International Society of the Learning Sciences

Investigations into the Effects of Hint Reading Time and Hint Type in a Digital Mathematics Game Through Response Time Decomposition

Ji-Eun Lee, Erin Ottmar, Kirk Vanacore, Alena Egorova
jlee13@wpi.edu, erottmar@wpi.edu, kpvanacore@wpi.edu, aegorova@wpi.edu
Worcester Polytechnic Institute

Anthony Botelho, University of Florida, abotelho@coe.ufl.edu

Ashish Gurung, Carnegie Mellon University, agurung@andrew.cmu.edu

Abstract: This study explored 7th-grade students' help-seeking behaviors in a digital mathematics learning game. In particular, we applied response time decomposition to examine the relationships among the amount of time a student spent reading a hint (i.e., hint reading time), hint type, and in-game performance. The results showed significant differences in hint reading time by the following action and hint type. Overall, hint reading time was positively related to problem solving efficiency scores. However, this association varied by hint type. The use of an indirect hint was negatively correlated with efficiency on the problem in which the hint was accessed but positively correlated with efficiency on the next problem. These findings suggest that prompting students to pause and think about the hint presented may support efficient problem solving. Although indirect hints may cause students to struggle through the problem, they show the benefits of that struggle in subsequent performance.

Introduction

Students' ability to recognize when help is needed and seek the most appropriate external resource, so-called *help-seeking* behavior, is an important component of the learning process and cognitive development (Aleven et al., 2016; McLaren et al., 2022). In learning technologies, students' help-seeking during problem solving can be used as a behavioral indicator that reflects students' level of knowledge or understanding. Several studies have investigated how students' help-seeking behaviors (e.g., requesting hints) are related to learning outcomes in learning technologies, yet the findings are inconsistent. While some studies showed that students who requested hints more frequently outperformed on algebra assignments than those who did not (Kehrer et al., 2013; Razzaq & Heffernan, 2010), many studies in the field found that students' hint requests were negatively related to their learning outcomes (McLaren et al., 2022; O'Rourke et al., 2014). Furthermore, our prior study that investigated students' help-seeking behaviors in a digital mathematics game also found that requesting more hints was associated with lower math posttest scores (Iannacchione et al., 2023).

Then, is seeking help not helpful for learning? Some researchers argued that the negative impacts of help-seeking in learning technologies on students' performance might not be due to help (e.g., hints) itself, but due to their maladaptive help-seeking behaviors, for example, *help abuse*, which indicates using hints in ways that are not likely to promote learning (e.g., clicking through hints and not reading them carefully) (Aleven et al., 2016; Long & Aleven, 2013). Thus, when exploring students' help-seeking behaviors in learning technologies, it may be important to account for the ways in which students use hints, rather than whether or not the students use hints. However, much of the research has focused on students' simple usage of help rather than how effectively they use hints in learning technologies. Another important aspect of student help-seeking is the *type of hints,* which may also impact learning. Previous research has found variability in the effectiveness of hints based on hint format and content (Arroyo et al., 2000; Muir & Conati, 2012).

This study examines how middle school students use help (i.e., hints) in a digital mathematics game, From Here to There! (FH2T). In particular, we apply *response time decomposition* and use the amount of time they spend reading a hint as an indicator of students' productive help-seeking behaviors (Gurung et al., 2021). The content type of the hints is also considered in the model. We will address the following research questions.

- RQ1: Do students use help productively in a digital mathematics game as determined by the amount of time they spend reading a hint?
    - RQ1-1: Are there differences in students' hint reading time by the next action after reading a hint?
    - RQ1-2: Are there differences in students' hint reading time by hint type?
- RQ2: Does hint reading time and hint type influence the in-game performance?

- RQ1-1: To what extent do hint reading time and hint type predict problem-solving efficiency?
- RQ1-2: To what extent do hint reading time and hint type predict the problem-solving efficiency on the next problem?

Literature Review

Hint reading time as an indicator of students' productive help-seeking

Students' response time data collected in learning technologies can be used to estimate students' invisible cognitive or metacognitive processes. Several studies have used the amount of time students spend reading a hint as an indicator of students' thinking or effort to understand the provided hints (Gurung et al., 2021; Shih et al., 2011). For example, Gurung and colleagues (2021) explored the students' effort to understand hints in a computer-based learning platform through response time decomposition. They measured the amount of time the students spent between two consecutive actions involving hint requests as a first action (e.g., hint request-attempt). Then, they used these data to model students' effort and examined the relationship between effort and performance. The study found that most students tended to spend a longer time reading a hint on correct attempts, and the students who exhibited high effort (i.e., longer hint reading time) were more likely to get the next problem correct. Similarly, Long and Aleven (2013) investigated the relationship between high school students' help-seeking behaviors and test performance in an intelligent tutoring system to teach Geometry. The findings revealed that time spent on each hint was positively related to test performance, while the frequency of hint requests negatively correlated with the performance. Together, these findings suggest that the amount of time students spend reading a hint can be used as an indicator of students' cognitive process and effort in thinking about hints, and it may have positive impacts on students' performance. Thus, we hypothesize that the amount of time the students spend reading hints is positively related to in-game performance in FH2T.

Relationship between hint type and student learning

Hints can be divided into various categories by their format (e.g., video, text; Gurung et al., 2021), the degree of interactivity (Arroyo et al., 2000), and so on. Hints can also be categorized into two dimensions in terms of content explicitness: *direct hints* (or concrete, specific, explicit, bottom-out hints), which tell students exactly what to do and help them get a correct answer/solution, and *indirect hints* (or abstract, general, inexplicit, instrumental hints) that provide students information relevant to the correct answer/solution but do not tell them what exactly to do (Arroyo et al., 2000; Muir & Conati, 2012).

A number of studies have investigated how these different hint types influence students' learning. For example, O'Rourke et al. (2014) examined how the hint systems in an educational video game to teach fractions impacted elementary school students' performance. The findings indicated that the students who used concrete hints performed better than those who used abstract hints, concluding that concrete hints were more helpful than abstract hints. Similarly, Arroyo et al. (2000) found that direct hints were more effective for students with lower cognitive abilities in a mathematics intelligent tutoring system. Further, a few studies revealed that hint type is also associated with hint reading time. Muir and Conati (2012) investigated the relationship between students' attention to hints measured using eye-tracking data and hint type in a digital mathematics game. A statistically significant interaction was revealed between students' fixation time on hints and hint type: the students had longer fixation time on definition hints (e.g., providing a definition of a factor) compared to tool hints (e.g., providing information about a tool in the game) and bottom-out hints. O'Rourke et al. (2014) found that students spent a significantly longer time reading concrete hints (3.67 seconds) than abstract hints (3.17 seconds).

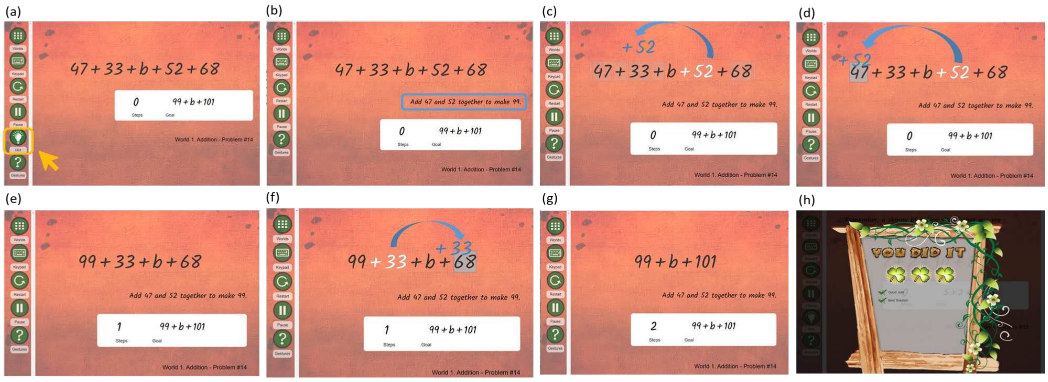
As such, some studies have shown that the hint type impacts students' hint reading time and performance; however, much less is known about the relationship between hint type and student learning and how hint reading time influences their relationship. Thus, this study examines the relationships among students' hint reading time, hint type, and their in-game performance.

Methods

Game description

FH2T (https://graspablemath.com/projects/fh2t) is a gesture-based digital interactive mathematics game developed based on cognitive and perceptual learning theories to improve students' algebraic learning. The objective of the game is to transform an algebraic expression (e.g., 47+33+b+52+68 in Figure 1) into a mathematically equivalent but perceptually different goal state expression (e.g., 99+b+101). The symbols and numbers in the game are reified as physical objects, which enables students to manipulate and transform them dynamically on the screen using a series of gesture actions. Each gesture-action leading to a valid transformation is considered a *step*. Students are encouraged to transform expressions from a starting state to a goal state using more efficient strategies involving fewer steps (Figure 1c-1g). Students receive rewards (clovers) if they reach the goal in the most efficient way using the minimum possible number of steps, called optimal step (Figure 1h). The number of clovers is deducted if they exceed the optimal step. If a student makes a mathematically invalid action (e.g., adding 7 and 2 before multiplying in 7+2*5), the term automatically snaps back to the starting position, signifying a mistake without indicating a valid action. The game also provides students a text-based hint for each problem upon students' requests (Figure 1a-1b). Only one hint is provided for each problem, and the hint button disappears after use. The game is composed of 14 levels that cover different math topics with increasing difficulty. Each level consists of 18 problems, and students can proceed to the next level if they complete 14 consecutive problems. Prior studies have shown that the game is effective in improving students' mathematical knowledge and problem-solving efficiency (Chan et al., 2021; Decker-Woodrow et al., 2021).

Figure 1
A Sample Problem, a Hint, and a Student's Action in From Here to There!



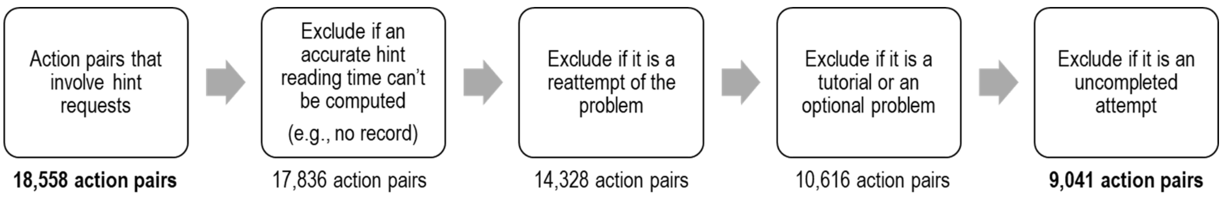
Participants and procedure

The data was drawn from a Randomized Control Trial (RCT) conducted in 2020-2021 that examined the efficacy of FH2T compared to two different learning technologies across four conditions (Decker-Woodrow et al., 2023). The participants of the RCT were 7th-grade students from 11 middle schools in one school district. The students were randomly assigned to one of the four conditions, and this study includes the students assigned to the FH2T condition (N = 1,430). The students took the pretest on their algebraic knowledge before the intervention and had nine 30-minute intervention sessions across the school year, administered online. The students played the game individually at their own pace so that all students had different numbers of problems completed. After the intervention, they had the posttest using the mirrored items to the pretest.

Data exploration and pre-processing

We first extracted action-level data from the FH2T database and created a column with the action pairs representing the sequence of the two actions involving hint requests as a first action (e.g., hint request [hereafter, hint]-reset). Note that the action pairs not involving a hint request were excluded as they were outside the scope of the study. The initial number of action pairs involving hint requests was 18,558. We first removed 722 action pairs for which hint reading time cannot be computed (e.g., hint-no record, hint-pause, hint-leave action pairs). As the game allows students to attempt the problem as many times as they want, we only included data on their first attempts. Thus, we excluded an additional 8,795 action pairs for re-attempts, tutorial/optional problems, and uncompleted attempts, resulting in an analytic sample of 9,041 action pairs (Figure 2). The unique number of students and the problems included in the analysis were 986 students and 156 problems, respectively.

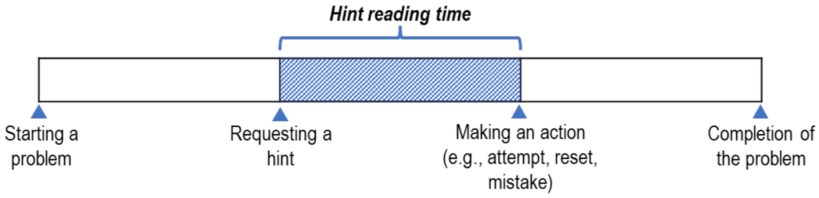
Figure 2
Data Preprocessing Process



Measures

Hint reading time (as an indicator of students' thinking/effort to understand the provided hints)
Hint reading time represents the time a student spends reading a hint and is measured through response time decomposition. Specifically, it was computed by subtraction on two time points: (1) requesting a hint and (2) making any following action after reading a hint (Figure 3). We applied a natural log transform to hint reading time as it was not normally distributed and used log-transformed data for further analyses.

Figure 3
A Visual Representation of the Response Time Decomposition of Hint Request and Reading



Action pairs
Action pairs represent two consecutive actions involving hint requests as a first action for each problem. Three action pairs were included in the analysis: making a mathematically valid attempt (hint-attempt), resetting to an initial state of the problem (hint-reset), and making a mathematically invalid attempt (hint-mistake). Table 1 shows the descriptions and examples of each action pair.

Table 1
Descriptions and Examples of Action Pairs

Action Pairs	Descriptions	Examples
hint-attempt	making a mathematically valid action after reading a hint	Multiplying before adding in 7+ 2*5 (7+2*5 → 7+10)
hint-mistake	making a mathematically invalid action after reading a hint	trying to add 7 and 2 before multiplying in 7+ 2*5
hint-reset	resetting the problem to an initial state after reading a hint	7+2*5 → 5+2+2*5

Hint content types
Based on prior literature, hints in the game were categorized into three types: direct, intermediate, and indirect hints (Table 2). Direct hints tell students exactly what to do, while indirect hints provide information relevant to the solution but do not tell students what to do. Intermediate hints are in between direct and indirect hints regarding the degree of explicitness. Two researchers individually hand-coded the hint type for all problems and had several rounds of discussions to reach an agreement. Note that only one text-based hint is provided for each problem. We then created three dummy variables to represent the type to which each hint belongs.

Table 2
Examples of Each Hint in FH2T

Hint type	Examples
Direct hint	● Click and move *a* to the first position on the left. ● Add 44 and 56 together to make 100.
Intermediate hint	● To match the goal, add twice and commute. ● Add and decompose to match the goal.
Indirect hint	● The order that you multiply numbers doesn't change the result. ● Which term has a factor of 11?

Efficiency scores (outcome variable for RQ2)
Efficiency refers to how efficiently a student solves a problem in the game. It is computed by dividing the fewest steps possible to reach the goal state (i.e., optimal step) by the number of steps made by the student for that problem. Thus, higher efficiency scores indicate more efficient problem solving, involving fewer computation

steps. For example, suppose a problem with a start state (5+13+4+5) and a goal state (10+7+4+6). If a student reaches the goal state using the optimal steps (e.g., 5+13+4+5 → [step 1] 5+7+6+4+5 → [step 2] 10+7+6+4 → [step 3] 10+7+6+4), the efficiency score for this student is equal to 1 (i.e., 3/3). If a student reaches the goal state using five steps (e.g., 5+13+4+5 → [step 1] 5+10+3+4+5 → [step 2] 5+10+1+2+4+5 → [step 3] 10+6+2+4+5 → [step 4] 10+6+7+4 → [step 5] 10+7+6+4), the efficiency score for this student is equal to 0.6 (i.e., 3/5). We used the efficiency scores of the problem in which the hint was accessed as an outcome variable for RQ 2.1, and the efficiency scores on the next problem in which the hint was accessed as an outcome variable for RQ 2.2. For example, if a student requests a hint for problem *n*, the efficiency score on the next problem is the score for problem *n*+1.

Prior efficiency scores (control variable for RQ2)

In order to control students' prior performance in prediction models, we computed the prior efficiency score by taking the average of all efficiency scores on problems completed by the student prior to each observed problem (i.e., the problem in which the hint was accessed). For problem *n*, their prior efficiency score for that problem is equal to $\sum_{i=1}^{n-1}(efficiency\ score_i)\ /\ (n-1)$.

Data analysis

For RQ1, we first considered conducting an ANOVA to test the differences in hint reading time by action pair and hint type. However, the results of the Shapiro-Wilk normality test and Bartlett test of homogeneity of variances showed that the data violated the assumptions for ANOVA. Thus, we conducted Kruskal-Wallis tests followed by Dunn's post hoc tests. For RQ2, we conducted multiple regression analyses to examine the influence of hint reading time and hint type on in-game performance. All analyses were conducted at the problem level (i.e., one row for each student's problem attempt).

Results 1: Students' hint usage in a digital mathematics game

RQ1.1: Differences in hint reading time by the action pair

We first examined whether there were differences in students' hint reading time by their next actions after reading a hint (i.e., action pair). As shown in Table 3, there was a statistically significant difference between hint reading time by the action pair ($\chi^2 = 70.767$, df = 2, $p <. 001$). We then conducted Dunn's post hoc test to identify which specific pairs differed significantly. The results showed that all action pairs were significantly different from each other. Specifically, "hint-attempt" had a significantly higher mean rank than the other two action pairs, indicating that the students who made a mathematically valid action after requesting a hint tended to spend a longer time reading hints than those who made an invalid action (mistake) or reset after reading hints. "Hint-reset" had a significantly lower mean rank than the other two action pairs.

Table 3
Kruskal Wallis Test Results of the Hint Reading Time Mean Scores by the Action Pairs

Action pair	*n*	Mean Rank	df	χ^2	*p*
hint-attempt	6,757	3,379			
hint-reset	877	439	2	70.706	< .001
hint-mistake	1,380	691			

RQ1.2: Differences in hint reading time by the hint content type

We examined whether there were differences in students' hint reading time by the hint content type. Table 4 shows a statistically significant difference in hint reading time among the hint types ($\chi^2 = 6.391$, df = 2, $p =. 04$).

Table 4
Kruskal Wallis Test Results of the Hint Reading Time Mean Scores by the Hint Types

Hint type	*n*	Mean Rank	df	χ^2	*p*
direct	6,160	3,080			
intermediate	2,634	1,318	2	6.391	.04
indirect	220	111			

The results of Dunn's post hoc test revealed that "indirect hint" had a significantly lower mean rank of hint reading time than those of "direct hint" and "intermediate hint.", showing that the students tended to spend a shorter time reading indirect hints compared to direct or intermediate hints in the game. There was no statistically significant difference in hint reading time between "direct hint" and "intermediate hint."

Results 2: Hint reading time and in-game performance

RQ2.1: Hint reading time and problem solving efficiency

We conducted a multiple regression analysis to test whether the potential predictors were related to problem-level efficiency scores. "Intermediate hint" was selected as a reference group. As shown in Table 5, hint reading time significantly predicted the efficiency score of the problem ($B = 0.005$, $p = .038$), indicating that spending a longer time reading a hint led to higher efficiency scores on a single problem after controlling for the prior efficiency scores. Direct hints also significantly and positively predicted the efficiency score after controlling for the prior efficiency scores ($B = 0.04$, $p < .001$), while indirect hints negatively predicted the efficiency score ($B = −0.13$, $p < .001$). The model explained 2.8% of the variance in the efficiency scores ($R^2 = 0.028$, $F(4, 8989) = 63.78$, $p <.001$).

Table 5
A Result of the Regression Analysis Predicting the Efficiency Score

Variable	Model 1.1			
	B	95% CI	β	*p*
Intercept	0.559***	[0.484, 0.635]	0.083	< .001
Hint reading time	0.005*	[0.000, 0.010]	0.004	.038
Direct hint	0.040***	[0.031, 0.049]	0.018	< .001
Indirect hint	−0.130***	[−0.156, −0.103]	−0.020	< .001
Prior efficiency	0.267***	[0.186, 0.348]	0.013	< .001

****p < .001, **p < .01, *p < .05*

Next, we tested the interactions between hint reading time (HRT) and direct hint (Model 1.2 in Table 6) and hint reading time and indirect hint on the efficiency scores (Model 1.3 in Table 6). There were no significant interaction effects on the efficiency scores between hint reading time and direct hint ($B = −0.003$, $p = .597$) and between hint reading time and indirect hint ($B = −0.014$, $p = .356$).

Table 6
A Result of the Regression Analysis Predicting the Efficiency Score with the Interaction Effects

Variable	Model 1.2				Model 1.3			
	B	95% CI	β	*p*	*B*	95% CI	β	*p*
Intercept	0.556***	[0.480, 0.632]	0.085	<.001	0.559***	[0.484, 0.634]	0.835	< .001
HRT	0.007*	[-0.001, 0.016]	0.004	.105	0.006*	[0.001, 0.011]	0.004	.02
Direct	0.045***	[0.024, 0.067]	0.019	<.001	0.040***	[0.031, 0.049]	0.019	< .001
Indirect	-0.130***	[-0.156,-0.103]	-0.020	<.001	-0.106***	[-0.163, -0.049]	-0.020	< .001
Prior efficiency	0.266***	[0.185, 0.348]	0.013	<.001	0.266***	[0.185, 0.348]	0.013	< .001
HRT× direct	-0.003	[-0.014, 0.008]	-0.001	.597	-	-	-	-
HRT× indirect	-		-	-	-0.014	[-0.045, 0.016]	-0.002	.356

****p < .001, **p < .01, *p < .05*

RQ2.2: Hint reading time and the problem solving efficiency on the next problem

We tested whether the potential predictors were related to the efficiency scores on the next problem, by selecting "intermediate hint" as a reference group (Table 7). Similar to the prior model, hint reading time ($B = 0.005$, $p = .033$) and direct hint ($B = 0.045$, $p < .001$) positively predicted the efficiency score on the next problem after controlling for the prior efficiency scores. Unlike the prior model, indirect hints significantly and positively predicted the efficiency score on the next problem ($B = 0.118$, $p < .001$). The model explained 2.6% of the variance in the efficiency score on the next problem ($R^2 = 0.026$, $F(4, 8770) = 58.73$, $p < .001$).

Table 7
A Result of the Regression Analysis Predicting the Efficiency Score on the Next Problem

Variable	Model 2.1			
	B	95% CI	β	*p*
Intercept	0.443***	[0.369, 0.518]	0.872	< .001
Hint reading time	0.005*	[0.000, 0.010]	0.004	.033
Direct hint	0.040***	[0.031, 0.049]	0.019	< .001
Indirect hint	0.118***	[0.092, 0.143]	0.018	< .001
Prior efficiency	0.429***	[0.349, 0.510]	0.021	< .001

****p* < .001, ***p* < .01, **p* < .05

Next, we tested the interaction effects between hint reading time (HRT) and direct hint (Model 2.2 in Table 8) and between hint reading time and indirect hint on the efficiency scores on the next problem (Model 2.3 in Table 8). The interaction effect between direct hint and hint reading time on the next efficiency scores was significant ($B = 0.025$, $p < .001$), indicating that the effect of the hint reading time on the next efficiency scores was greater for direct hints than intermediate hints. In other words, students with a longer hint reading time tended to have higher efficiency scores on the next problem when the direct hint was provided compared to when the intermediate hint was provided. There was no significant interaction effect between hint reading time and indirect hint on the next efficiency scores ($B = 0.000$, $p = .983$).

Table 8
A Result of the Regression Analysis Predicting Efficiency Score on the Next Problem with the Interaction Effects

Variable	Model 2.2				Model 2.3			
	B	95% CI	β	*p*	*B*	95% CI	β	*p*
Intercept	0.470***	[0.395, 0.546]	0.872	< .001	0.443***	[0.369, 0.518]	0.872	< .001
HRT	-0.012**	[-0.021, -0.003]	0.004	.008	0.005*	[0.000, 0.010]	0.004	.035
Direct	-0.004	[-0.024, 0.017]	0.018	.727	0.040***	[0.031, 0.049]	0.019	< .001
Indirect	0.116***	[0.090, 0.141]	0.018	< .001	0.118***	[0.062, 0.174]	0.018	< .001
Prior efficiency	0.433***	[0.353, 0.514]	0.021	< .001	0.429***	[0.349, 0.510]	0.020	< .001
HRT× direct	0.025***	[0.014, 0.036]	0.009	< .001	-	-	-	-
HRT× indirect	-		-	-	0.000	[−0.030, 0.030]	0.000	.983

****p* < .001, ***p* < .01, **p* < .05

Discussion

While help-seeking behavior is an important component of the learning process and cognitive development, research has shown inconsistent findings on the effect of help in learning technologies; most have found negative effects of requesting help through hints on learning outcomes (Iannacchione et al., 2023; McLaren et al., 2022). This study examined how 7th-grade students used hints in a digital mathematics game, FH2T, with a focus on the productivity of hint usage using response time decomposition and hint content type.

We found that mathematically valid attempts after reading hints tended to be preceded by longer hint reading times than when students made mistakes or reset the problem, which was aligned with Gurung et al.'s (2021) finding. In addition, students' hint reading time was positively related to in-game performance, which was consistent with the results of other studies (Gurung et al., 2021; Long & Aleven, 2013; Shih et al., 2011). This finding also aligns with our prior work showing that longer pause time before problem solving is related to more efficient use of strategies (Chan et al., 2023). Further, our findings support the idea of Aleven et al. (2016) that the negative impacts of help-seeking on students' performance might be due to their maladaptive help-seeking behaviors, not the use of hints. Thus, one implication of this finding is that prompting students to pause and think about the hint presented may support more efficient problem solving in mathematics.

Regarding hint type, in accordance with O'Rourke et al.'s (2014) findings, the students had a longer hint reading time on direct hints than on indirect hints. In addition, the use of direct hints was positively related to the efficiency scores on both the problem in which the hint was accessed and the following problem. The most interesting finding was the differential effects of indirect hints on in-game performance. While indirect hints were negatively related to the efficiency scores on the problem in which the hint was accessed, they were positively related to the efficiency scores on the next problem. Further, providing indirect hints was more predictive of the efficiency scores on the next problem than direct hints. These findings suggest that reading indirect hints may provide a productive struggle for students (Warshauer et al., 2015), where the students' struggles to restructure their knowledge toward new understanding may lead to stronger problem solving efficiency later.

Several limitations of the study need to be considered. First, the current study examined the effect of hint reading time and type on in-game performance using problem-level data. Further research is needed to assess the effect of students' hint reading time and type on learning outcomes (e.g., posttest scores) using student-level data. Another limitation of the study is an imbalance in the distribution of three hint types. A further controlled study is suggested to confirm the relationship between the hint type and performance.

References

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 26*, 205-223.

Arroyo, I., Beck, J. E., Woolf, B. P., Beal, C. R., & Schultz, K. (2000). Macroadapting Animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In *Intelligent Tutoring Systems: 5th International Conference, ITS 2000 Montréal, Canada, June 19–23, 2000 Proceedings 5* (pp. 574–583). Springer Berlin Heidelberg.

Chan, J. Y. C., Lee, J. E., Mason, C. A., Sawrey, K., & Ottmar, E. (2022). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology*, *114*(1), 56–71.

Decker-Woodrow, L. E., Mason, C. A., Lee, J. E., Chan, J. Y. C., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA open*, *9*.

Gurung, A., Botelho, A. F., & Heffernan, N. T. (2021, April). Examining student effort on help through response time decomposition. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 292-301). ACM.

Iannacchione, A., Ottmar, E., Ngo, V., Mason, C. A., Chan, J. Y. C., Smith, H., ... & Shaw, S. T. (2023). Examining relations between math anxiety, prior knowledge, hint usage, and performance of math equivalence in two different online learning contexts. *Instructional Science*, *51*(2), 285–307.

Kehrer, P., Kelly, K., & Heffernan, N. (2013). Does immediate feedback while doing homework improve learning? In Proceedings of *the Twenty-Sixth International Artificial Intelligence Research Society Conference* (pp. 542–545). AAAI Press.

Long, Y., & Aleven, V. (2013). Skill Diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education, AIED 2013*, (pp 249–258), New York: Springer.

McLaren, B. M., Richey, J. E., Nguyen, H., & Hou, X. (2022). How instructional context can impact learning with educational technology: Lessons from a study with a digital learning game. *Computers & Education, 178,* 104366.

Muir, M., & Conati, C. (2012). An analysis of attention to student–adaptive hints in an educational game. In *Intelligent Tutoring Systems: 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18, 2012. Proceedings 11* (pp. 112–122). Springer Berlin Heidelberg.

O'Rourke, E., Ballweber, C., & Popovií, Z. (2014). Hint systems may negatively impact performance in educational games. In *Proceedings of the first ACM conference on Learning@Scale conference* (pp. 51–60).

Razzaq, L., & Heffernan, N. T. (2010). Hints: Is it better to give or wait to be asked? In *Intelligent Tutoring Systems: 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part I 10* (pp. 349-358). Springer Berlin Heidelberg.

Shih, B., Koedinger, K. R., & Scheines, R. (2011). A response time model for bottom-out hints as worked examples. *Handbook of Educational Data Mining*, 201–212. ACM.

Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, *18*, 375-400.

Acknowledgment

This work was supported by National Science Foundation: [Grant Number 2142984] and Institute of Education Sciences: [Grant Number R305A180401].