Characterizing and Influencing Students' Tendency to Write Self-explanations in Online Homework

Yuya Asano
yuya.asano@mail.utoronto.ca
University of Toronto
Toronto, Ontario, Canada

Jaemarie Solyst
jsolyst@cs.toronto.edu
University of Toronto
Toronto, Ontario, Canada

Joseph Jay Williams
williams@cs.toronto.edu
University of Toronto
Toronto, Ontario, Canada

ABSTRACT

In the context of online programming homework for a university course, we explore the extent to which learners engage with optional prompts to *self*-explain answers they choose for problems. Such prompts are known to benefit learning in laboratory and classroom settings [4], but there are less data about the extent to which students engage with them when they are optional additions to online homework. We report data from a deployment of self-explanation prompts in online programming homework, providing insight into how the frequency of writing explanations is correlated with different variables, such as how early students start homework, whether they got a problem correct, and how proficient they are in the language of instruction. We also report suggestive results from a randomized experiment comparing several methods for increasing the rate at which people write explanations, such as including more than one kind of prompt. These findings provide insight into promising dimensions to explore in understanding how real students may engage with prompts to explain answers.

CCS CONCEPTS

• **Applied computing → E-learning**.

KEYWORDS

self-explanation, prompts, engagement, online homework, randomized experiment

ACM Reference Format:
Yuya Asano, Jaemarie Solyst, and Joseph Jay Williams. 2020. Characterizing and Influencing Students' Tendency to Write Self-explanations in Online Homework. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20), March 23–27, 2020, Frankfurt, Germany.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3375462.3375511

1 INTRODUCTION & RELATED WORK

Prompting students to explain the meaning of what they are learning in their own words, or self-explanation, has been shown to be beneficial for learning in both laboratory [16] and classroom settings [5, 13]. For example, prompting people to explain the answer to a problem can help them identify underlying principles [17]. However, there is far less research investigating learners' behavior when prompts to explain are embedded in real-world digital educational environments, such as online homework.

In contrast to traditional in-person environments, students in online courses often lack motivation, which results in undesirable learning behaviours. For example, they tend to fail to finish watching videos [7] and have lower participation rates in discussion forums [8], compared to students enrolled in in-person courses. Online courses can also have much higher attrition rates than traditional courses [8]. Furthermore, some students even "game" educational systems to get the correct answers without thoughtful effort, e.g. by taking advantage of the automatic feedback they are provided with or continually guessing at multiple-choice questions [3]. Considering these obstacles in online learning environments, it would be valuable to understand the extent to which people would engage with optional prompts to explain because they learn differently in online environments than in-person.

Past work on prompting and supporting self-explanations in digital environments often has extensive support built in to engage students in explaining. Studies in intelligent tutoring systems often require students to write explanations before they can move between steps [12]. Or, citing a concern with students not providing free responses, systems do not fully elicit explanations but give students multiple-choice options to choose an explanation for a concept [1].

Other approaches to engage students in self-explanation use natural language processing (NLP) techniques to coach students in constructing explanations [2]. These can engage students in providing explanations through checking the nature and quality of these explanations, giving feedback, and guiding students to generate self-explanations. Such data-intensive approaches have also been applied to far more complex kinds of reflective writing, even at the level of analytics for writing essays [10].

Despite the progress in this past work, there is surprisingly little work looking at students' engagement with open-ended prompts to explain, in the *absence* of requirements to provide explanations or sophisticated NLP techniques for ensuring students write text. This is important because adding optional open-ended prompts to explain is a pedagogical technique that could be easily applied by any instructor even without having a sophisticated custom-built system or needing to incorporate grading and checking of explanations into a marking scheme. Understanding how and whether students would engage with such prompts could be of broad applicability.

For example, what kinds of variables about student state correlate with students' tendency to respond to self-explanation prompts? How might instructors aim to increase the rate at which students respond to self-explanation prompts?

LAK '20, March 23–27, 2020, Frankfurt, Germany

ACM ISBN 978-1-4503-7712-6/20/03…$15.00
https://doi.org/10.1145/3375462.3375511

Work on retrieval-practice comes the closest, in comparing the impact of asking people to write summaries of videos with that of providing summaries [15]. In contrast, we investigate students' engagement to open-ended prompts to explain the answers they chose to online homework problems in a for-credit university course. We explore the complexities of encouraging students to engage in self-explanation in an online homework platform for introductory programming. Specifically, we investigate the following research questions:

(1) When do students tend to respond to self-explanation prompts?
(2) How might we encourage students to respond to self-explanation prompts?

To answer these questions, we embedded prompts asking novice programming students to explain their answers for online homework problems. This paper reports on how the proportion of students writing self-explanations covaries with several variables about student state. We found that students were less likely to participate in self-explanation if they spent less time on solving homework questions, were not confident in their proficiency in the language of instruction, or scored poorly on homework questions. Additionally, students who performed well on the intervention questions and prior related problems generated more detailed explanations. We also employed various prompting strategies to address our second research question. Although many of these prompting strategies did not result in statistically significant differences, the data provide insights into which findings are promising. For example, giving context for the self-explanation prompt might have discouraged students from writing and lowered the specificity of explanations, while the specificity of explanations might have been increased by a prompt with a metacognitive element of identifying points of confusion [4].

2 METHODS

In this section, we summarize the details of our experiments, including participants, the "contextual" variables to capture characteristics of students or their state, and the experimental factors varied to encourage self-explanation.

2.1 Deployment in Online Homework System

We embedded self-explanation prompts, which we refer to as interventions, in online homework problems in an introductory programming course at a large research university in North America. The digital homework system consists of pre-lecture and post-lecture activities, in which students are assigned tasks every week before and after the lectures. Both activities have multiple-choice and open-ended Python coding questions (309 and 61, respectively), but only pre-lecture activities have instructional videos (133 in total). 46% of the problems were for course credit and accounted for 14% of the students' final grade. The only feedback the system gives to the students is the correctness of their answer and their score. We deployed the interventions on one not-for-credit multiple-choice problem (P1) and one for-credit multiple-choice problem (P2), which were in different pre-lecture activities in the same week (Figure 1).

There were 205 students enrolled in the course. At the beginning of the course, we sent out a survey to ask students about their backgrounds and for consent to use their data in this study. We removed students from our analysis who did not consent to be a part of our study and used data from 95 students on P1 and 126 students on P2.

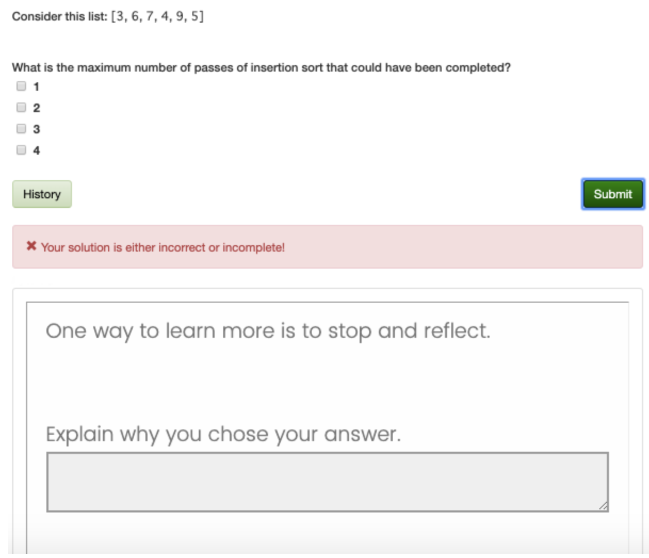

Figure 1: The screenshot of P1 and our intervention. Students saw an intervention every time they hit "Submit" button and saw the same intervention even when they submitted multiple times on the same question. Prompts were shown through Qualtrics survey [14].

2.2 Contextual Variables about Student State

To understand covariation between tendency to write explanations and student state, we used *Contextual Variables* to categorize student state.

Logged data from the homework system were used to obtain the following contextual variables: Time Solving Problems, Time Until Deadline, whether they Solved Problem Correctly at the first attempt, and Accuracy on Related Prior Problems (Table 1). The time spent on solving problems was estimated using the time of the previous submission since both P1 and P2 immediately followed other problems. Related prior problems were already encoded by the instructors who had designed and used the system over more than five years. Accuracy on Related Prior Problems was calculated by averaging the score of each attempt for each prior problem and then averaging those averages across the prior problems.

A measure of proficiency in the language of instruction (English) was obtained by asking if the students *"find it challenging to read and listen in English. For example, in reading text and listening to videos in [the name of the online homework system]"* in the survey. They answered on a 9-point Likert scale (from 9: strongly disagree to 1: strongly agree).

Measures for each student, except for their scores on P1 and P2, were compared with the corresponding medians to categorize students into two groups for each contextual variable in Table 1. The contextual variable Solved Problem Correctly was determined by whether students got a full score in P1 or P2.

2.3 Experimental Factors

When we showed self-explanation prompts to students, we also randomly assigned students to receive differently-worded prompts, additional messages, and text entry boxes of different sizes as shown in Figure 1. The Context for Prompts told students that *"one way to learn more is to stop and reflect"*. It provided them with some context for why the prompt was being included before they read. It incorporated some strategies in the ARCS model, which is aimed at increasing and maintaining the motivation of learners [9]. The prompt was given to 59 and 78 students in P1 and P2, respectively. We showed an Elaboration on Prompt, saying *"for example, what are the key concepts you used? Are there concepts you're confused about?"*, to 50 students in P1 and 65 students in P2. It contained a metacognitive element used to elicit self-explanation in a prior study [4]. All students saw the default justification-based prompt, *"explain why you chose your answer"* [6]. We also randomized whether students (50 in P1 and 62 in P2) were asked to write as if they were explaining to a friend. The Explain to Friend phrase, *"as if you were talking to a friend in class"*, was added to the end of the default prompt. In addition, we varied the Size of Text Box in P2. In P1, all students received a multi-line box, but in P2, 67 students received a single-line box while the remaining (59) received a multi-line box.

The unequal number of participants per condition occurred not because of differential dropout, but because we randomized the conditions based on their internal identification code in the homework system. In addition to the students who did not consent to their data being used for research (mentioned in section 2.1), we dropped some of the students from our analysis because they moved onto the next problem in less than one second after completing it and were unlikely to have seen the prompts, but this dropout is not the cause of the uneven distribution of the participants in each condition.

2.4 Specificity of Students' Explanations

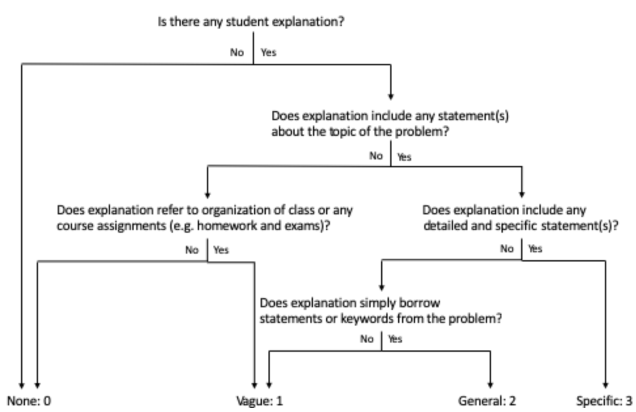

Figure 2: The flowchart of our coding schema for Explanation Specificity. This was adapted from Menekse et al. [11]

We coded the student-generated explanations to assess their specificity. We adapted the coding schema suggested by Menekse et al. [11]. Explanations were assigned scores ranging from 0 to 3. Figure 2 illustrates our coding schema.

A score of 0 means that the student left the text box blank (i.e. no explanation was written) or that their explanation did not engage with the core concepts of the problem. Explanations that received a score of 0 includes *"because it's correct"* and *"this is what the [the name of the homework system] video taught me"*. Since the students who did not write an explanation received a score of 0, our results were skewed toward 0.

Explanations received a score of 1 if they were not specific to the current problems or only took words or phrases directly from the problem statement. For example, in P2 a student wrote *"because bubble sort is an algorithm that supposedly 'sorts' items from smallest to largest. Bubble sort is used once, so the last item on the list would be the largest"*. Although this looks specific to the topic of the problem, which is bubble sort, this explanation discusses the general fact that any sorting algorithm eventually arranges the elements of its input in ascending order.

A score of 2 was assigned to the explanations that discussed the topic of the problem but were not detailed. One example from P1 is *"the fourth item in the list is not sorted"*. This student seemed to refer to the property of insertion sort, in that it sorts the first n elements relatively after the n^{th} pass, but did not say explicitly how this statement related to the answer of the problem.

A score of 3 was given to explanations that were detailed and specific to the problem. In P2, a student said *"... for each two items in a list, it compares which is largest and ensures the largest is to the right of the two, going through the whole list the largest item in the list should be at the far right"*. This explanation clearly shows how the student translated their understanding of bubble sort to their answer.

3 RESULTS

In this section, we first present the correlations between characteristics of students and the frequency of writing self-explanations (at all) and the tendency to write more specific explanations. We then present the effects of the experimental factors on the frequency and specificity of self-explanations.

3.1 Relationship between Student State and Self-Explanation Frequency and Specificity

The results indicate that only a minority of students wrote responses to the self-explanation prompts. Only 18 students wrote explanations in both problems, and 27.7% of students wrote an explanation at least once (in either problem). Over 96% of explanations were generated directly following the student's first submission for a given problem. Therefore, we focus on students' first submissions in the following analysis.

We found that students were more likely to write responses to self-explanation prompts when they started homework early, spent more time solving the homework questions, and had a good performance in related prior and target questions (Table 1). For example, of students who were above the median in Time Solving Problem, the proportion writing explanations was 38% (P1) and 32% (P2). Of students who were below the median in Time Solving Problems, the proportion was 7% (P1) and 17% (P2) (two proportion z-test, $z = 3.57, 1.94$, $p < 0.01, 0.05$, two-tailed).

Table 1: The proportion of students who wrote self-explanations and their average specificity (with and without blank explanations), by the contextual variables representing student state. Numbers in parentheses represent standard error of the mean. The symbol * indicates a significant statistical difference at $p < 0.05$ level, between the two rows within the same context. The difference between proportions was tested using a two-proportion z-test, and the difference in specificity was tested using a Mann-Whitney U-test. Note that blank explanations received a specificity score of 0.

Student Contextual Variables		Proportion of Students who Wrote Explanation P1	P2	Explanation Specificity P1	P2	Explanation Specificity (blanks removed) P1	P2
Time Solving Problem	<median	0.07* (0.03)	0.17* (0.05)	0.12 (0.07)	0.22 (0.08)	1.67 (0.88)	1.3 (0.37)
	>median	0.38* (0.06)	0.32* (0.05)	0.56 (0.14)	0.44 (0.11)	1.45 (0.30)	1.36 (0.28)
Time Until Deadline	<median	0.15* (0.05)	0.20 (0.05)	0.28 (0.10)	0.33 (0.11)	1.86 (0.51)	1.67 (0.41)
	>median	0.33* (0.06)	0.30 (0.05)	0.43 (0.13)	0.35 (0.09)	1.31 (0.34)	1.15 (0.25)
Self-rated Proficiency in Language of Instruction (English)	<median	0.14* (0.04)	0.13* (0.04)	0.14 (0.07)	0.21 (0.08)	1.0 (0.44)	1.44 (0.38)
	>median	0.36* (0.06)	0.35* (0.06)	0.60 (0.15)	0.46 (0.12)	1.69 (0.35)	1.3 (0.28)
Solved Problem Correctly	No	0.14* (0.04)	0.16* (0.04)	0.11* (0.04)	0.16* (0.06)	0.75 (0.31)	1.0 (0.33)
	Yes	0.38* (0.08)	0.35* (0.06)	0.72* (0.19)	0.53* (0.13)	1.87 (0.36)	1.5 (0.28)
Accuracy on Related Prior Problems	<median	0.15* (0.04)	0.22 (0.05)	0.15* (0.06)	0.16* (0.06)	1.0 (0.42)	0.73* (0.25)
	>median	0.37* (0.07)	0.29 (0.05)	0.63* (0.17)	0.54* (0.13)	1.73 (0.36)	1.88* (0.31)

Students who started the homework earlier—above the median of Time Until Deadline—engaged in self-explanation more often than those who did not in P1, but that difference did not approach significance in P2 (two proportion z-test, $z = 1.98, 1.33$, $p = 0.05, 0.18$, two-tailed). Furthermore, 36% (P1) and 35% (P2) of students with higher Self-rated Proficiency in Language of Instruction (English) wrote self-explanations in P1 and P2 respectively, while only 14% (P1) and 13% (P2) of students with lower proficiency did so (two proportion z-test, $z = 2.45, 2.66$, $p = 0.01, 0.01$, two-tailed).

We also found that a student's understanding of the current and previous course material affected their tendency to write explanations (Table 1). For instance, students were more likely to engage with self-explanation prompts if they Solved Problem Correctly (two proportion z-test, $z = 2.71, 2.56$, $p = 0.01, 0.01$, two-tailed). Additionally, students with higher Accuracy on Related Prior Problems tended to participate in self-explanation in P1 (two proportion z-test, $z = 2.45$, $p = 0.01$, two-tailed). This difference was not statistically significant in P2 (two proportion z-test, $z = 0.83$, $p = 0.41$, two-tailed).

Explanation Specificity was influenced by student performance, in both variable Solved Problem Correctly and its Accuracy on Related Prior Problems (Table 1). Without dropping blank explanations, the average specificity for students who Solved Problem Correctly at the first try was 0.72 (P1) and 0.53 (P2), but, for those who did not, it was 0.11 (P1) and 0.16 (P2) (Mann-Whitney U-test, $U = 1312.0, 2264.5$, $p = 0.01, 0.03$). Students who had higher Accuracy on Related Prior Problems created more specific explanations ($M = 0.63, 0.54$), compared to those who had lower accuracy ($M = 0.15, 0.16$). These differences were statistically significant (Mann-Whitney U-test, $U = 1304.0, 2231.0$, $p = 0.02, 0.05$). However, after removing students who did not write any explanations, most differences were not statistically significant except for the Accuracy on Related Prior Problems in P2 (Mann-Whitney U-test, $U = 191.0$, $p = 0.01$).

3.2 Effects of Different Prompts and Messages

Counterintuitively, the Context for Prompt had marginally significant negative effects on the proportion of students who wrote explanations but did not change the Explanation Specificity (Table 2). Only 14% and 17% of students wrote explanations when they received the Context for Prompt, in P1 and P2, respectively, while 31% of students did so when they did *not* get it,in both P1 and P2, (two proportion z-test, $z = -1.83, -1.77$, $p = 0.07, 0.08$, two-tailed). The impact on Explanation Specificity was not statistically significant; the results of Mann-Whitney U-test were $U = 961.5, 1722.0$ and $p = 0.21, 0.24$ with blank explanations and $U = 53.5, 111.0$ and $p = 0.53, 0.51$ without them.

The Elaboration on Prompt did not change the Proportion of Students who Wrote Explanations as shown in Table 2 (two proportion z-test, $z = 1.49, 1.03$, $p = 0.14, 0.30$, two-tailed). Also, there was no significant effect of the Elaboration on Prompt on the Explanation Specificity either before (Mann-Whitney U-test, $U = 1237.5, 2130.5$, $p = 0.18, 0.26$) or after (Mann-Whitney U-test, $U = 68.5, 140.0$, $p = 0.74, 0.59$) removing blank explanations from the analysis.

Asking students to Explain to Friend in class had no effect on the Proportion of Students who Wrote Explanation (two proportion z-test, $z = -1.39, -0.51$, $p = 0.16, 0.61$, two-tailed), or on Explanation Specificity (non-blank and blank explanations) (Mann-Whitney U-test, $U = 1101.0, 1970.5$, $p = 0.78, 0.92$), or Explanation Specificity (only non-blank explanations) (Mann-Whitney U-test, $U = 79.0, 134.0$, $p = 0.21, 0.81$).

The Size of Text Box (single-line vs multi-line) factor in P2 (all P1 problems had a multi-line text box) had no significant effect on the Proportion of Students who Wrote Explanations (two proportion z-test, $z = -0.42$, $p = 0.68$, two-tailed). However, in examining the Explanation Specificity (blanks removed) for people who wrote explanations, there was a significant reduction in Explanation Specificity for the single-line text box (Mann-Whitney U-test, $U = 194.5$, $p = 0.01$).

Table 2: The proportion of students who wrote self-explanations and their average specificity (with and without blank explanations), by the experimental factors. Numbers in parentheses represent standard error of the mean. The symbol * indicates a significant statistical difference at $p < 0.05$ level, between the two rows within the same factor. The difference between proportions was tested using a two-proportion z-test, and the difference in specificity was tested using a Mann-Whitney U-test. Note that blank explanations received a specificity score of 0.

		Proportion of Students who Wrote Explanation		Explanation Specificity		Explanation Specificity (blanks removed)	
Experimental factors		P1	P2	P1	P2	P1	P2
Context for Prompt	Not received	0.31 (0.05)	0.31 (0.05)	0.42 (0.11)	0.38 (0.09)	1.39 (0.30)	1.25 (0.24)
	Received	0.14 (0.05)	0.17 (0.05)	0.25 (0.12)	0.27 (0.11)	1.8 (0.73)	1.62 (0.53)
Elaboration on Prompt	Not received	0.18 (0.05)	0.22 (0.05)	0.24 (0.10)	0.26 (0.09)	1.33 (0.47)	1.21 (0.33)
	Received	0.31 (0.06)	0.30 (0.05)	0.49 (0.13)	0.43 (0.11)	1.57 (0.36)	1.44 (0.30)
Explain to Friend	Not received	0.30 (0.06)	0.27 (0.05)	0.36 (0.11)	0.35 (0.10)	1.2 (0.34)	1.29 (0.32)
	Received	0.18 (0.05)	0.23 (0.05)	0.36 (0.12)	0.33 (0.10)	2.0 (0.46)	1.4 (0.32)
Size of Text Box	Single-line	N/A	0.27 (0.05)	N/A	0.2 (0.07)	N/A	0.75* (0.27)
	Multi-line	N/A	0.24 (0.05)	N/A	0.46 (0.12)	N/A	1.94* (0.30)

4 DISCUSSION, LIMITATIONS & FUTURE WORK

The current work aimed at a better understanding of how students in a real course would respond to optional prompts to explain their answers when solving multiple-choice questions in online programming homework. We found that over 25% of students wrote explanations when prompted, even though no marks were awarded. However, student interviews and surveys suggested that even when students do not type responses, some of them are reflecting on the questions without writing responses because they find typing laborious. What students type is useful evidence but likely underestimates the depth and quality of the reflection they engage in.

Students were more likely to write a self-explanation when they had spent more time solving the question, started more in advance of the deadline, rated themselves as proficient in the language of instruction, got the answer right at the first attempt, and did well on similar previous problems. While in hindsight these might seem to be intuitive results, we believe that empirical evidence for these patterns in a real-world university course is useful. Firstly, there has been relatively little work reporting on the levels of engagement with writing self-explanations, in contexts where prompts are optional additions to online programming homework. This means there are little data for instructors and researchers to understand how frequently people engage in writing such explanations and how writing such explanations is correlated with our reported contextual variables about student state.

Secondly, there are reasons to predict we might *not* have found these correlations in our specific setting. One might also predict there would be no correlation or the opposite relationship, in different settings. For example, it could have been the case that students who spend more time on a question or perform better on related questions might not have been any more likely to write explanations because writing an explanation could be related to factors independent of diligence, such as personal preferences for articulating thoughts verbally through typing. Students who perform better might sometimes be *less* likely to write explanations than those who don't perform well because they see the explanation as being obvious. It is also noteworthy that students were more likely to write an explanation when they got an answer correct than when they got it incorrect, on average. However, several students reported in an end of course survey just the opposite—they would *prefer* to write explanations and reflect when the answer was *not* obvious or they were confused about what the right answer was.

Even the intuitive finding that learners who rated their proficiency in the language of instruction (English) lower were 22% less likely to write an explanation is valuable because it provides compelling quantitative evidence for supporting this population. It raises important questions about how to encourage people who report (and may be) less proficient in the language of instruction to engage in self-explanation. One potential solution might be allowing them to reflect in a way that does not involve laborious typing, such as fill-in-the-blank or multiple-choice explanations [1].

4.0.1 Experimental Results on Promoting Self-Explanation. There were some surprising trends in the effects of our experiments on students' responses to self-explanation prompts. For example, the Context for Prompt factor varied whether or not we gave students context for why the self-explanation prompt was included, by including the message *"one way to learn more is to stop and reflect"* above the prompt. This was included because feedback from the instructor and a few students suggested it was jarring or confusing to suddenly see the prompt *"explain why you chose your answer"* below a problem with no explication of why this was added to a problem. However, the data suggest that including this message might *reduce* the proportion of students who engage in self-explanation. This points to an interesting direction for future work, to more directly understand this phenomenon and the contexts in which it occurs. One possibility is that for some students the cost of scrolling down this message to read the prompt outweighs the benefit of having the system explain why a self-explanation prompt has been added. Another possibility is that students were frustrated by the statement of an overly obvious point, which points the way to better ways to design such messages.

There were also interesting trends that did not approach statistical significance. Including an elaboration on the prompt, saying that *"for example, what are the key concepts you used? Are there concepts you're confused about?"*, might increase the chances of writing an explanation. Meanwhile, being asked to explain to a friend might reduce the chances of writing an explanation. The current data indicate both are interesting variables to explore in the future.

Finally, we anticipated that reducing the more imposing multi-line text box to a single-line text box might increase how many people would write explanations, but there was no evidence for this. However, it did significantly reduce the specificity of explanations that people did write. This suggests that trying to lower the burden of explanation with a shorter box may be outweighed by lower quality explanations, at least in this context.

4.0.2 Limitations. One limitation of the correlations we report is that more research is needed to establish these as causal relationships and to use these promising patterns to spell out implications for instructional design. Another limitation is that more work needs to explore how the rate at which people write explanations changes over time, as such prompts become ubiquitous in a course and students experience them repeatedly. Will that make students less likely to write as they become habituated? Or will they be more likely to write since they will be in the habit and become better at writing explanations?

Our measure of explanation specificity is also limited as it is a fairly simple one, which was chosen because the optional explanations people wrote were fairly sparse. Future work can explore the value of using more sophisticated NLP and other approaches to measure the quality of explanations. Moreover, we recognize that the question of what makes an explanation high-quality is a very complex one and do not expect our current measures to capture all of that complexity.

4.0.3 Future Work & Conclusions. We believe the current results provide an insight into the engagement of real-world students with optional self-explanation prompts, highlighting several promising directions for future research. Future learning analytics researchers can expand on the data we report about when (or which) students write explanations. They can probe whether these point to true causal relationships or explore additional variables. In particular, it will be valuable to more precisely understand why some students are more likely to write explanations in certain circumstances. Moreover, how do the contexts in which students are more likely to write explanations align with the contexts when they are most likely to learn from writing explanations?

We also reported data from experimental variations of self-explanation prompts (providing context, explaining to a friend, elaborating on prompts, and varying the size of the text box). The results provided a combination of suggestive evidence about counter-intuitive effects, as well as the lack of evidence for effects one might have anticipated (non-significant effects). This can help instructors and researchers to make decisions about what variables are useful for further exploration and/or use our operationalization of these variables to design potentially more impactful versions. As is often the case with experiments, these results can be the first step in a series of data about potential effects across different contexts, to move towards a broader account of which instructional interventions for self-explanation are effective, under which circumstances.

ACKNOWLEDGMENTS

The authors gratefully acknowledge grants from the Office of Naval Research (#N00014-18-1-2755) and the National Science Foundation (1724889). We'd also like to thank Sam Maldonado and Lisa Zhang for their contributions to implementing our online interventions, as well as instructors that gave us support and students who participated in these studies.

REFERENCES

[1] Vincent A.W.M.M. Aleven and Kenneth R. Koedinger. 2002. An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science* 26, 2 (2002), 147 – 179. https://doi.org/10.1016/S0364-0213(02)00061-7

[2] Laura K. Allen, Cecile Perret, Aaron Likens, and Danielle S. McNamara. 2017. What'd You Say Again?: Recurrence Quantification Analysis As a Method for Analyzing the Dynamics of Discourse in a Reading Strategy Tutor. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. ACM, New York, NY, USA, 373–382. https://doi.org/10.1145/3027385.3027445

[3] Ryan Shaun Baker, Albert T. Corbett, and Kenneth R. Koedinger. 2004. Detecting Student Misuse of Intelligent Tutoring Systems. In *Intelligent Tutoring Systems*, James C. Lester, Rosa Maria Vicari, and Fábio Paraguaçu (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 531–540.

[4] Michelene T.H. Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian Lavancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18, 3 (1994), 439 – 477. https://doi.org/10.1016/0364-0213(94)90016-7

[5] Jennifer L. Chiu and Michelene T. H. Chi. 2014. *Supporting Self-Explanation in the Classroom.* American Psychological Association, 271 – 286. https://teachpsych.org/Resources/Documents/ebooks/asle2014.pdf#page=97

[6] Cristina Conati and Kurt Vanlehn. 2001. Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education,* 11, (01 2001). https://doi.org/10.1007/s40593-015-0074-8

[7] Brent J. Evans, Rachel B. Baker, and Thomas S. Dee. 2016. Persistence Patterns in Massive Open Online Courses (MOOCs). *The Journal of Higher Education* 87, 2 (2016), 206–242. https://doi.org/10.1080/00221546.2016.11777400

[8] Khe Foon Hew and Wing Sum Cheung. 2014. Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review* 12 (2014), 45 – 58. https://doi.org/10.1016/j.edurev.2014.05.001

[9] John M. Keller. 1987. Development and use of the ARCS model of instructional design. *Journal of instructional development* 10, 3 (01 Sep 1987), 2. https://doi.org/10.1007/BF02905780

[10] Vitomir Kovanović, Srećko Joksimović, Negin Mirriahi, Ellen Blaine, Dragan Gašević, George Siemens, and Shane Dawson. 2018. Understand Students' Self-reflections Through Learning Analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. ACM, New York, NY, USA, 389–398. https://doi.org/10.1145/3170358.3170374

[11] Muhsin Menekse, Glenda Stump, Stephen Krause, and Michelene Chi. 2011. The effectiveness of students' daily reflections on learning in engineering context. *ASEE Annual Conference and Exposition, Conference Proceedings* (2011).

[12] Timothy J. Nokes, Robert G. M. Hausmann, Kurt VanLehn, and Sophia Gershman. 2011. Testing the instructional fit hypothesis: the case of self-explanation prompts. *Instructional Science* 39, 5 (01 Sep 2011), 645–666. https://doi.org/10.1007/s11251-010-9151-4

[13] Tenaha O'Reilly, Roger S. Taylor, and Danielle S. McNamara. 2006. Classroom Based Reading Strategy Training: Self-Explanation vs. a Reading Control. *Annual Meeting of the Cognitive Science Society* 28. https://pdfs.semanticscholar.org/39ea/31c4e59c31c90e8ea356fd3ff5f4432cea5e.pdf

[14] Qualtrics. 2005–2019. Qualtrics software. https://www.qualtrics.com/.

[15] Tim van der Zee, Dan Davis, Nadira Saab, Bas Giesbers, Jasper Ginn, Frans van der Sluis, Fred Paas, and Wilfried Admiraal. 2018. Evaluating Retrieval Practice in a MOOC: How Writing and Reading Summaries of Videos Affects Student Learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. ACM, New York, NY, USA, 216–225. https://doi.org/10.1145/3170358.3170382

[16] Joseph J. Williams and Tania Lombrozo. 2010. The Role of Explanation in Discovery and Generalization: Evidence From Category Learning. *Cognitive Science* 34, 5 (2010), 776–806. https://doi.org/10.1111/j.1551-6709.2010.01113.x

[17] Joseph Jay Williams, Tania Lombrozo, Anne Hsu, Bernd Huber, and Juho Kim. 2016. Revising Learner Misconceptions Without Feedback: Prompting for Reflection on Anomalies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 470–474. https://doi.org/10.1145/2858036.2858361