How Does Order of Gameplay Impact Learning and Enjoyment in a Digital Learning Game?

Yeyu Wang[1(✉)], Huy Nguyen[2], Erik Harpstead[2], John Stamper[2], and Bruce M. McLaren[2]

[1] University of Pennsylvania, Philadelphia, PA, USA
wangyeyu215@gmail.com
[2] Carnegie Mellon University, Pittsburgh, PA, USA

Abstract. When students are given agency in playing and learning from a digital learning game, how do their decisions about sequence of gameplay impact learning and enjoyment? We explored this question in the context of *Decimal Point*, a math learning game that teaches decimals to middle-school students. Our analysis is based on students in a high-agency condition, those who can choose the order of gameplay, as well as when to stop. By clustering student mini-game sequences by edit distance – the number of edit operations to turn one sequence into another – we found that, among students who stopped early, those who deviated more from a canonical game sequence reported higher enjoyment than those who did not. However, there were no differences in learning gains. Our results suggest that students who can self-regulate and exercise agency will enjoy the game, but the type and number of choices may also have an impact on enjoyment factors. At the same time, more investigation into the amount and means of delivering instruction to maximize learning efficiency within the game is necessary. We conclude by discussing digital learning game design lessons to create a game that more closely aligns with students' learning needs and affective states.

Keywords: Digital learning game · Decimal number · Edit distance · Clustering

1 Introduction

An important aspect of digital learning game design is deciding which gameplay elements the players (i.e., students) can control. In a typical game environment, players are offered a lot of agency - the capability to make their own decisions about how, what, and when they play. However, agency, which is often associated with engagement and enjoyment [41], may or may not be helpful to learning. Another nuance present in digital learning games is whether students should be given instructionally relevant choices, since young learners often have difficulty in making effective instructional decisions [33], in many cases resorting to unthoughtful choices [44].

One way to enhance students' experience and outcomes, while still giving them control over instructionally relevant aspects of gameplay, is to provide a recommendation

© Springer Nature Switzerland AG 2019
S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 518–531, 2019.
https://doi.org/10.1007/978-3-030-23204-7_43

feature within the game that can suggest the (potentially) optimal next step, without reducing the students' sense of agency. To achieve this, an important step is examining the influences of different problem sequences and identifying those that are most beneficial in terms of learning, enjoyment, or ideally both. We examined this question in *Decimal Point*, a digital learning game composed of a variety of mini-games designed to help middle-school students learn decimals [30]. While the original version of the game features a canonical sequence of mini-games that aims at interleaving various problem types and visual themes, it is not designed to be optimal for both learning and enjoyment for all students. To build a recommender capability as outlined above, we would need to identify the features of a good sequence while, at the same time, noting that these features may vary based on individual students.

To tackle this issue, prior studies of *Decimal Point* have compared learning and enjoyment between a high- and low-agency condition [23, 34]. The high-agency group could play the mini-games in any order and also had the option to stop playing early or play extra games. In contrast, the low-agency group had to play all mini-games in a fixed order. Expanding on this work, we focused solely on the high-agency students and explored potential differences among them in our analysis. In other words, given that high-agency students can make their own choices about mini-game selection, how would different selection orders (i.e., *game sequences*) impact their experience? More specifically, we investigated the following research questions:

RQ1: *How do students' game sequences impact their self-reported enjoyment of the digital learning game?*

RQ2: *How do students' game sequences impact their learning outcomes from the digital learning game?*

2 Background

2.1 The *Decimal Point* Game

Decimal Point is a single-player game that helps middle-school students learn about decimal numbers and their operations (e.g., adding, ordering, comparing). The game is based on an amusement park metaphor (Fig. 1), where students travel to different areas of the park, each with a theme (e.g., Haunted House, Sports World), and play a variety of mini-games, each targeting a common decimal misconception [19, 25, 52].

In the original game [30], students were prompted to play the mini-games in a predefined, canonical sequence, according to the dashed line shown in Fig. 1A (starting from the upper left). This sequence was originally developed to maintain thematic cohesion and to interleave problem types, which has been shown to improve mathematics learning [37, 38]. However, it is unclear whether a different sequence could be more or less beneficial to students. A subsequent study by [34] further explored agency by comparing two versions of *Decimal Point*: high-agency and low-agency. In the high-agency condition, students could play the mini-games in any order, could stop halfway through (i.e., after 12 mini-games) or play extra rounds after finishing all 24 mini-games. In the low-agency condition, students played all mini-games in a fixed order, without the option to stop early or play more.

Fig. 1. The different game maps used in (A) low-agency and high-agency with line, and (B) high-agency without line. The filled circles denote completed mini-games.

The authors reported no differences in learning or enjoyment between the two conditions, and had two conjectures regarding the high-agency students. First, they may have been implicitly guided to follow the canonical sequence by the dashed line on the map (Fig. 1A), hence their experience was comparable to that of students in the low-agency condition. Second, high-agency students may not have felt that their specific mini-game choices were consequential, as they would either stop early or eventually end up having played all mini-games, same as the low-agency students. In other words, to these students, different game sequences may have seemed to result in the same outcome.

The first conjecture was confirmed by post-hoc analyses reported in [34] and in a follow-up study by [23]. [34] reported that 68% of high-agency students played only 24 mini-games, similar to those in the low-agency condition, in approximately the same order. The study in [23] introduced a new high-agency condition without the dashed line (Fig. 1B) and it was observed that students in this condition deviated from the canonical path significantly more than those in the original high-agency condition.

As the next step, in this paper, we investigate the second conjecture – whether different game sequences selected by students in the high-agency conditions (with and without the dashed line) can have an impact on learning and enjoyment.

2.2 Related Work

The high-agency version of *Decimal Point* has many characteristics of an exploratory learning environment (ELE) [1], where students are free to explore instructional materials rather than follow a predefined learning path. Other notable digital learning games of this type include *Physics Playground* [49], *iSTART-2* [50], *Quest Atlantis* [5] and *Crystal Island* [43]. Among these, *Crystal Island* has been the subject of an experimental manipulation similar to that of *Decimal Point*, with three agency conditions: (1) high-agency, where students could freely explore the game world and choose which activities to do and in what order, (2) low-agency, where students did the

same activities but had to follow a fixed order, and (3) no-agency, where students only observed a video of an expert playing the game. Study results from [43] showed that low-agency students demonstrated the greatest learning gains but also exhibited undesirable behaviors such as a propensity for guessing, suggesting that some degree of agency may be beneficial, but too much is not.

An important task in ELEs is modeling students' learning to provide effective interventions based on fine-grained interactions with the learning environment [1]. A useful metric that can be derived from these sequential data is the *distance* - a measure of how similar two sequences are. Prior research has shown that in digital learning games, the distances from students' problem-solving sequences to an expert solution sequence are correlated with their learning gains [42] and test performance [20]. It is also possible to compute the distances among students' own sequences to cluster them. Analysis of the resulting clusters has been instrumental in several ELE assessment tasks: identifying player strategies in an algorithmic puzzle-based game [24], distinguishing between low and high achieving students in a problem-solving tabletop application [29], exploring the solution space in an open-ended physics game [22], and so on.

Another focus of the current work is student enjoyment and how it may be influenced by gameplay choices. In general, digital learning games are effective at promoting engagement and enjoyment by giving students control over the learning environment [45, 50, 51]. However, the effect of student choices is also subject to several nuances. First, it can vary based on individual students' self-regulation skills [32]. Second, students need to feel that their choices are meaningful and acquire a sense of agency (for a detailed discussion of agency within *Decimal Point*, refer to [23] and [34]). Third, the type and number of choices may affect their utility. In particular, choices that reflect personal interest will have the greatest effect, yet a large number of choices can become discouraging [36]. We will elaborate on these nuances in our later discussion.

3 Context

The work reported in this paper is a post-hoc analysis of data collected from two prior studies of *Decimal Point* [23, 34]. We briefly introduce the way these studies were conducted here before describing our analysis approach.

The two prior studies involved a total of 484 students. In this work, we focused on only 287 of those students in two conditions, high-agency with line (HAL) and high-agency without line (HANL), since these were the groups of students who could make their own mini-game selections, as opposed to those in the low-agency condition who could not make such choices. We further removed students who did not finish all of the pre- and posttest materials and evaluation surveys, which are used to measure learning and enjoyment outcomes, reducing the sample to 235 students (110 male, 125 female). The digital learning game and study materials included the following:

Pretest, Immediate Posttest and Delayed Posttest. The pretest, immediate posttest and delayed posttest (one week after the posttest), were administered online. The tests are isomorphic to one another (i.e., the same types of problems in the same order) and contain decimal items similar to those found in the game (e.g., ordering, comparing,

and adding decimals). The tests were also counterbalanced across students (e.g., ABC, ACB, BAC, etc. for pre, post, and delayed). Learning gains from pretest to posttest and pretest to delayed posttest are used to measure learning outcome.

Intervention. Students playing the high-agency versions of the game were shown a game map depicted in Fig. 1A (for the HAL group) or B (for the HANL group), where they could make their mini-game selections. There is also a dashboard that provides information about the types of game activity, and shows current mini-game completion progress. After playing half of the mini-games, students would be notified that they could choose to stop playing at any time from this point. Once students finished all 24 mini-games, the map interface would be reset to allow each game to be played once more (with the same game mechanics but different question content). Hence, the number of mini-games played by each student ranges from 12 to 48.

Evaluation Questionnaire and Survey. After finishing the game, students were given an evaluation questionnaire and post-survey, which asked them to (1) rate their overall experiences using a 5-point Likert scale, with a variety of game enjoyment questions (e.g., "I liked doing this lesson"), (2) select their most favorite mini-game, and (3) reflect on their agency experience (e.g., "if you did this activity again, would you play fewer, the same, or more number of mini-games? Why?"). The scores from (1) are averaged to produce a measure of self-reported enjoyment.

4 Results

4.1 Game Sequence Clustering

Since there is no expert sequence in *Decimal Point* (as we previously mentioned, it is unclear if the canonical sequence is optimal), we did not measure deviation from expert solution like other studies [20, 42]. Instead, our goal was to look at trends in learning and enjoyment among students who played through the mini-games in a similar way. We took a clustering approach to create groups of students who played a similar sequence of mini-games and looked for differences between these groups. To be consistent with prior studies, and because it was shown to be useful for analyzing our type of sequential data [23, 34], we used the Damerau-Levenshtein edit distance [13] as a measure of similarity between sequences. This metric counts the minimal operations required to change one sequence to another using insertions, deletions, substitutions, and transpositions. The smaller the edit distance, the more similar two sequences are to one another. If the value is zero, the two sequences are identical; if the value is the sum of the two sequence lengths, they are completely distinct.

We then applied k-medoids clustering [6] with the pairwise edit distance matrix of all game sequences as input. In this way, students who played similar game sequences (i.e., have a smaller edit distance between one another) would be grouped within the same cluster. We experimented with different values of k (number of clusters) for k-medoids clustering. After searching from 2 to 20, we selected the optimal k value of 4, based on the best average Silhouette Coefficient [40]. The four cluster medoids are illustrated in Fig. 2. We named each cluster based on the key mnemonic features of its

medoid. The first is Canonical Sequence (CS) with a medoid sequence identical to the canonical, following the dashed line in Fig. 1A. The second is Initial Exploration (IE) because students played a few mini-games out of order at the beginning of gameplay before returning to the canonical sequence. The third and fourth are Half on Top (HT) and Half on Left (HL) respectively because their medoids only span a portion of the game map (the top half and left half, respectively). Descriptive statistics for all clusters are included in Table 1.

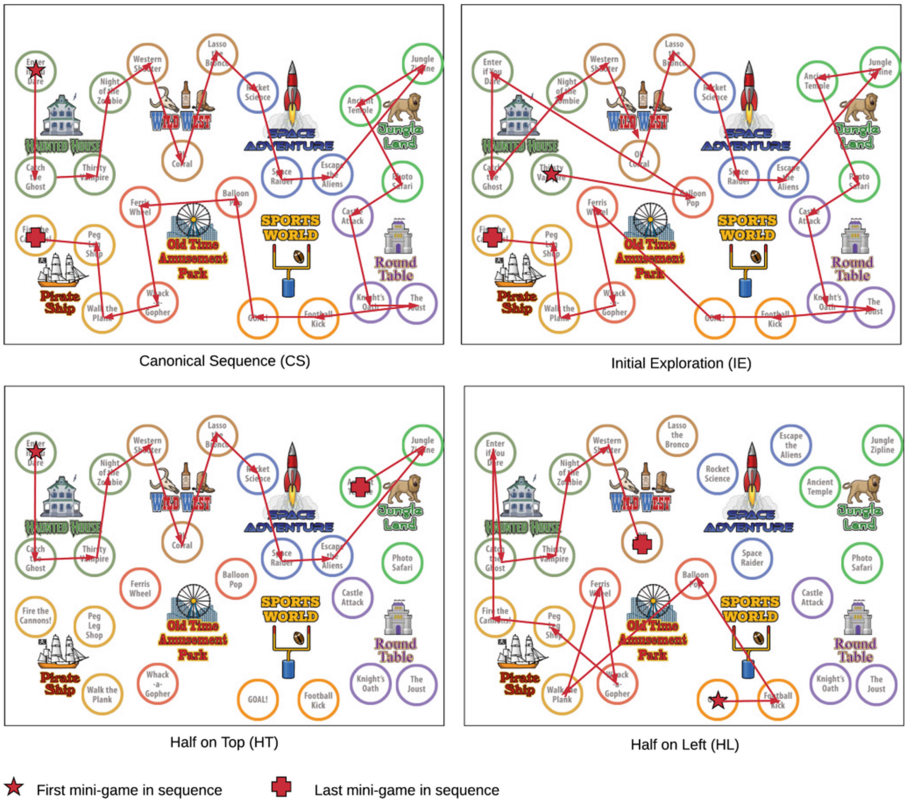

Fig. 2. Visualizations for the medoid game sequences in four clusters. Here the maps are shown without the line for clarity.

Table 1. Descriptive statistics for the four clusters.

Cluster	# of Students	# of Mini-games	Pretest	Immediate posttest	Delayed posttest	Enjoyment
CS	89	24.9 (4.6)	36.8 (13.1)	42.4 (10.3)	44.3 (10.1)	3.7 (0.8)
IE	14	26.9 (9.2)	37.3 (11.1)	42.4 (8.6)	42.9 (9.2)	3.8 (0.6)
HT	100	17.2 (6.3)	33.9 (12.9)	39.7 (11.7)	40.8 (12.3)	3.4 (1.0)
HL	32	19.6 (6.0)	40.0 (9.8)	43.3 (10.6)	44.3 (10.5)	3.8 (0.8)

To identify differences in learning and enjoyment across clusters, we conducted the Kruskal-Wallis test [14]. Kruskal-Wallis was chosen because our data did not satisfy the normality assumptions of an ANOVA. In the case of a significant difference, we used Dunn's post hoc [17] to perform pairwise comparisons between clusters. The effect size was also considered, following the thresholds of Cliff's Delta [39]. In this way, we examined our two research questions:

RQ1: *How do students' game sequences impact their self-reported enjoyment of the digital learning game?* Kruskal-Wallis test revealed a significant difference across the four clusters ($H = 10.248$, $p = 0.017$). Using Dunn's post hoc test with a Benjamini-Hochberg correction [8], we observed that Cluster HL had significantly higher enjoyment scores than HT, with a small effect size (Cliff's $d = 0.310$, $p = 0.007$), as shown in Table 2.

RQ2: *How do students' game sequences impact their learning outcomes from the digital learning game?* Kruskal-Wallis test showed no significant difference across clusters in gaining scores from pretest to immediate posttest ($H = 3.086$, $p = 0.378$) and from pretest to delayed posttest ($H = 2.585$, $p = 0.414$). Thus, clusters do not have a significant effect on students' learning outcomes.

Table 2. Multiple comparisons for enjoyment scores in different level (** - significant, p < adjusted α; ^ - small effect size).

Pairwise median of enjoyment scores for different clusters				Cliff's d	p-value	Adjusted α
IE	3.813	HL	3.813	–0.096	0.731	0.042
IE	3.813	CS	3.750	0.055	0.748	0.050
IE	3.813	HT	3.500	0.257	0.130	0.025^
HL	3.813	CS	3.750	0.115	0.326	0.033
HL	3.813	HT	3.500	0.310	0.007	0.008**^
CS	3.750	HT	3.500	0.195	0.018	0.017^

4.2 Post Analysis

Prior studies have established agency as a sense of freedom and control by the student [54], and in the context of our digital learning game, the amount of deviation from the canonical path [23]. Given that there is a significant difference in enjoyment scores between two clusters, we further explored the relationship between game sequence, agency, and enjoyment through the following two metrics.

Theme Transition Frequency (TTF). We expected that students who exercised agency would look at the entire map and explore different theme areas, as opposed to selecting a mini-game nearest to their current location or staying within one theme. While students could stay within a theme that they liked, we believed they were unlikely to enjoy every theme; therefore, we still expected to see more exploration. To measure this behavior, we defined a new metric, called *theme transition frequency*, as

the number of transitions between consecutive mini-games with different themes divided by the total number of transitions, for a given student. A value close to 1 means that the student tends to alternate between themes; a value close to 0 means that the student sticks to the same theme until all mini-games in that theme are completed. Next, we conducted Kruskal-Wallis test and found a significant difference in TTF across the four clusters ($H = 52.421$, $p < 0.0005$). To compare the TTF between pairs of clusters, we applied Dunn's post hoc test with Benjamini-Hochberg correction [8]. Cluster IE had significantly higher TTF than cluster CS, with a large effect size (Cliff's $d = 0.527$, $p = 0.004$). Cluster HL had significantly higher TTF than cluster CS with a large effect size (Cliff's $d = 0.749$, $p < 0.0005$) and higher than cluster HT with a small effect size (Cliff's $d = 0.244$, $p < 0.022$). Cluster HT had significantly higher TTF than cluster CS, with a medium effect size (Cliff's $d = -0.454$, $p = 0.017$) (Table 3).

Table 3. Multiple comparisons for TTF in different level (** - significant, p < adjusted α; ^ - small effect size, ^^ - medium effect size, ^^^ - large effect size)

Pairwise Median of TTF for Different clusters				Cliff's d	p-value	Adjusted α
IE	0.506	HL	0.675	–0.304	0.149	0.042
IE	0.506	CS	0.304	0.527	0.004	0.025**^^^
IE	0.506	HT	0.545	–0.041	0.996	0.050
HL	0.675	CS	0.304	0.749	< 0.0005	0.008**^^^
HL	0.675	HT	0.545	0.244	0.022	0.033**^
CS	0.304	HT	0.545	–0.454	< 0.0005	0.017**^^

Mini-game Preference. As the only difference in enjoyment we identified was among those who stopped early, in the HL and HT clusters, we conjectured that students may have had a stronger sense of enjoyment earlier in gameplay than towards the end. However, we did not have a mechanism to detect affective states over time. Therefore, as a proxy in examining this behavior, for each student, we looked at her self-reported favorite mini-game on the post-survey and where it occurred in her game sequence. More specifically, each student was labeled as one of three categories: (1) prefer one of the first three mini-games played, (2) prefer one of the last three mini-games played, and (3) prefer one between the first and last three mini-games. We then tested if the favorite mini-game is equally likely to appear in every part of the sequence. Since there are 24 mini-games in total, the null hypothesis is that the distribution of the three groups is 12.5%, 12.5% and 75% of the number of students respectively. We conducted a Chi-Square goodness of fit test [15] and found that this hypothesized distribution differs significantly from the empirical distribution of 30.2%, 59.6%, and 10.2% respectively ($\chi^2 = 67.42$, $df = 3$, $p < 0.0005$). In particular, the first category, despite covering only the first three mini-games, accounted for almost one-third of the most favorite mini-game responses, much higher than its expected portion of 12.5%. This result implies that students tended to prefer their initial gameplay experience.

5 Discussion

In this work we explored the question of whether different game sequences lead to different learning and/or enjoyment outcomes for students in the high-agency condition who could decide on their mini-game selections. Across the four identified clusters of game sequences – CS, IE, HT, HL – we found no differences in learning, but Cluster HL had significantly higher enjoyment scores than Cluster HT. We discuss this key result, as well as our other results, in the following paragraphs.

With respect to learning, we saw that the varied numbers of mini-games played by students across the clusters did not result in learning differences. This outcome is consistent with [23], and the authors' proposed explanations are also applicable in our case. Students who stopped early may have been able to self-regulate their learning and learned as much as those who played all mini-games, resulting in more efficient learning [31]. Alternatively, it is possible that there is more instructional content than required for mastery in the game, so students who played all of the mini-games essentially over-practiced rather than being less efficient. There have been debates about the varying effects of over-practice; some researchers claim that it leads to decreased learning efficiency [10, 12], while others suggest it yields higher levels of fluency [27] and better long-term outcomes [16]. In our case, it appears that over-practice, if present, had a neutral effect, since students who potentially over-practiced achieved the same learning gains as those who did not. A step toward better understanding this would be to construct a knowledge component (KC) model of students' in-game learning [21] so that learning efficiency and over-practice can be validated through Bayesian knowledge tracing [12] and learning curve analysis [18, 28, 53]. Such a KC model could also be displayed to students to facilitate awareness of progress and self-regulation, in the form of an open learner model [9].

With respect to enjoyment, while students in HL and HT both played approximately half of the mini-games, the former played the most mini-games out of order, while the latter tended to follow the canonical sequence. This distinction, demonstrated by our analysis of theme transition frequency, suggests that the HL group exercised more agency and enjoyed the game more than HT. On the other hand, we expected that students in CS and IE would have more enjoyment than those in HL and HT, because the former group also had the option to stop early yet chose to continue playing. However, we did not observe this difference. One possible explanation is that students in CS and IE did not stop early because they were not good at self-regulating their learning, rather than because they were enjoying the game more. This idea is supported by the observation from Fig. 2 that the game sequences in CS and IE are close to the canonical sequence, suggesting that students did not exercise agency in mini-game selection. A second explanation is that the novelty of the game environment may wear off towards the end, i.e., students may have experienced a "burnout effect" with diminished feeling of progression [11], which influenced their rating of overall enjoyment. Survey responses of mini-game preference did in fact show that students tended to favor the initial mini-games. A potential reason for this phenomenon is the nature of choices in *Decimal Point*. According to [7], engaging in choices or self-control is effortful and draws on limited resources. Therefore, a large number of choices

can become overwhelming [26, 46], and making several independent choices in a limited time may result in fatigue or ego-depletion [36]. In the high-agency condition, students first have to select one of the 24 mini-games, then one of the 23 remaining mini-games, and so on. Those who played all mini-games had to make 24 such selections within the timeframe of the study, so they may have experienced ego-depletion, which resulted in reduced enjoyment. Also, towards the end of gameplay, students do not have as many options to pick from because the completed mini-games are blocked from re-selection; however, this lack of choice may instead lead to decreased sense of agency. [36] suggested that there is an optimum number of choices that balances between the cognitive load from too many choices and the lack of agency from too few. Identifying this number for *Decimal Point* is left for future work.

In summary, we derived the following game design lessons from our analyses. First, one should aim for just the right amount of instructional content so that students can master the materials yet not incur the potential negative effects of over-practice. It can be difficult to initially estimate how much content is sufficient, but educational data mining techniques (e.g., learning curve analysis [21, 28]) can help revise and improve the materials in subsequent iterations. In addition, like *Decimal Point*, a game could allow students to control how much practice they are given, with proper scaffolding to assist them in self-regulating (e.g., an open learner model [9]). Second, when providing students with instructionally relevant choices, one should take into account factors such as agency, burnout and ego-depletion in designing the type and number of choices [11, 36]. Third, when collecting data from survey questions, one should note that students tend to report on their most recent experience, near the end of gameplay, rather than the overall experience.

Finally, we should point out that in this work, posttest scores and survey responses were used to measure the impacts of game sequence clusters. While these metrics are consistent with our prior studies [23, 34], it is possible that more fine-grained measures, for example those taken after each in-game action or mini-game played, would provide a better understanding of the influences of game sequences. In particular, we can use moment-by-moment learning models [3] to understand whether immediate or delayed learning takes place, and learning curve analysis [12] to track students' performance over time. For enjoyment, we will analyze learner affect by integrating automated affect detectors [2, 4, 35] in our data collection and analysis procedures, which can yield more reliable results than survey responses alone. This direction is consistent with the view of digital learning game researchers that students' learning and enjoyment should be assessed by in-game data rather than external measures [47, 48].

6 Conclusion

Our work investigated the effects of game sequences in *Decimal Point*. There were no differences in learning across sequence clusters, However, among students who chose to stop playing early, at around half of the mini-games, those who deviated more from the canonical order and switched between theme areas reported higher enjoyment scores. These results lead to important questions about the amount of instructional content, the nature of choices, and the interplay of various engagement factors in the

context of digital learning games. We intend to investigate these questions in future work to better understand the dynamics of students' game experience. This, in turn, will help us develop better AI techniques to personalize the game for increased enjoyment and learning.

Acknowledgements. This work was supported by NSF Award #DRL-1238619. The opinions expressed are those of the authors and do not represent the views of NSF. Special thanks to J. Elizabeth Richey for offering valuable feedback. Thanks to Scott Herbst, Craig Ganoe, Darlan Santana, Farias, Rick Henkel, Patrick B. McLaren, Grace Kihumba, Kim Lister, Kevin Dhou, John Choi, and Jimit Bhalani, all of whom made important contributions to the development of the Decimal Point game.

References

1. Amershi, S., Conati, C.: Combining unsupervised and supervised classification to build user models for exploratory. JEDM J. Educ. Data Min. **1**, 18–71 (2009)
2. Baker, R.S.J.d., et al.: Towards sensor-free affect detection in cognitive tutor algebra. In: Proceedings of the 5th International Conference on Educational Data Mining, pp. 126–133 (2012)
3. Baker, R.S., Goldstein, A.B., Heffernan, N.T.: Detecting learning moment-by-moment. Int. J. Artif. Intell. Educ. **21**, 5–25 (2011)
4. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Larusson, J. A., White, B. (eds.) Learning Analytics, pp. 61–75. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-3305-7_4
5. Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., Solomou, M.: Game-based curriculum and transformational play: designing to meaningfully positioning person, content, and context. Comput. Educ. **58**, 518–533 (2012)
6. Bauckhage, C.: Numpy/scipy recipes for data science: k-medoids clustering. Researchgate. Net, February 2015
7. Baumeister, R.F., Bratslavsky, E., Muraven, M.: Ego depletion: is the active self a limited resource? In: Self-Regulation and Self-Control, pp. 24–52. Routledge, New York (2018)
8. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. **57**, 289–300 (1995)
9. Bull, S., Nghiem, T.: Helping learners to understand themselves with a learner model open to students, peers and instructors. In: Proceedings of Workshop on Individual and Group Modelling Methods that Help Learners Understand Themselves, International Conference on Intelligent Tutoring Systems, pp. 5–13. Citeseer (2002)
10. Cen, H., Koedinger, K.R., Junker, B.: Is over practice necessary?-Improving learning efficiency with the cognitive tutor through educational data mining. Front. Artif. Intell. Appl. **158**, 511 (2007)
11. Cook, D.: What are game mechanics. Lost Gard (2006)
12. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. User Model. User-Adapt. Interact. **4**, 253–278 (1994)
13. Damerau, F.J.: A technique for computer detection and correction of spelling errors. Commun. ACM **7**, 171–176 (1964)
14. Daniel, W.W.: Kruskal–Wallis one-way analysis of variance by ranks. Appl. Nonparametric Stat. 226–234 (1990)

15. Dodge, Y.: The Concise Encyclopedia of Statistics. Springer, New York (2008). https://doi.org/10.1007/978-0-387-32833-1
16. Doroudi, S., Holstein, K., Aleven, V., Brunskill, E.: Towards understanding how to leverage sense-making, induction and refinement, and fluency to improve robust learning. In: Proceedings of the 8th International Conference on Educational Data Mining, pp. 376–379 (2015)
17. Dunn, O.J.: Multiple comparisons using rank sums. Technometrics **6**, 241–252 (1964)
18. Eagle, M., et al.: Estimating individual differences for student modeling in intelligent tutors from reading and pretest data. In: Micarelli, A., Stamper, J., Panourgia, K. (eds.) ITS 2016. LNCS, vol. 9684, pp. 133–143. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39583-8_13
19. Glasgow, R., Ragan, G., Fields, W.M., Reys, R., Wasman, D.: The decimal dilemma. Teach. Child. Math. **7**, 89 (2000)
20. Hao, J., Shu, Z., von Davier, A.: Analyzing process data from game/scenario-based tasks: an edit distance approach. JEDM-J. Educ. Data Min. **7**, 33–50 (2015)
21. Harpstead, E., Aleven, V.: Using empirical learning curve analysis to inform design in an educational game. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, pp. 197–207. ACM (2015)
22. Harpstead, E., MacLellan, C.J., Koedinger, K.R., Aleven, V., Dow, S.P., Myers, B.: Investigating the solution space of an open-ended educational game using conceptual feature extraction (2013)
23. 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, pp. 121–129. ACM (2019)
24. Horn, B., Hoover, A.K., Barnes, J., Folajimi, Y., Smith, G., Harteveld, C.: Opening the black box of play: Strategy analysis of an educational game. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, pp. 142–153. ACM (2016)
25. Isotani, S., McLaren, B.M., Altman, M.: Towards intelligent tutoring with erroneous examples: a taxonomy of decimal misconceptions. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6095, pp. 346–348. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13437-1_66
26. Iyengar, S.S., Lepper, M.R.: Rethinking the value of choice: a cultural perspective on intrinsic motivation. J. Pers. Soc. Psychol. **76**, 349 (1999)
27. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. Cogn. Sci. **36**, 757–798 (2012)
28. Koedinger, K.R., Stamper, J.C., McLaughlin, E.A., Nixon, T.: Using data-driven discovery of better student models to improve student learning. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 421–430. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_43
29. Maldonado, R.M., Yacef, K., Kay, J., Kharrufa, A., Al-Qaraghuli, A.: Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In: Educational Data Mining 2011 (2010)
30. McLaren, B.M., Adams, D.M., Mayer, R.E., Forlizzi, J.: A computer-based game that promotes mathematics learning more than a conventional approach. Int. J. Game-Based Learn. IJGBL **7**, 36–56 (2017)
31. McLaren, B.M., Lim, S.-J., Koedinger, K.R.: When and how often should worked examples be given to students? New results and a summary of the current state of research. In: Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 2176–2181 (2008)

32. McNamara, D.S., Shapiro, A.M.: Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. J. Educ. Comput. Res. **33**, 1–29 (2005)
33. Metcalfe, J., Kornell, N.: The dynamics of learning and allocation of study time to a region of proximal learning. J. Exp. Psychol. Gen. **132**, 530 (2003)
34. Nguyen, H., Harpstead, E., Wang, Y., McLaren, B.M.: Student agency and game-based learning: a study comparing low and high agency. In: Penstein Rosé, C., et al. (eds.) AIED 2018. LNCS (LNAI), vol. 10947, pp. 338–351. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93843-1_25
35. Ocumpaugh, J.: Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual. New York, NY Manila Philippines. Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences, vol. 60 (2015)
36. Patall, E.A., Cooper, H., Robinson, J.C.: The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. Psychol. Bull. **134**, 270 (2008)
37. Patel, R., Liu, R., Koedinger, K.R.: When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication. In: CogSci (2016)
38. Rohrer, D., Dedrick, R.F., Burgess, K.: The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. Psychon. Bull. Rev. **21**, 1323–1330 (2014)
39. Romano, J., Kromrey, J.D., Coraggio, J., Skowronek, J.: Appropriate statistics for ordinal level data: should we really be using t-test and Cohen'sd for evaluating group differences on the NSSE and other surveys. In: Annual Meeting of the Florida Association of Institutional Research, pp. 1–33 (2006)
40. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
41. Ryan, R.M., Rigby, C.S., Przybylski, A.: The motivational pull of video games: a self-determination theory approach. Motiv. Emot. **30**, 344–360 (2006)
42. Sawyer, R., Rowe, J., Azevedo, R., Lester, J.: Filtered time series analyses of student problem-solving behaviors in game-based learning. In: Proceedings of the 11th International Conference on Educational Data Mining, pp. 229–238 (2015)
43. Sawyer, R., Smith, A., Rowe, J., Azevedo, R., Lester, J.: Is more agency better? The impact of student agency on game-based learning. In: André, E., Baker, R., Hu, X., Rodrigo, Ma.M.T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 335–346. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61425-0_28
44. Schneider, W.: The development of metacognitive knowledge in children and adolescents: major trends and implications for education. Mind Brain Educ. **2**, 114–121 (2008)
45. Schønau-Fog, H., Bjørner, T.: "Sure, I Would Like to Continue" a method for mapping the experience of engagement in video games. Bull. Sci. Technol. Soc. **32**, 405–412 (2012)
46. Schwartz, B.: Self-determination: the tyranny of freedom. Am. Psychol. **55**, 79 (2000)
47. Shaffer, D.W., Gee, J.P.: The right kind of GATE: computer games and the future of assessment. In: Mayrath, M.C., Clarke-Midura, J., Robinson, D.H., Schraw, G. (eds.) Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications From Modern Research. Information Age Publications, Charlotte, NC, 211–228 (2012)
48. Shute, V.J., Ventura, M.: Stealth Assessment: Measuring and Supporting Learning in Video Games. MIT Press, Boca Raton (2013)
49. Shute, V.J., Ventura, M., Kim, Y.J.: Assessment and learning of qualitative physics in newton's playground. J. Educ. Res. **106**, 423–430 (2013)
50. Snow, E., Jacovina, M., Varner, L., Dai, J., McNamara, D.: Entropy: a stealth measure of agency in learning environments. In: Educational Data Mining 2014 (2014)

51. Spires, H.A., Rowe, J.P., Mott, B.W., Lester, J.C.: Problem solving and game-based learning: effects of middle grade students' hypothesis testing strategies on learning outcomes. J. Educ. Comput. Res. **44**, 453–472 (2011)
52. 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: PME Conference, pp. 4–217 (2001)
53. Stamper, J.C., Koedinger, K.R.: Human-machine student model discovery and improvement using dataShop. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS (LNAI), vol. 6738, pp. 353–360. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_46
54. Wardrip-Fruin, N., Mateas, M., Dow, S., Sali, S.: Agency reconsidered. In: DiGRA Conference. Citeseer (2009)