When More Intelligent Tutoring in the Form of Buggy Messages Does not Help

Conference Paper · June 2015

DOI: 10.1007/978-3-319-19773-9_112

Neil T. Heffernan
Worcester Polytechnic Institute

When More Intelligent Tutoring in the Form of Buggy Messages Does Not Help

Douglas Selent(✉) and Neil Heffernan

Worcester Polytechnic Institute, Worcester, MA, USA
{dselent,nth}@wpi.edu

Abstract. This paper reports on the null results from two large scale randomized controlled trials that were run in the ASSISTments online tutoring system. Both studies attempted to use reactive buggy messages to help students learn; one in the form of short 20–30 second videos and another in the form of large color-coded text. Bug messages were supplied for common wrong answers for one-step equation problems in both studies. Despite the large amount of prior research done on error analysis, both interventions using the predicted common wrong answers were unsuccessful at helping students.

Keywords: Buggy message · Randomized controlled trial · Video feedback

1 Introduction

A "buggy message" provides a unique and specific to the incorrect answer entered by the student. The notion of "bugs" was first introduced by Brown and Burton in 1978 and the early 1980's with their series of "BUGGY" programs [1]. The main goal of BUGGY was to show teachers that incorrect answers supplied by the student are not random and are a result of systematic errors and misconceptions. An extension of the BUGGY program called DEBUGGY and IDEBUGGY (an online interactive version of DEBUGGY) attempted to match the incorrect answer to a set of previously defined bug rules [3].

Brown and VanLehn continued this work using repair theory to explain why students make some errors but not others and the cause of the errors [2]. After running some empirical studies VanLehn found that errors were not stable and students did not consistently make the same mistake twice [8].

Sleeman also created a program to discover bugs. Sleeman ran an experiment on a group of 24 14-year-old students on algebra problem using the Leeds Modeling System (LMS) bug database to establish classes for the different types of errors [6]. Further analysis was done on the types of errors students made in attempt to explain bug-migration (inconsistent student errors). Sleeman proposed the mechanism of Mis-generalization (incorrectly inferring rules) to explain bug migration and why LMS originally missed several bugs [7].

Throughout the 80's and 90's there was a large amount of work focused on student modeling. This work includes student modeling based on the incorrect answers and

C. Conati et al. (Eds.): AIED 2015, LNAI 9112, pp. 768–771, 2015.
DOI: 10.1007/978-3-319-19773-9_112

automatically discovering the incorrect processes that generated the incorrect answers. Sison summarizes several machine learning programs that attempt automatically construct student models [5].

The two studies in this paper follow up on the randomized controlled trial run by Selent and Heffernan, who tried to help students with buggy messages [4]. Since it was believed that their initial experiment failed due to poor tutoring content, two additional randomized controlled trials were run with focus on improving the tutoring content. One experiment used buggy messages in the form of short 20-30 second videos. Combining both buggy messages with short videos has not been done before and is a novel form of tutoring. Another experiment provides buggy messages with large color-coded text providing both context to the problem and the reason why the answer is incorrect. The hypothesis is that the answer-specific help would help students more than on-demand help. This paper reports on the null results of the two randomized controlled trials run in the ASSISTments online tutoring system.

2 Experiment Design

2.1 Study 1 (Text Buggy Messages)

A skill builder problem set was created for problems on solving one-step equations with integers, which is typically taught in seventh grade for students of ages 12-13. The control group received problems where students had the option to click on a hint button. A student could click on the hint button up to three times. The first hint contained a similar example problem, showing all the steps to reach the solution. The second and third hint contained steps to solve the current problem each time the hint button was clicked, until the last hint revealed the answer to the problem. This currently is the best form of help for this problem set in the ASSISTments system, and therefore, was used as the problem set for the control group. The experiment group received the same problems as the control group with the addition of buggy text messages that appeared as soon as a student submitted an incorrect answer. These buggy messages showed the steps of the problem the student did correctly and the step that the student performed incorrectly as well as providing the correct rules the student needed to fix their mistake. If a student entered an incorrect answer that was not predicted, a generic message was shown to the student to notify the student that he/she had answered the problem incorrectly. This experiment design compares a strong control group, where students have a set of hints that walk through an example problem, to an experiment group that adds buggy messages to the current problems.

2.2 Study 2 (Video Buggy Messages)

Similar to the first study, a skill builder problem set was created for problems on solving one step equations with integers. The control group received problems where students had the option to click on a hint button which gave the answer. This is the standard form of help in ASSISTments, therefore was used as the control. The experiment group received problems that did not have the option to click on a hint button, but received help in the form of video buggy messages. Students received a short 20-30 second video when they entered a predicted incorrect answer. This video explained what process the student used to arrive at their incorrect answer and how to start on the correct solution path. The videos were created by Andrew Burnett, a former middle school teacher, where Andrew explained the problem using a white board. The videos did not give the solution to the student. This experiment design compares a weak control group, where students only had the option to see the answer, to an experiment group that had proactive help in the form of video buggy messages hints but did not have on-demand hints.

3 Analysis

Students who answered three problems correctly in a row and completed the problem set without any tutoring were removed. Since these students did not receive any tutoring, they did not experience the conditions of the experiment. A χ^2 test was used to ensure the number of students was not significantly different between the two groups for the initial condition assignment for study 1, χ^2 (1, n=490) = 2.359, p>0.05, and study 2, χ^2 (1, n=1379) = 0.059, p>0.05. A χ^2 test was also used to ensure the number of students that were removed was not significantly different between the two groups for study 1, χ^2 (1, n=268) = 1.493, p>0.05, and study 2, χ^2 (1, n=730) = 0.005, p>0.05. After removing these students a total of 222 (control = 104, experiment = 118) students remained for the first study and a total of 649 (control = 328, experiment = 321) students remained for the second study.

The following statistics were used to measure the helpfulness of the tutoring.

1. The number of students that finish the problem set
2. Number of problems it takes a student to complete the problem set
3. Correctness on the next problem following a student's first incorrect response
4. Attempts on the next problem following a student's first incorrect response
5. Hints on the next problem following a student's first incorrect response

The completion percentage for study 1 was 64% for the control group and 68% for the experiment group and the completion percent for study 2 was 88% for the control group and 86% for the experiment group. A χ^2 showed that there were no significant differences in completion rates between the control group and the experiment group for study 1, χ^2 (1, n=147) = 1.15, p>0.05 and study 2, χ^2 (1, n=564) = 0.348, p>0.05.

A two-tailed t-test showed that there were no significance differences for the number of problems it took students to complete the problem set in study 1 ($\mu_1 = 6.0$, $\mu_1 = 5.8$, $n_1 = 67$, $n_2 = 80$, $p = 0.69$) and study 2 ($\mu_1 = 7.3$, $\mu_1 = 6.9$, $n_1 = 289$, $n_2 = 275$, $p = 0.30$). A MANOVA was done using SPSS to see if there was a significant different between the control and experiment conditions on the next question after a student's first incorrect response. For the first study a MANOVA was run with condition as the only factor and correctness, attempts, and hints as dependent measures. The results show no main effect with Wilks' $\lambda = 0.98$, $F(3, 143) = 0.989$, and $p = 0.4$. For the second study the same analysis was done, except hint use was excluded as a measure since the experiment group did not have the option to ask for hints. The results show no main effect with Wilks' $\lambda = 0.991$, $F(2, 561) = 2.674$, and $p = 0.07$.

4 Conclusions and Future Work

In this paper two studies were run in the ASSISTments tutoring system using tutoring with different types of buggy messages. Although both studies resulted in a null result, we choose to report on them for the following reasons. Despite the large amount of research done on error analysis, there were few interventions done using buggy messages and none done at the scale in this paper. We show that neither color-coded text messages nor short videos were effective. It is future work to focus less on how or why a student arrived at a specific wrong answer and focus other methods of tutoring to communicate better to the students.

Acknowledgement. We acknowledge funding for ASSISTments from the NSF (1316736, 1252297, 1109483, 1031398, 0742503, 1440753), the U.S. Dept. of Ed. GAANN (P200A120238), ONR's "STEM Grand Challenges," and IES (R305A120125, R305C100024).

References

1. Brown, J.S., Burton, R.R.: Diagnostic Models for Procedural Bugs in Basic Mathematical Skills*. Cognitive science **2**(2), 155–192 (1978)
2. Brown, J.S., VanLehn, K.: Repair theory: A generative theory of bugs in procedural skills. Cognitive science **4**(4), 379–426 (1980)
3. Burton, R.B.: DEBUGGY: diagnosis of errors in basic mathematical skills. In: Sherman, D., Brown, J.S. (eds.) Intelligent tutoring systems. Academic Press, London (1981)
4. Selent, D., Heffernan, N.: Reducing student hint use by creating buggy messages from machine learned incorrect processes. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 674–675. Springer, Heidelberg (2014)
5. Sison, R., Shimura, M.: Student modeling and machine learning. International Journal of Artificial Intelligence in Education (IJAIED) **9**, 128–158 (1998)
6. Sleeman, D.: Basic algebra revisited: a study with 14-year-olds. International Journal of Man-Machine Studies **22**(2), 127–149 (1985)
7. Sleeman, D.: Mis-generalisation: an explanation of observed mal-rules. In: Proceedings of the Sixth Annual Conference of the Cognitive Science Society, pp. 51–56 (1984)
8. Van Lehn, K.: Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. The Journal of Mathematical Behavior (1982)
9. https://sites.google.com/site/assistmentsdata/projects/selent2015