WHEN IS IT HELPFUL TO RESTATE STUDENT RESPONSES WITHIN A TUTORIAL DIALOGUE SYSTEM?

Pamela W. Jordan, Patricia L. Albacete, Sandra Katz

Learning Research and Development Center, University of Pittsburgh

Paper presented at the 17th International Conference on Artificial Intelligence in Education

Madrid, Spain, June 2015

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130441 to the University of Pittsburgh. The opinions are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

When Is It Helpful to Restate Student Responses within a Tutorial Dialogue System?

Pamela Jordan, Patricia Albacete, and Sandra Katz

Learning Research and Development Center
University of Pittsburgh, Pittsburgh PA 15260, USA,
pjordan@pitt.edu

Abstract. Tutorial dialogue systems often simulate tactics used by experienced human tutors such as restating students' dialogue input. We investigated whether the amount of tutor restatement that supports student inference interacts with students' incoming knowledge level in predicting how much students learn from a system. We found that students with lower incoming knowledge benefit more from an increased level of these types of restatement while students with higher incoming knowledge benefit more from a decreased level of such restatements.

Keywords: tutorial dialogue, restatement, inference, prior knowledge

1 Introduction

A tutor restating part of a student's dialogue contribution can serve many purposes and at the surface level can range from exact repetitions to semantic reformulations [6]. Some of the purposes for restatement that are found in human tutoring are acknowledging the correct parts of a student's response [4, 3], marking (i.e., focusing on part of a response) and modeling a better answer [4, 2]. Because restatements of correct responses have been shown to correlate with learning [4], this opens the possibility that restatements could cause learning by strengthening correct knowledge. While restatements of various types have been incorporated into a number of tutorial dialogue systems (e.g. Circsim-Tutor [5], AutoTutor [9], Beetle II [4]), restatement has not been tested in isolation from other tactics to determine whether it has any causal connection to learning.

Here, we explore a different type of restatement that has the purpose of showing consequence [6]–that is, making an inference explicit. We test two alternative hypotheses about this type of restatement: 1) that it will equally benefit all students and 2) that its effect varies according to students' incoming knowledge. If it strengthens learning of correct knowledge, then it should benefit all students equally. However, we expect students with lower incoming knowledge to benefit more from an increased level of consequence restatement while higher incoming knowledge students would benefit more from a decreased level of such restatements. Our expectation is motivated by prior research which found that unpacking the inferences in text supports comprehension among low-knowledge readers, while less cohesive (higher inference-inducing) text is better for high-knowledge readers [8]. Reduced cognitive load is a proposed alternative explanation for the “reverse cohesion effect”, particularly for high-knowledge readers when reading a less coherent text. Cognitive load increases when they have to reconcile their existing schema about the topic discussed in the text with the background material provided in a “highly coherent” text [7].

2 Methods

Participants The study was conducted in high school physics classes at three schools in the Pittsburgh PA area with 168 students participating. Students were randomly assigned to one of two conditions: high restatement (N= 88; 30 females, 58 males) and low restatement (N= 80; 27 females, 53 males).

Materials We used an existing version of the Rimac natural-language tutoring system to conduct our experiment. A brief description of the system can be found in [1]. It engages students in post-problem solving reflection dialogues on the concepts involved in solving quantitative problems. Rimac’s content was developed in consultation with high school physics teachers. For this experiment we used its dynamics content which covers three problems with two reflection questions per problem, and a 21 item pretest and isomorphic post-test which included nine multiple-choice problems and 12 open-response problems.

To create the high restatement system, three dialogue authors added consequence inference restatements of student responses when it would result in either: 1) an explicit concluding statement at the end of a sub-dialogue that draws upon the student’s responses during the sub-dialogue or 2) an explicit if-then statement that draws the “if” or “then” part from the student’s immediately preceding response. An example of the latter context for consequence inference restatements is shown below for the high restatement condition:

T: While the arrow is flying is anything touching or in contact with it?
S: No [there is nothing touching the arrow during its flight]
T: I agree. Hence since **there is nothing touching the arrow during its flight** there is no contact force applied to it.

To create the low restatement version authors identified all restatements involving inference and either deleted or replaced the restatement with a pronoun, taking care not to disturb the coherency of the dialogue. The low restatement version of the above example is identical to the high restatement version, except for the second tutor turn, which would read: “I agree. Hence there is no contact force applied to it.”

Procedure On the first day, the teacher gave the pretest in class and assigned the three dynamics problems, referred to in the Materials section, for homework. During the next one to two class days (depending on whether classes were approximately 45 min. or 80 min. long), students used Rimac in class. For each homework problem, students watched a video “walkthrough” of a sample solution and then engaged in the problem’s reflective dialogues. The videos focused on procedural/problem-solving knowledge, while the dialogues focused on conceptual knowledge. Finally, at the next class meeting, teachers gave the post-test, which was isomorphic to the pretest.

3 Results

Learning Performance To determine whether interaction with the tutoring system, regardless of condition, promoted learning, we compared gains from pretest to post-test using paired samples t-tests. When students in each condition were considered separately, we found a statistically significant difference for all problems together (H: $t(87)=3.56$, $p<.01$; L: $t(79)=4.49$, $p<.01$), multiple-choice problems (H: $t(87)=2.73$, $p<.01$; L: $t(79)=2.39$, $p<.02$), and open-response problems (H: $t(87)=3.13$, $p<.01$; L: $t(79)=4.8$, $p<.01$), where H=high restatement, L=low restatement. These results suggest that students in both conditions learned from both versions of the system.

High Restatement vs. Low Restatement To test whether students who used the high restatement version of the system performed differently from students who used the low restatement version, we compared students' gains from pretest to post-test using independent samples t-tests. We found no significant differences between conditions for any subset of problems even when controlling for pretest (there were no differences for mean time on task). This suggests that this type of restatement does not support learning by strengthening correct knowledge. If it did, then we would expect to see a difference between conditions for learners of all prior knowledge levels.

Prior knowledge-treatment interaction To investigate whether there was a prior knowledge treatment interaction, we performed a multiple regression analysis using condition, prior-knowledge (as measured by pretest) and condition * prior-knowledge (interaction) as explanatory variables, and gain as the dependent variable. When all problems were considered together, we found a significant interaction between condition and prior knowledge in their effect on gains (t=-2.126, p=0.04). Likewise, we found a significant interaction when we considered only gains on open-response problems (t=-2.689, p=0.01). However, for multiple-choice problems we did not find a significant interaction.

Graphing gain vs. prior knowledge for all problems suggested that students with pretest scores that are 35% correct or less benefit more from the high restatement version of the system. However students with pretest scores above 35% correct benefit more from the low restatement version. Graphing gain vs. prior knowledge for open-response problems suggested that students with pretest scores of 23% or less on open-response items benefit more from higher restatement and students with pretest scores greater than 23% benefit more from lower restatement. Consistent with the results reported in [8], both findings offer evidence to support our hypothesis that lower knowledge students benefit more from high restatement in inferential contexts while higher knowledge students benefit more from low restatement.

4 Conclusions and Future Work

We found that students learned from the tutoring system, across conditions, as measured by normalized gain scores. There was no difference between conditions, which suggests that this type of restatement may not cause learning by strengthening correct knowledge. However, we did find a prior knowledge treatment interaction which supported our hypothesis that lower knowledge students would benefit more from a high restatement system while higher knowledge students would benefit more from a low restatement system. Thus system designers may need to be judicious in their use of restatement as it may dampen learning if there is a mismatch to students' prior knowledge levels.

In future research, we plan to determine if the benefits of the high and low restatement versions of Rimac can be used advantageously in a system that adapts to students' knowledge levels and to formulate and test additional hypotheses for other types of restatement (e.g., that have other purposes).

Acknowledgements. We thank Stefani Allegretti, Michael Lipschultz, Diane Litman, Dennis Lusetich, Svetlana Romanova, and Scott Silliman for their contributions. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130441 to the University of Pittsburgh. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute or the U.S. Department of Education.

References

1. Albacete, P., Jordan, P., Katz, S.: Is a dialogue-based tutoring system that emulates helpful co-constructed relations during human tutoring effective? In: AIED: 17th International Conference on Artificial Intelligence in Education (2015)
2. Becker, L., Ward, W., Vuuren, S.V., Palmer, M.: Discuss: A dialogue move taxonomy layered over semantic representations. In: IWCS 2011: The 9th International Conference on Computational Semantics. Oxford, England (January 2011)
3. Chi, M.T.H., Roy, M.: How adaptive is an expert human tutor? In: 10th International Conference on Intelligent Tutoring Systems (ITS). pp. 401–412 (2010)
4. Dzikovska, M., Campbell, G., Callaway, C., Steinhauser, N., Farrow, E., Moore, J., Butler, L., Matheson, C.: Diagnosing natural language answers to support adaptive tutoring. In: International FLAIRS Conference (2008)
5. Freedman, R.: Using a reactive planner as the basis for a dialogue agent. In: International FLAIRS Conference (2000)
6. Hyland, K.: Applying a gloss: Exemplifying and reformulating in academic discourse. Applied Linguistics 28(2), 266–285 (2007)
7. Kalyuga, S., Ayres, P.: The expertise reversal effect. Educational Psychology 38, 23–31 (2003)
8. McNamara, D., Kintsch, E., Songer, N., Kintsch, W.: Are good texts always better? text coherence, background knowledge, and levels of understanding in learning from text. Cognition and Instruction 14, 1–43 (1996)
9. Person, N., Graesser, A., Kreuz, R., Pomeroy, V.: Simulating human tutor dialog moves in autotutor. International Journal of Artificial Intelligence in Education 12(23-39) (2003)