DO SUMMARIES SUPPORT LEARNING FROM POST-PROBLEM REFLECTIVE DIALOGUES?

Sandra Katz, Patricia Albacete, Pamela Jordan

Learning Research and Development Center, University of Pittsburgh

Do Summaries Support Learning from Post-problem Reflective Dialogues?

Sandra Katz, Patricia Albacete, and Pamela Jordan

Learning Research and Development Center, University of Pittsburgh
{katz, palbacet, pjordan}@pitt.edu

Abstract. This poster reports on a study that compared three types of summaries at the end of natural-language tutorial dialogues and a no-dialogue control, to determine which type of summary, if any, best predicted learning gains. Although we found no significant differences between conditions, analyses of gender differences indicate that female students benefit most from the most concise summary (restatement of a reflection question and its answer).

1 Introduction

Natural-language tutoring systems typically wrap up a discussion about a problem or complex question with a summary of the line of reasoning (LOR) that leads to its solution (e.g., [1-3]). However, observations of human tutoring reveal that tutors seldom present complete LOR summaries, or give other types of dialogue summaries. For example, the tutor might remind the student of the main question and its conclusion but leave out the detailed, intermediate LOR. We refer to these as *Conclusion* summaries. Alternatively, the tutor might present the question and its answer, as in Conclusion summaries, but add some "take home advice", such as how the discussion could be applied more generally to similar types of problems. We call these *Advice* summaries.

This poster reports on a study that compared the potential benefits of LOR summaries with these alternative types of dialogue summaries and a no-summary control. We hypothesized that a full LOR summary would be more beneficial for students with low prior knowledge than for higher incoming knowledge students. The former type of student may make more mistakes and need help pulling together the LOR. We hypothesized that a Conclusion summary would be more beneficial for students with mid-level incoming knowledge because they are likely to be able to self-explain the connection between the question and its conclusion. Finally, we hypothesized that an Advice summary would be more beneficial for high prior knowledge students because they are ready to generalize from a line of reasoning that they may be able to self-generate.

2 A Study of Summarization and Student Characteristics

Research Platform. Rimac, a web-based natural-language tutoring system for conceptual physics, served as a research platform for this study. Rimac engages students in conceptual discussions (reflective dialogues) after they solve physics problems [3].

Participants. One hundred and ninety students, from three high schools in southwestern Pennsylvania, USA, participated in the study. However, the data from only 96 of these students was complete and used for analysis. Students were randomly assigned to one of the four conditions, within each class: No-summary (24), Line of Reasoning summary (23), Conclusion summary (25), and Advice summary (24).

Procedure. The study took place during class. Students took a 21-item pretest that covered dynamics concepts. Students in the summary conditions then solved three problems on paper and, after each problem, reviewed the video of a sample solution and engaged with the automated tutor in several reflective dialogues per problem. Students in the No-summary condition solved an additional problem which was isomorphic to one other problem, to control for time on task. Finally, students took a posttest that was isomorphic with the pretest.

Results and Discussion. Students across conditions learned from interacting with the tutor. However, contrary to our hypotheses, there were no interactions between students' prior knowledge and learning gains. Furthermore, there were no significant differences in test gain scores between conditions ($F_{(3,92)}= 0.289$, $p=.833$). This suggests that end-of-dialogue summarization is not a predictor of learning gains.

Although we did not find aptitude-treatment interactions, we observed differences in gain scores between genders. The mean gain was significantly greater for female students (($t(94)=2.096$, $p=.039$). Within conditions, this difference held only for Conclusion summaries ($t(23)=2.081$, $p=.049$), with a trend for these summaries to be better for females than males, but only for test items rated as difficult ($t(23)=2.000$, $p=.057$).

It is possible that students would learn more from dialogue summaries if they participated in generating them—for example, if the system prompted students to fill in missing pieces of information. We are conducting a study to address this question.

3 References

1. Evens, M. and J. Michael, *One-on-one tutoring by humans and computers*. 2006: Psychology Press.
2. Graesser, A.C., et al., *AutoTutor: A simulation of a human tutor*. Cognitive Systems Research, 1999. **1**(1): p. 35-51.
3. Katz, S. and P.L. Albacete, *A tutoring system that simulates the highly interactive nature of human tutoring*. Journal of Educational Psychology, 2013. **105**(4): p. 1126.