R. S.: The impact of off-task and gaming behaviors on learning: immediate or aggregate?. In Proceeding of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, pp. 507-514, IOS Press (2009).
18. Common Sense Media. The common sense census: Media use by tweens and teens, https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens, last accessed 2021/2/12.
19. Crocco, F., Offenholley, K., & Hernandez, C.: A proof-of-concept study of game-based learning in higher education. Simulation & Gaming, 47(4), 403-22 (2016).
20. D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A.: Confusion can be beneficial for learning. Learning and Instruction, 29, 153-170 (2014).
21. Di Leo, I., Muis, K. R., Singh, C. A., & Psaradellis, C.: Curiosity… Confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. Contemporary Educational Psychology, 58, 121-137 (2019).
22. Fishman, B., Riconscente, M., Snider, R., Tsai, T., & Plass, J.: Empowering educators: Supporting student progress in the classroom with digital games. Ann Arbor: University of Michigan. gamesandlearning.umich.edu/agames (2014).
23. Forlizzi, J., McLaren, B., Ganoe, C., McLaren, P., Kihumba, G., & Lister, K.: Decimal Point: Designing and developing an educational game to teach decimals to middle school students. In: Busch, C. (ed.) Proceedings of the 8th European Conference on Games Based Learning (ECGBL-2014), 128-135 (2014).
24. Gagnon, D. J., Harpstead, E., & Slater, S.: Comparison of off the shelf data mining methodologies in educational game analytics. Proc. EDM, 38-43 (2019).
25. Gamesandlearning.org http://www.gamesandlearning.org/2014/06/09/teachers-on-using-games-in-class/, last accessed 2021/2/12.
26. Glasgow, R., Ragan, G., Fields, W.M., Reys, R., & Wasman, D.: The decimal dilemma. Teaching children mathematics, 7(2), 89-93 (2000).
27. Habgood, M.P.J., & Ainsworth, S.E.: Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. Journal of the Learning Sciences, 20(2), 169-206 (2011).
28. Harpstead, E., MacLellan, C. J., Aleven, V., & Myers, B. A.: Using extracted features to inform alignment-driven design ideas in an educational game. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3329-3338, ACM (2014).
29. Harpstead, E., Richey, J. E., Nguyen, H., & McLaren, B. M.: Exploring the subtleties of agency and indirect control in digital learning games. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 121-129, ACM (2019).
30. Hayes, A. F.: Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Publications (2017).
31. Irwin, K.C.: Using everyday knowledge of decimals to enhance understanding. Journal for Research in Mathematics Education, 32(4), 399-420 (2001).
32. Johnson, C. I., & Mayer, R. E.: Adding the self-explanation principle to multimedia learning in a computer-based game-like environment. Computers in Human Behavior, 26, 1246-1252 (2010).
33. Ke, F.: Designing and integrating purposeful learning in game play: A systematic review. Educational Technology Research and Development, 64(2), 219-44 (2016).
34. Lee, D. M. C., Rodrigo, M. M. T., Baker, R. S., Sugay, J. O., & Coronel, A.: Exploring the relationship between novice programmer confusion and achievement. In International conference on affective computing and intelligent interaction (pp. 175-184). Springer, Berlin, Heidelberg (2011).

35. Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., ... & Graesser, A. Inducing and tracking confusion with contradictions during complex learning. International Journal of Artificial Intelligence in Education, 22(1-2), 85-105 (2013).
36. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R.: Sequences of frustration and confusion, and learning. In Educational Data Mining (2013).
37. Mayer, R. E.: Computer games in education. Annual Review of Psychology. 70:531–49 (2019).
38. McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J.: A computer-based game that promotes mathematics learning more than a conventional approach. International Journal of Game-Based Learning (IJGBL), 7(1), 36-56 (2017).
39. Mogessie, M., Richey, J. E., McLaren, B. M., Andres-Bray, J. M. L., & Baker, R. S.: Confrustion and gaming while learning with erroneous examples in a decimals game. In International Conference on Artificial Intelligence in Education, pp. 208-213, Springer, Cham (2020).
40. Nicholson, S.: A user-centered theoretical framework for meaningful gamification. Games+ Learning + Society, 8(1), 223-230 (2012).
41. Nicholson, S.: Two paths to motivation through game design elements: Reward-based gamification and meaningful gamification. Proceedings of the iConference 2013, pp. 671-672 (2013).
42. Paquette, L., Baker, R. S., & Moskal, M.: A system-general model for the detection of gaming the system behavior in CTAT and LearnSphere. In International Conference on Artificial Intelligence in Education, pp. 257-260, Springer, Cham (2018).
43. Paquette, L., de Carvahlo, A., Baker, R., & Ocumpaugh, J.: Reengineering the feature distillation process: A case study in detection of gaming the system. In Educational Data Mining (2014).
44. Parong, J, Wells, A., & Mayer, R. E.: Replicated evidence towards a cognitive theory of game-based training. Journal of Educational Psychology, 112(5), 922-937 (2020).
45. Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M., Star, J. R., Baker, R. S., & McLaren, B. M.: More confusion and frustration, better learning: The impact of erroneous examples. Computers & Education, 139, 173-190 (2019a).
46. Richey, J. E., McLaren, B. M., Andres-Bray, M., Mogessie, M., Scruggs, R., Baker, R., & Star, J.: Confrustion in learning from erroneous examples: Does type of prompted self-explanation make a difference?. In International Conference on Artificial Intelligence in Education, pp. 445-457, Springer, Cham (2019b).
47. Riconscente, M. M. Results from a controlled study of the iPad fractions game Motion Math. Games and Culture, 8(4), 186-214 (2013).
48. Rodrigo, M.M.T., Baker, R.S., Jadud, M.C., Amarra, A.C.M., Dy, T., Espejo-Lahoz, M.B.V., Lim, S.A.L., Pascua, S.A.M.S., Sugay, J.O., & Tabanao, E.S.: Affective and behavioral predictors of novice programmer achievement. Proceedings of the 14th ACM-SIGCSE Annual Conference on Innovation and Technology in Computer Science Education, 156-160 (2009).
49. San Pedro, M.O.Z., Baker, R.S.J.d., Bowers, A.J., Heffernan, N.T.: Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. Proceedings of the 6th International Conference on Educational Data Mining, 177-184 (2013).
50. Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., et al.: Investigating optimal learning moments in U.S. and Finnish science classes. Journal of Research in Science Teaching, 53(3), 400–421 (2015).

51. Seaborn, K., & Fels, D. I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies, 74, 14-31 (2015).
52. Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B.: Application of learning analytics in educational videogames. Entertainment Computing, 5(4), 313-322 (2014).
53. Siew, N. M., Geofrey, J., & Lee, B. N.: Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using Dragonbox 12+ app. Electronic Journal of Mathematics & Technology, 10(2) (2016).
54. Sitzmann, T.: A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychology, 64, 489–528 (2011).
55. Slater, S., Ocumpaugh, J., Baker, R., Scupelli, P., Inventado, P.S., & Heffernan, N.: Semantic features of math problems: Relationships to student learning and engagement. Proceedings of the 9th International Conference on Educational Data Mining, 223-230 (2016).
56. Stacey, K., Helme, S., & Steinle, V.: 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, pp. 217-224, Utrecht, PME (2001).
57. Suh, S., Kim, S.W., & Kim, N.J.: Effectiveness of MMORPG-based instruction in elementary English education in Korea. Journal of Computer Assisted Learning, 26, 370-378 (2010).
58. Tokac, U., Novak, E., & Thompson, C. G.: Effects of game-based learning on students' mathematics achievement: A meta-analysis. Journal of Computer Assisted Learning, 35(3), 407-420 (2019).
59. Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M.: Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*(3), 229-243 (2006).
60. Wouters, P., & van Oostendorp, H.: (Eds.). Instructional techniques to facilitate learning and motivation of serious games. New York: Springer (2017).
61. Wu, C. H., Huang, Y. M., & Hwang, J. P.: Review of affective computing in education/learning: Trends and challenges. British Journal of Educational Technology, 47(6), 1304–1323 (2015).
62. Yip, F.W.M., & Kwan, A.C.M.: Online vocabulary games as a tool for teaching and learning English vocabulary. Educational Media International, 43, 233-249 (2006).