Nguyen, H., Hou, X., Stec, H., Di, S., Stamper, J., & McLaren, B.M. (2023). Examining the benefits of prompted self-explanation for problem-solving in a decimal learning game. *Proceedings of 24th International Conference on Artificial Intelligence in Education (AIED 2023).*

Examining the Learning Benefits of Different Types of Prompted Self-explanation in a Decimal Learning Game

Huy A. Nguyen[1], Xinying Hou[2], Hayden Stec[1], Sarah Di[1], John Stamper[1] and Bruce M. McLaren[1]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
[2] University of Michigan, Ann Arbor MI 48104, USA
hn1@cs.cmu.edu

Abstract. While self-explanation prompts have been shown to promote robust learning in several knowledge domains, there is less research on how different self-explanation formats benefit each skill set in a given domain. To address this gap, our work investigates 214 students' problem-solving performance in a learning game for decimal numbers as they perform self-explanation in one of three formats: multiple-choice (N = 52), drag-and-drop (N = 72) and open-ended (N = 67). We found that self-explanation performance in all three formats was positively associated with problem-solving performance. At the same time, we observed better outcomes with the drag-and-drop format than the open-ended format for solving decimal addition problems that do not remind students about carrying, but worse outcomes than the multiple-choice and open-ended format for other problem types. These results point to the nuanced interactions between the problem type and self-explanation format that should be further investigated to identify when and how self-explanation is most helpful for learning.

Keywords: decimal learning, problem-solving, self-explanation format

1 Introduction

Prompted self-explanation, where the student self-explains what they have learned or how they solved a problem, is a highly robust and effective learning strategy [2]. There are different ways to implement this activity in a learning system, such as multiple-choice, drag-and-drop and open-ended, which vary in their constructive or active nature [2]. Yet there is mixed evidence regarding their relative effectiveness. While [2] and [8] pointed to open-ended prompts being more beneficial than multiple-choice prompts, findings from [4] showed that the multiple-choice prompts led to better transfer. Drag-and-drop self-explanation is less explored, with one study showing it did not differ in learning benefits from the other formats [8], but drag-and-drop questions have become increasingly popular in computer-based assessments, with the advantage of promoting interactivity and reducing random guessing [1].

Our conjecture regarding these mixed results is that the effectiveness of each self-explanation format could differ across knowledge domains, or even across problem types within a domain. To examine this possibility, we performed a post-hoc analysis of our prior study [8] which compares the learning outcomes from multiple-choice, drag-and-drop and open-ended self-explanation prompts in a digital learning game for decimal numbers. In the game, students solve a variety of decimal problem types (e.g., locate decimal numbers on a number line, sort sequences of decimal numbers) and perform self-explanation after every two rounds of problem-solving [7]. While we have found that open-ended self-explanation prompts led to higher delayed posttest scores than multiple-choice prompts [8], our prior work did not investigate whether the relative benefits of these formats also differ by the decimal problem types. Thus, our current work focuses on addressing this gap.

2 A Digital Learning Game for Decimal Numbers

Decimal Point is a web-based single-player game that teaches decimal numbers and operations to middle-school students [7]. The game features an amusement park metaphor with 8 theme areas and 24 mini-games. Each mini-game consists of two problem-solving activities, designed to help students practice decimal procedures, and a self-explanation activity, designed to reinforce their learning [2]. Each problem-solving activity belongs to one of five problem types listed in Table 1. As an example, in a *Sequence* mini-game called *Escape the Aliens*, the two problem-solving activities involve filling in the 4th and 5th numbers in the sequences 0.0, 0.35, 0.70, ___, ___ and 0.0, 4.4, 8.8, ___, ___. After solving these problems, students are asked to self-explain how they would find the 6th number in the second sequence (Table 2).

Table 1. Descriptions and examples of the five decimal problem types in *Decimal Point*.

Problem type	Description	Example
Addition	Enter the carry and result digits when adding two decimal numbers.	Enter the carry and result digits when adding 7.50 to 3.90.
Bucket	Label each of the given decimals as “less than” or “greater than” a threshold.	**Given decimals**: 0.51, 0.132, 0.9, -0.833. **Threshold**: -0.4.
Number Line	Locate the position of a given decimal on a number line.	Place 0.111 on a number line from -1 to 1.
Sequence	Fill in the next two numbers in a given sequence of decimals.	Fill in the next two numbers: 0.33, 0.66, 0.99, ___, ___.
Sorting	Sort a given list of decimal numbers in ascending or descending order.	Sort the list 0.9063, 0.39, 0.291, 0.7 from small to large.

Students need to play through all 24 mini-games in a fixed order to finish the game. They also get immediate feedback about the correctness of their answer in both the problem-solving and self-explanation activities. If they make a mistake, they can retry any number of times until arriving at the correct answer, and must do so to advance to the next mini-game. Additionally, in the study of *Decimal Point* which we base our analysis on [8], students are randomly assigned to play one of three versions of the game with different self-explanation formats. In the control multiple-choice condition, students select a correct answer from three given options. In the drag-and-drop condition, students fill in the blanks of an answer statement by dragging items from a given word bank with 4 or 5 phrases. In the open-ended condition, students enter their responses in an open text box. To deter students from providing unthoughtful answers, we implemented a preliminary grading mechanism, using a keyword matching technique commonly applied to automated grading of short answers [10]. In particular, the game considers the student's answer as acceptable if it contains more than three words, with at least one of the words belonging to a set of keywords that would be found in a correct explanation. The keywords are pre-determined for every mini-game and are not revealed to the students. Table 2 shows the self-explanation question for the *Escape the Aliens* mini-game described earlier in each of the three game versions.

Table 2. A self-explanation question example in each of the three formats.

Multiple-choice	Drag-and-drop	Open-ended
The next number in the pattern can be found by adding 17.6 + 4.4. What is the answer and how do you know? ▢ 22.00 because you should carry a 1 over to the ones place. ▢ 21.10 because .6 + .4 = .10 and 4 + 17 = 21. ▢ 22.10 because .6 + .4 = .10 and you also carry a 1.	The next number in the pattern is 17.6 + 4.4, which equals _____ because when we add, we _____. **Word bank**: 22.0, 21.0, 22.10, carry once, add the right side separately.	The next number in the pattern can be found by adding 17.6 + 4.4. What is the answer and how do you know? _________

3 Methods

Our analysis involves 357 5th and 6th graders from a prior classroom study [8] in the spring of 2021, which includes a pretest, game play and immediate posttest in the first week, followed by a delayed posttest one week later. 143 students who either did not finish all of the materials or had already played through the game in a previous year were excluded. Among the remaining 214 students, there were 75 in the multiple-choice condition, 72 in the drag-and-drop condition, and 67 in the open-ended condition.

While students' self-explanations can be automatically graded in the multiple-choice and drag-and-drop formats, the open-ended responses require manual grading. Thus, we coded all of the 5,142 self-explanation responses from students as either *correct*, *incorrect* or *off-topic* (i.e., the response is unrelated to the question). Following [3], the coding process started with one member of the research team analyzing 1.5% of the responses to develop a rubric for the correct, incorrect and off-topic labels. Two other members then applied this rubric to independently code 20% of the responses, achieving an inter-rater reliability of $\kappa = 62.7\%$. Next, the two members met to resolve disagreements in their codes and refine the rubric, which was then applied by the first member to code the remaining 80% of the data. This coding process yielded 1,000 correct, 4,076 incorrect and 66 off-topic responses.

4 Results

Our first analysis investigated how students' self-explanation performance relates to their problem-solving performance. As we had reported the comparison of test scores by condition in [8], with significantly higher delayed posttest scores in the open-ended than in the multiple-choice condition, in this work we focus only on in-game performance. To measure student performance in the problem-solving and self-explanation activities, we computed the rate at which their first attempt is correct (i.e., the number of times their first attempt in a question is correct, divided by the number of questions). We only considered the first attempt because subsequent attempts may be influenced by the game's corrective feedback and therefore not reflective of the student's performance; this approach is also common in learning assessment on digital platforms [11]. Next, we built a regression model that predicts students' problem-solving performance (measured as the rate of correctness at the first problem-solving attempt, i.e., *PS-success-rate*) based on their prior knowledge (measured by pretest scores), study condition and self-explanation performance (measured as the rate of correctness at the first self-explanation attempt, i.e., *SE-success-rate*). The overall model achieved an R^2 value of .498, with the coefficient terms reported in Table 3. We observed that, when controlled for prior knowledge, students' self-explanation performance in all three formats had a significant and positive association with their overall problem-solving performance.

Table 3. Results of the regression model predicting *PS-success-rate*. Rows with (***) indicate significant coefficients ($p < .001$)

	Coefficient	Std Err	t
Intercept	0.405	0.022	18.284 (***)
Condition (drag-and-drop)	0.004	0.014	0.275
Condition (open-ended)	0.026	0.018	1.439
SE-success-rate	0.129	0.035	3.694 (***)
Pretest Score	0.007	0.001	9.981 (***)

Table 4. Descriptive statistics for *PS-success-rate* by problem type and study condition. Rows with (**) indicate significant condition differences ($p < .01$).

Problem type	Multiple-choice *M (SD)*	Drag-and-drop *M (SD)*	Open-ended *M (SD)*
Addition	0.660 (0.255)	0.660 (0.279)	0.698 (0.248)
Bucket	0.683 (0.248)	0.661 (0.225)	0.716 (0.238)
Number Line (**)	0.649 (0.227)	0.530 (0.237)	0.626 (0.233)
Sequence (**)	0.706 (0.098)	0.736 (0.099)	0.688 (0.104)
Sorting	0.605 (0.191)	0.603 (0.181)	0.629 (0.186)

Next, we examined how students' problem-solving performance in each decimal problem type differs across conditions. To measure problem-solving performance, we again used the rate of correctness at the first problem-solving attempt (*PS-success-rate*), but computed a separate value for each of the five decimal problem types (see Table 4 for descriptive statistics). We then conducted a series of ANCOVAs comparing students' rate of correctness in each decimal problem type across the three study conditions, with pretest scores as covariates to account for prior knowledge. Our results showed significant condition effects on *PS-success-rate* in the *Number Line* ($F(2, 210) = 5.136$, $p = .007$) and *Sequence* ($F(2, 210) = 5.391$, $p = .005$) problem types. Within the *Number Line* problems, post-hoc (Tukey) comparisons revealed significant differences between the drag-and-drop and multiple-choice conditions (p = .006), as well as between the drag-and-drop and open-ended conditions (p = .043), with the drag-and-drop condition leading to the lowest *PS-success-rate*. Within the *Sequence* problems, we similarly observed a significant difference between the drag-and-drop and open-ended conditions (p = .014), but with higher *PS-success-rate* in the drag-and-drop condition.

5 Discussion and Conclusion

This work examined the effects of prompted self-explanation formats on problem-solving performance in a digital learning game, *Decimal Point*. Our results showed that, when controlling for prior knowledge, self-explanation performance across all three formats is a significant and positive predictor of problem-solving performance. This benefit is consistent with prior cognitive theories and empirical evidence of self-explanation [2]. In the case of *Decimal Point*, the self-explanation questions were designed to target common decimal misconceptions [7]. Therefore, students who answered these questions correctly, especially on the first attempt, had acquired a good understanding of decimal concepts, which translated to better problem-solving. This connection suggests that self-explanation could also be helpful in other domains where students' learning difficulties lie primarily in their common misconceptions.

On the other hand, our analysis of individual decimal problem types revealed varying effects of the self-explanation format. First, drag-and-drop self-explanation prompts led to higher *PS-success-rate* than open-ended prompts in the *Sequence* problem type, which focuses on decimal additions. Because *Sequence* problems do not require specifying the carry digits during addition, the most common error in these problems is forgetting to carry across the decimal point [7]. In that case, the drag-and-drop prompts are better at addressing this misconception, thanks to the given word bank that always contained the term "carry," whereas the open-ended prompts provided no such reminders (Table 2). At the same time, we found that the *PS-success-rate* was lowest in the drag-and-drop condition across all other problem types, especially the *Number Line* type where the condition differences were significant. One explanation is that the drag-and-drop format could cause reading difficulties due to the blanks in the given self-explanation statements. Combined with evidence that young students' reading skills have declined during the COVID-19 pandemic [5], this suggests students may not have learned effectively from drag-and-drop self-explanation prompts because they had to exert more cognitive efforts on reading comprehension [6]. In turn, they would perform worse in most decimal problem types, especially *Number Line* which was identified as the most difficult type [9]. Finally, we note that there were no differences in the benefits of multiple-choice and open-ended self-explanation for solving any decimal problem type. On the other hand, our prior results did report higher delayed posttest scores from students who performed open-ended self-explanation [8], suggesting that the benefit of the open-ended format only manifests in the long term, but not during immediate game play.

In conclusion, our findings support the benefits of prompted self-explanation in decimal problem-solving, but also point to variances across problem types and self-explanation formats. At the same time, this work presents certain limitations that future research should address. First, the grading rubric of open-ended self-explanations can be refined to improve the graders' agreement and more accurately evaluate the students' work. Second, while several mini-games share the same decimal problem type, there may still be inherent differences in the game design or narrative that could influence students' performance. A more nuanced understanding of these latent factors will help identify when and how each self-explanation format is best for learning.

References

1. Arslan, B., Jiang, Y., Keehner, M., Gong, T., Katz, I.R., Yan, F.: The Effect of Drag-and-Drop Item Features on Test-Taker Performance and Response Strategies. Educational Measurement: Issues and Practice. 39, 2, 96–106 (2020).
2. Chi, M.T., Wylie, R.: The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational psychologist. 49, 4, 219–243 (2014).
3. DeCuir-Gunby, J.T., Marshall, P.L., McCulloch, A.W.: Developing and using a codebook for the analysis of interview data: An example from a professional development research project. Field methods. 23, 2, 136–155 (2011).
4. Johnson, C.I., Mayer, R.E.: Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. Computers in Human Behavior. 26, 6, 1246–1252 (2010).
5. Kuhfeld, M., Lewis, K., Peltier, T.: Reading achievement declines during the COVID-19 pandemic: evidence from 5 million US students in grades 3–8. Reading and Writing. 1–17 (2022).
6. McCray, G., Brunfaut, T.: Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. Language Testing. 35, 1, 51–73 (2018). https://doi.org/10.1177/0265532216677105.
7. 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).
8. McLaren, B.M., Richey, J.E., Nguyen, H.A., Mogessie, M.: Focused Self-Explanations Lead to the Best Learning Outcomes in a Digital Learning Game. In: Proceedings of the 17th International Conference of the Learning Sciences. pp. 1229–1232 ISLS (2022).
9. Nguyen, H., Hou, X., Stamper, J., McLaren, B.M.: Moving beyond Test Scores: Analyzing the Effectiveness of a Digital Learning Game through Learning Analytics. In: Proceedings of the 13th International Conference on Educational Data Mining. (2020).
10. Rokade, A., Patil, B., Rajani, S., Revandkar, S., Shedge, R.: Automated grading system using natural language processing. In: 2018 Second international conference on inventive communication and computational technologies (ICICCT). pp. 1123–1127 IEEE (2018).
11. Yu, H.-F., Lo, H.-Y., Hsieh, H.-P., Lou, J.-K., McKenzie, T.G., Chou, J.-W., Chung, P.-H., Ho, C.-H., Chang, C.-F., Wei, Y.-H.: Feature engineering and classifier ensemble for KDD cup 2010. In: Proceedings of the KDD Cup 2010 Workshop. (2010).