Exploring Additional Personalized Support While Attempting Exercise Problems in Online Learning Platforms

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

Madhurima Dutta
University of Toronto
Toronto, Canada
madhurima.dutta@mail.utoronto.ca

Trisha Thakur
University of Toronto
Toronto, Canada
trisha.thakur@mail.utoronto.ca

Jaemarie Solyst
Carnegie Mellon University
Pittsburgh, USA
jsolyst@andrew.cmu.edu

Stephanie Cristea
University of Toronto
Toronto, Canada
stephanie.cristea@mail.utoronto.ca

Helena Jovic
University of Toronto
Toronto, Canada
helena.jovic@mail.utoronto.ca

Andrew Petersen
University of Toronto Mississauga
Mississauga, Canada
andrew.petersen@utoronto.ca

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

ABSTRACT

In online asynchronous learning environments, students are assigned exercises, but it is not clear how to incorporate the kinds of actions an in-person tutor might take such as explaining, providing more practice, prompting for reflection, and motivating. We explore approaches to adding "Drop-Downs" that appear after a student submits an answer and that contain additional information to support learning. We conducted randomized A/B experiments exploring the impact of these Drop-Downs on student learning in the online portion of a flipped CS1 course. The deployed Drop-Downs in this course provided explanations, reflective prompts, additional problems, and motivational messages. The results suggest that students benefit from various Drop-Downs in different contexts, indicating the possibility of personalizing content based on the student's state. We discuss the resulting design implications of Drop-Downs in online learning systems.

Author Keywords

A/B Testing; Field deployment; Randomized experiments; Online learning; Personalization; Intervention

CCS Concepts

•**Applied computing** → **E-learning;**

L@S '21, June 22–25, 2021, Virtual Event, Germany.

ACM ISBN 978-1-4503-8215-1/21/06.
http://dx.doi.org/10.1145/3430895.3460145

INTRODUCTION AND RELATED WORK

An increasing number of educational institutions have adopted online asynchronous learning [9] and flipped classrooms [16] to deliver lectures to a larger number of students. One of the most common ways to engage students with asynchronous lectures is to provide them with exercise problems to apply what they have learned [8, 2]. However, in classrooms, instructors do not provide just one intervention; they tailor their approach to the student, deciding when and what kind of support to provide based on contextual factors such as the student's frustration and the amount of time before the deadline. Therefore, we investigate the impact of alternative ways of supporting students to meet their requests in massive open online courses for more feedback and support.

Micro-interventions in educational settings have been an active research area. For example, studies show that prompting students to self-explain has been shown to help them recognize patterns in learning new material and underlying principles and their own misconceptions [18, 1, 4]. Motivational messages are effective in reducing attrition rates [17], building close relationships between students and instructors [14], fostering self-regulated strategies, and boosting academic performance [7]. Several research papers have shown that it is beneficial for students to solve related, additional problems as they help students transfer knowledge used in one question to another [6, 10, 11]. Adaptive feedback, such as explanations and hints, is a popular mechanism for supporting students in online education. Such feedback is divided into two subgroups: step-by-step hints and summative feedback [5]. Suzuki et al. further examined the design spaces of programming hints: locating student errors, presenting examples that break their programs, tracing their execution, and showing example usage of func-

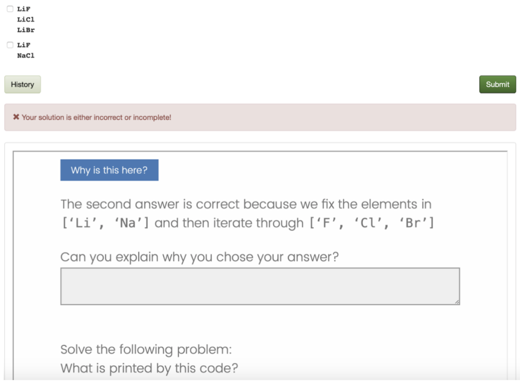

Figure 1. An example of a Drop-Down. Students saw it after clicking the green "Submit" button at the top right of this figure. This example shows the code prompt, the short explanation, and the additional problem.

What is printed by this code?

```
for metal in ['Li', 'Na']:
    for gas in ['F', 'Cl', 'Br']:
        print(metal + gas)
```

Figure 2. The code presented in the pre-test problem. Students were asked to select the option that would be printed when the code was run.

tions [15]. Some of these types of hints are implemented in existing systems [12, 13]. Yet, how different students react to these interventions is not still well understood.

To better understand the personalization of interventions, we conducted a randomized A/B comparison of different Drop-Downs in a CS1 course at a large North American research university. Due to a large number of students with a non-English background, we further analyzed the interaction effects with English proficiency while providing additional support. The contribution of our paper is the empirical result for the deployment of Drop-Downs that suggest interaction effects between their contents and students' language proficiency.

METHODS

We deployed a variety of Drop-Downs into an online exercise system (PCRS) used in a Python-based CS1 course at a research-intensive North American university. The goal of the study was to gather preliminary data on the impact of various types of Drop-Downs in a real classroom setting. 772 (of 1219) students in the course saw the Drop-Downs and consented for their data to be analyzed.

Students in the course are required to watch course-related video content and complete weekly assigned exercises on an online system (PCRS). We deployed our intervention in a set of multiple-choice questions and related video content provided as an introduction to the programming concepts for the week. Students are given unlimited attempts for each question and receive full marks (0.5% of course grade) for attempting all of them by Monday, before the course meeting.

Design of the Intervention

Our intervention was presented as a Drop-Down like the one in Figure 1. These Drop-Downs were deployed after the first multiple-choice exercise in the section. The exercise (see Figure 2) asked for the expected output of the nested loop code. This is the pre-test. Several types of Drop-Downs were deployed: self-explanation, motivational messages, an additional problem that is very similar to the pre-test, and an explanation of the solution. These variant factors were assigned independently and randomly varied. Students did not receive additional credit for responding to the Drop-Downs.

The self-explanation factor had three levels: no prompt, code prompt, and instructor prompt. The code prompt was, "Can you explain why you chose your answer?" The instructor prompt was, "Can you explain why you chose your answer? Imagine you were explaining to another student, or to your instructor." These prompts were meant to urge students to reflect on their solution to the attempted problem.

The motivational message shown to participants was, "It is completely normal to feel the urge to guess on PCRS problems, but remember, guessing on PCRS problems until you get the answer prevents you from learning, and can affect you in the final exam. Try to think for at least 30-60 seconds before you choose an answer!" These prompts were intended to reduce maladaptive behaviours like guessing.

Two explanations of the solution were presented. The short version of the explanation was, "The second answer is correct because we fix the elements in ['Li', 'Na'] and then iterate through ['F', 'Cl', 'Br']." The longer explanation of this answer was, "The second answer is correct and here's why; At line 1, we choose one of 'Li' and 'Na' and store it in variable metal. At line 2, we choose one of 'F', 'Cl', and 'Br' and store in variable gas without changing variable metal. This means that while executing the inner loop (i.e. for gas in ['F', 'Cl', 'Br']) we have the same value in metal. Therefore, 'Li' and 'Na' must not alternate." A student could receive a short explanation, a long explanation, or no explanation at all.

After the students received the intervention, they encountered a post-test problem that contains the same nested list structure as the pre-test problem. As PCRS allows students to attempt a problem multiple times, performance was measured in terms of the difference in the average number of attempts students took to arrive at the correct answer for the pre and post-test problem. An improvement from pre-test to post-test would result in a negative value using this measure.

Contextual Variables

Outside of PCRS, we asked students to complete a survey in the middle of the semester to gather their demographics, and 894 of them consented for their survey data to be used in research (M = 400, F = 471, Other = 259). We considered students' self-reported English proficiency because previous research shows it correlates with their engagement with additional support [3]. The following question on the survey were used as a contextual variable:

C1 I find it challenging to read and listen in English. For example, in reading text and listening to videos in PCRS.

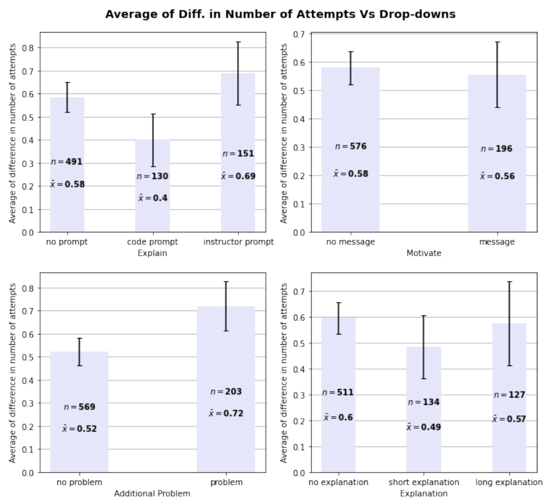

Figure 3. The main effect of the experimental factors. The lower the value of y, the fewer post-test submissions relative to the pre-test.

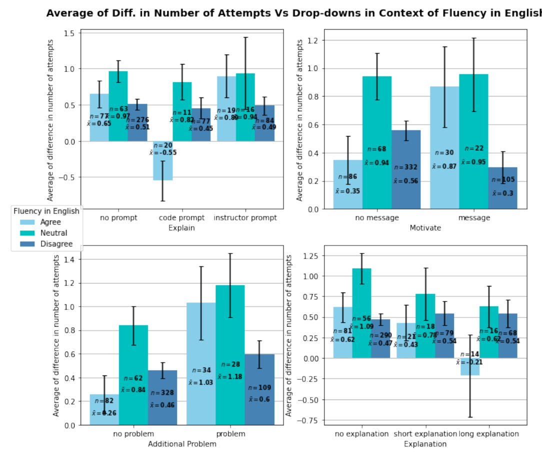

Figure 4. Performance of students with various types of Drop-downs based on their self-reported fluency in English.

The answers for this question ranged from Strongly Agree to Strongly Disagree on a 7-point Likert scale that were then binned into three categories. Strongly Agree, Agree, and Somewhat Agree were labelled Agree, Neither Agree nor Disagree was labelled Neutral, and the rest were labelled Disagree. 643 students responded to this survey question and solved the pre-test and post-test problems.

RESULTS AND DISCUSSION

To preview, the results of the deployment show that Drop-Downs does not seem to have a direct impact on student performance on the post-test problem. However, when looked at within the context of self-reported student fluency in English, those who find English challenging benefited most from the code self-explanation prompt. In addition, the motivational message had marginal positive impact on those who are fluent in English and negative impact on those who are not.

Main Effects

As shown in Figure 3, we did not find any significant effect of the Drop-Downs on student performance for any experimental factors (Explain: $H(2) = 1.326$, $p = 0.515$, $\varepsilon^2 = 0.002$; Motivational message: $H(1) = 0.586$, $p = 0.444$, $\varepsilon^2 = 0.001$; Explanation: $H(2) = 0.558$, $p = 0.757$, $\varepsilon^2 = 0.001$) [1], except that the additional problem had a marginally significant negative effect ($H(1) = 3.046$, $p = 0.081$, $\varepsilon^2 = 0.004$). While the lack of impact for the general problem is disappointing, it highlights the potential importance of contextualization, as described in the next subsection.

Fluency in English

When we grouped students by self-reported fluency in English, we found the code prompt of the Explain factor significantly positively affected students with lower English proficiency and that the motivational message had marginally different effects depending on students' fluency, as shown in Figure 4.

The result of the Kruskal-Wallis test revealed that the three variant groups of the Explain factor had significantly varying effects on performance for students who agreed to facing challenges in comprehending the questions in English ($H(2) = 11.628$, $p = 0.003$, $\varepsilon^2 = 0.101$) but not for others (Neutral: $H(2) = 1.098$, $p = 0.577$, $\varepsilon^2 = 0.012$; Disagree: $H(2) = 0.214$, $p = 0.898$, $\varepsilon^2 = 0.000$). A Mann-Whitney U test with a Bonferroni correction was then conducted to study the pairwise differences amongst the three variant groups. The results indicate that the code prompt had a significant positive effect on student performance compared to the instructor prompt and no prompt variants ($U(N_{no\_prompt} = 77, n_{code} = 20) = 449.500$, $p = 0.001$, $d_{cohen} = 0.606$ and $U(n_{code} = 20, n_{instruct} = 19) = 92.500$, $p = 0.002$, $d_{cohen} = 0.976$). The average difference of number of attempts between the pre- and post-test problems was significantly lower with the code prompt than the no prompt case ($\bar{x}_{no\_prompt} = 0.65$, $\bar{x}_{code} = -0.55$) and the instructor prompt ($\bar{x}_{instruct} = 0.83$, $\bar{x}_{code} = -0.55$). However, there was no significant difference between having the instructor prompt and having no prompts at all ($U(n_{no\_prompt} = 77, n_{instruct} = 19) = 649.000$, $p = 0.209$, $d_{cohen} = 0.155$).

The other Drop-Down types had no significant effect in the context of student fluency in English (Additional problem: Agree: $H(1) = 3.097$, $p = 0.078$, $\varepsilon^2 = 0.027$, Neutral: $H(1) = 1.356$, $p = 0.244$, $\varepsilon^2 = 0.015$, Disagree: $H(1) = 1.319$, $p = 0.251$, $\varepsilon^2 = 0.003$; Explanation: Agree: $H(2) = 1.656$, $p = 0.437$, $\varepsilon^2 = 0.014$, Neutral: $H(2) = 1.852$, $p = 0.396$, $\varepsilon^2 = 0.001$, Disagree: $H(2) = 0.427$, $p = 0.808$, $\varepsilon^2 = 0.021$). However, it is worth noting a slight improvement in student performance with the motivational message for those who are fluent in English and a slight decline for those who are not (Agree: $H(1) = 1.724$, $p = 0.189$, $\varepsilon^2 = 0.015$; Neutral: $H(1) = 0.038$, $p = 0.845$, $\varepsilon^2 = 0.000$; Disagree: $H(1) = 2.939$, $p = 0.086$, $\varepsilon^2 = 0.007$).

Limitations

There are several limitations in our study. First, our measure of English proficiency was subjective. Our C1 (described in

[1] We ran nonparametric statistical tests because the data was not normally distributed.

section 2.0.2) merged non-native English speakers who are confident with English and native speakers. However, some students may be overconfident about their English skills. Second, our pre-test and post-test measured slightly different concepts. Even though both exercises were about "for" loops, the pre-test checks the understanding of the order of the executions, while the post-test checks understanding of the total number of executions. Finally, we deployed the Drop-Downs only in one question. Although the effect of interventions might depend on attributes of the problem, we did not compare the effectiveness of Drop-Downs across problems.

Implications for Future Work

This paper investigates the possibility of personalizing additional support in exercise problems on online learning platforms. The deployment of additional support via Drop-Downs in a university course with hundreds of students revealed that even though most of the interventions did not uniformly work for everyone, there was suggestive evidence that the English proficiency of students changed the effectiveness of the support. We hope practitioners will better tailor online education based on students' background in the language of instruction. We also hope researchers will build on this study to better understand why such personalization occurs and further explore the difference between the needs of students with higher language proficiency and lower proficiency.

ACKNOWLEDGMENTS

We would like to thank instructors of the CS1 course and Sam Maldonado for helping us to deploy our interventions and the CS1 students who consented to have their data analyzed for this study. This work was funded by the Office of Naval Research grant number #N00014-18-1-2755 and the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number #RGPIN-2019-06968 to Joseph Jay Williams.

REFERENCES

[1] Laura K Allen, Danielle S McNamara, and Matthew McCrudden. 2015. Change your Mind: Investigating the Effects of Self-Explanation in the Resolution of Misconceptions. In *CogSci*.

[2] Mohamed Ally. 2004. Foundations of educational theory for online learning. *Theory and practice of online learning* 2 (2004), 15–44.

[3] 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 Tenth International Conference on Learning Analytics & Knowledge*. 448–453.

[4] Michelene TH Chi, Miriam Bassok, Matthew W Lewis, Peter Reimann, and Robert Glaser. 1989. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science* 13, 2 (1989), 145–182.

[5] Tyne Crow, Andrew Luxton-Reilly, and Burkhard Wuensche. 2018. Intelligent tutoring systems for programming education: a systematic review. In *Proceedings of the 20th Australasian Computing Education Conference*. 53–62.

[6] Lyn D English. 1998. Reasoning by analogy in solving comparison problems. *Mathematical Cognition* 4, 2 (1998), 125–146.

[7] Tiong-Thye Goh, Boon-Chong Seet, and Nian-Shing Chen. 2012. The impact of persuasive SMS on students' self-regulated learning. *British Journal of Educational Technology* 43, 4 (2012), 624–640.

[8] Thomas Janicki and Jens O Liegle. 2001. Development and evaluation of a framework for creating web-based learning modules: a pedagogical and systems perspective. *Journal of Asynchronous Learning Networks* 5, 1 (2001), 58–84.

[9] A Frank Mayadas, John Bourne, and Paul Bacsich. 2009. Online education today. *Science* 323, 5910 (2009), 85–89.

[10] Laura R Novick. 1995. Some determinants of successful analogical transfer in the solution of algebra word problems. *Thinking & Reasoning* 1, 1 (1995), 5–30.

[11] Laura R Novick and Keith J Holyoak. 1991. Mathematical problem solving by analogy. *Journal of experimental psychology: Learning, memory, and cognition* 17, 3 (1991), 398.

[12] Thomas W Price, Samiha Marwan, Michael Winters, and Joseph Jay Williams. 2020. An Evaluation of Data-Driven Programming Hints in a Classroom Setting. In *International Conference on Artificial Intelligence in Education*. Springer, 246–251.

[13] Kelly Rivers and Kenneth R Koedinger. 2017. Data-driven hint generation in vast solution spaces: a self-improving python programming tutor. *International Journal of Artificial Intelligence in Education* 27, 1 (2017), 37–64.

[14] Firat Sarsar. 2017. Student and Instructor Responses to Emotional Motivational Feedback Messages in an Online Instructional Environment. *Turkish Online Journal of Educational Technology-TOJET* 16, 1 (2017), 115–127.

[15] Ryo Suzuki, Gustavo Soares, Elena Glassman, Andrew Head, Loris D'Antoni, and Björn Hartmann. 2017. Exploring the design space of automatically synthesized hints for introductory programming assignments. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2951–2958.

[16] Bill Tucker. 2012. The flipped classroom. *Education next* 12, 1 (2012), 82–83.

[17] Lya Visser. 1998. The development of motivational communication in distance education support. (1998).

[18] 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*. 470–474.