Running Head: EVALUATION OF RMT IN THE CLASSROOM

Research Methods Tutor:

Evaluation of a Dialogue-Based Tutoring System in the Classroom

Elizabeth Arnott

Chicago State University

Peter Hastings

David Allbritton

DePaul University

FOR SCIP SPECIAL ISSUE

Corresponding Author:

Elizabeth Arnott

Chicago State University Department of Psychology

9501 S King Dr.

Chicago, IL 60628

Telephone: 773-821-2437

Email: earnott@csu.edu

Abstract

Research Methods Tutor (RMT) is a dialogue-based intelligent tutoring system for use in conjunction with undergraduate psychology research methods courses. RMT includes five topics that correspond to the curriculum of introductory research methods courses: ethics, variables, reliability, validity, and experimental design. We evaluated the effectiveness of the RMT system in the classroom using a non-equivalent control group design. Students in three classes (n = 73) used RMT, and students in two classes (n = 52) did not use RMT. Results indicated that the use of RMT yielded strong learning gains of .71 SD above classroom instruction alone. Further, the dialogue-based tutoring condition of the system resulted in higher gains than the textbook-style condition (CAI version) of the system. Future directions for RMT include the addition of new topics and tutoring elements.

Research Methods Tutor:
Evaluation of a Dialogue-Based Tutoring System in the Classroom

A course in research methodology is a part of the required curriculum for psychology majors at most institutions. Although an understanding of the research process is a fundamental aspect of the comprehension of psychology as a disciple, many undergraduates struggle with research methods courses. Research methods courses tend to be more technical, quantitative, and applied than other types of psychology courses. Similar to most college-level courses, time spent in class is rarely enough to provide the students with sufficient practice, but unlike other courses, research methods is not something the students can learn without practice applying their knowledge to research scenarios. As the students are unlikely to encounter research scenarios in their everyday lives, they often lack the ability to sufficiently practice this skill. This paper describes the evaluation of Research Methods Tutor (RMT), an intelligent tutoring system that engages students in one-on-one dialogues about various topics in undergraduate psychology research methods.

There is considerable evidence for the effectiveness of one-on-one tutoring. Studies of tutored students have shown that they can achieve learning gains up to 2.3 standard deviations above classroom instruction alone (Bloom, 1984). The extent to which the student is an active participant in a dialogue has been shown to positively correlate with learning outcomes (Wood & Middleton, 1975; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001). Tutorial dialogues allow interaction between the tutor and student, and, therefore, can yield a number of potential advantages over more traditional learning methods. Tutorial dialogue involves cooperation to solve a wide

variety of problems (Graesser, Person, and Magliano, 1995). This cooperation can allow the tutor to assess the student's current level of knowledge and appropriately react to any changes in knowledge level (Anderson, Corbett, Koedinger, & Pelletier, 1995). Tutors can act to repair any errors in student understanding, providing immediate feedback and decreasing time necessary for concept mastery (Corbett & Anderson, 1991). Dialogue also allows tutors to model appropriate strategies when the student is unable to generate them on his/her own (Lesgold, Lajoie, Bunzo, & Eggan, 1992).

Although engaging in tutorial dialogue has many potential advantages, many students do not have access to skilled tutors. Tutoring can involve prohibitive expenses and time commitments, especially for non-traditional students. Intelligent Tutoring Systems (ITS's) avoid the practical disadvantages of one-on-one human tutoring. ITS's can provide some of the learning benefits of one-on-one human tutoring with little or no cost to the student, and they can be accessed at any time, which provides flexibility for working students or students with children. A large scale study on the effectiveness of an algebra tutoring system in high school settings found that students who used the tutor had basic skills test scores that were approximately 100% higher than a comparison class that did not use the tutor (Koedinger, Anderson, Hadley, & Mark, 1997). Dialogue-based ITS's support natural language interaction with students and can allow students to experience collaborative problem solving and feedback similar to that provided by a human tutor. In laboratory experiments, one dialogue-based ITS, AutoTutor, has been shown to produce learning gains of up to one standard deviation above reading a textbook alone (Graesser, Jackson, Mathews, Mitchell, Olney, Ventura, Chipman, Franceschetti, Hu, Louwerse, Person, & the Tutoring Research Group, 2003).

Description of the System

RMT is a dialogue-based intelligent tutoring system that is designed for use as an adjunct to introductory psychology research methods courses. Like its predecessor AutoTutor, RMT engages students in a natural language dialogue, evaluating student responses against sets of expected answers (Wiemer-Hastings, Graesser, Harter, & the Tutoring Research Group, 1998). The tutor asks the student a question, and the student types a response into the text box on the screen. RMT makes the comparison between responses and expected answers using latent semantic analysis, or LSA (Landauer, Foltz, & Laham, 1998), which creates a high-dimensional vector representation of both the expected answer and student's response based on a body of domain-relevant texts. The cosine of the vectors represents the similarity of the student's answer to the expected answer.

The RMT system includes five topics from the curriculum of typical introductory psychology research methods courses: ethics, variables, reliability, validity, and experimental design. Students are assigned a topic module to complete while they are learning about the same concept in the classroom. Following Bloom's (1956) taxonomy, each topic module contains a mix of conceptual, analytic, and synthetic questions. Conceptual questions are questions that have a single correct answer ("What is an independent variable?"). Analytic questions are those that require a student to not only know about concepts but to apply those concepts to new situations ("What is the independent variable in this experiment?"). Synthetic questions require students to possess a more advanced understanding of the concepts and to construct solutions to new problems ("Construct a study that contains an independent variable and a dependent variable.").

RMT includes two instructional conditions that are used for the purpose of system assessment. In the tutoring instructional condition, the system interacts with the student. It engages in a natural language dialogue with the student, asking questions ("What is reliability?"), providing prompts ("The reliability of a measure is the extent to which the measure is…") and hints ("Internal validity is about finding what type of relationship between your independent variable and dependent variable?"), and providing summaries of the key concepts. The computer-aided instruction (CAI) condition is considerably less interactive. In this condition, the system covers the same information, but presents it in a textbook-style fashion and asks multiple choice questions at the end of each section to help ensure that the student reads the material.

In addition to the instructional conditions, there are also two presentation modes in the RMT system. The "face" of the RMT agent presentation mode is an animated pedagogical agent named Mr. Joshua (Figure 1). Mr. Joshua appears on-screen and communicates with the student using synthesized speech and a number of hand and facial gestures, including nodding and turning his head, blinking his eyes, and moving his hands.

Insert Figure 1 about here.

The text-only presentation mode has no agent. The questions and responses of the tutor simply appear on-screen in the form of written text. While the text-only version is technologically much simpler, it has been shown that in some situations, learners pay little attention to text presented on ITS screens (Salvucci & Anderson, 1998). In addition, textual displays combined with additional figures may visually overload the

student and "short circuit" visual processing (Clark & Meyer, 2002). Thus, a secondary goal of our research is to determine whether or not a talking head tutor is superior to text-only tutoring and, if so, under what conditions.

The primary goal of the assessment was to examine the overall effectiveness of the RMT system. We hypothesized that the classes that used RMT would exhibit greater learning gains than classes that did not use RMT. We also assumed that greater interaction between the system and the student would result in increased learning, and, thus, predicted that students would show greater evidence of learning in the tutoring condition than in the CAI condition. Finally, based on previous findings concerning students' attention to on-screen text, we predicted that students using the animated agent would outperform students who used the text-only presentation mode.

Method

Participants

During the winter and spring quarters of 2006, RMT was assessed using five introductory research methods courses at DePaul University. The students in three of these classes (n = 83) used RMT throughout the quarter as part of the course requirement. The students in the other two classes (n = 53) did not use RMT and served as a non-equivalent control group. Four of the five courses (2 RMT and 2 control) were taught by the same instructor. Each quarter the instructor taught one evening course and one daytime course. RMT was used in the daytime course during the winter and in the evening course during the spring.

Materials

A 106-item paper-and-pencil test was used to assess learning. The pretest was administered on the first day of class, and the same test was administered during the last class period. Students were given one hour to complete each test. The pretest/posttest included multiple-choice questions that corresponded to each of the topic modules, with approximately 21 questions per topic.

Procedure

On the first day of class, the pretest was given to students in both RMT and non-RMT classes. As each topic was covered in the course, students in the RMT classrooms were assigned a module to complete. Modules were completed in the following order: ethics, variables, reliability, validity, and experimental design. All RMT students used both the tutoring and CAI instructional conditions and were assigned to these instructional conditions in a counterbalanced order (students used one condition for three of the topics and one condition for the other two topics). There were equal numbers of students who used the tutoring and CAI conditions for each topic module.

In order to ensure that all students had similar course experiences, students in all five research methods classes were asked to register with the RMT system and install the RMT software at the beginning of the term. Most students did so successfully (106 of 136). Those who could not install the software were generally students who did not have access to a computer on which they could download software (i.e., they used on-campus computer labs). Students in the RMT classes who could not install the necessary software to run the agent version of the system were automatically assigned to the text-only presentation mode. Thus, assignment to presentation mode was non-random; students self-selected into a presentation mode. Students in non-RMT sections ceased to use the system after the registration and installation phase.

During the spring of 2007, additional control data was collected. These students were also enrolled in introductory research methods courses at DePaul and did not use the RMT system in conjunction with a course. All students took the pretest at the beginning of the term and the posttest at the end of the term.

Results

The primary question we investigated was: Do students who use the RMT system show higher learning gains from pretest to posttest than students who do not use RMT? In addition, we were interested in two secondary questions: 1) Do learning gain differences exist between those using the tutoring and CAI conditions? 2) Do learning gain differences exist between those using the agent and text-only presentation modes?

Before the analysis we excluded the data from any student who was not able to complete both the pretest and posttest. Overall ten students were eliminated from the RMT classes and one student was eliminated from the control classes, leaving 73 students in the RMT condition and 52 students in the control condition. In order to investigate our hypothesis that students who used RMT gained more at posttest than those who did not, we conducted an ANCOVA with gain score (posttest – pretest) as the dependent variable, pretest score as the covariate, and classroom condition (RMT versus control) as the independent variable. We found that RMT classes had significantly higher gain scores than control classes, $F(1, 122) = 17.24$, $p < .01$. RMT classes had an average gain of .109 (10.9 percentage point gain from pretest to posttest) and a standard deviation of .118, while the control classes showed an average gain of .03 (3 percentage point gain from pretest to posttest) and a standard deviation of .094. The NRP (National Reading Panel, 2000) effect size corresponding to this difference was .75 standard deviations, and

the eta-squared for the effect was $\eta^2 = .124$. This difference remained statistically significant when only the four sections taught by the same instructor were analyzed, $F(1, 94) = 5.99$, $p = .016$, NRP effect size = .49, $\eta^2 = .06$.

When the control data from 2007 was added, we again compared RMT classrooms (n = 73) to control classrooms (n = 85). Using an ANCOVA, we found additional support that RMT classes have significantly higher gain scores than non-RMT classes, $F(1, 155) = 23.21$, $p < .01$. RMT classes had an average gain of .109 (10.9 percentage point gain from pretest to posttest), while control classes had an average gain of .02 (2 percentage point gain from pretest to posttest). The NRP effect size corresponding to this difference was .76 standard deviations, and $\eta^2 = .13$.

The effectiveness of the tutoring condition (compared to the CAI condition) was evaluated by conducting a within-subjects comparison of gain scores in tutoring condition modules and CAI modules. A repeated-measures ANCOVA was conducted with condition gain score as the dependent variable, pretest score as the covariate, and instruction condition (tutoring versus CAI) as the independent variable. There was a significant overall difference in average gain score between tutoring and CAI versions of the tutor, $F(1, 71) = 4.627$, $p = .035$. The NRP effect size was .34 standard deviations, and $\eta^2 = .061$. Students had an average gain of .135 (13.5 percentage point gain from pretest to posttest) for modules in the tutoring condition, and .088 (8.8 percentage point gain from pretest to posttest) for modules in the CAI condition.

Finally, we examined the difference in learning gains between students who used the agent and those who used the text-only version of the system. Students who self-selected into the agent condition showed a marginally significant gain over students who self-selected into the text-only version, $F(1, 74) = 3.701$, $p = .058$, $\eta^2 = .048$. The agent

condition had a mean gain of .119 (11.9 percentage point gain from pretest to posttest) and the text-only condition had a mean gain of .06 (6 percentage point gain from pretest to posttest).

Discussion

During the winter and spring of 2006, RMT was assessed using five sections of introductory research methods. We found that the use of RMT resulted in higher learning gains than classroom instruction alone, with an overall NRP effect size of .71 standard deviations. This effect size is less than the 2.3 standard deviation maximum that Bloom reported for human tutors, but it ranks among the best results for ITS's. Although the effect size was not as great as the 1 standard deviation increase reported for AutoTutor (Graesser, et al., 2003), it is impressive evidence for the system's effectiveness, especially given three key differences between the RMT and AutoTutor assessments: 1) AutoTutor was evaluated in a laboratory setting, with the pre-test, two 2-hour tutoring sessions, and the post-test all conducted in a one week period. In our study, students in both the RMT and non-RMT conditions were also studying the subject matter in a regular course during a 10-week quarter. 2) RMT participants only used the system for 3-5 hours over the course of that quarter. 3) RMT was used by the students in their "natural environment" where they may have been distracted or may not have given RMT the full attention that they would have in a controlled lab setting.

In addition to the overall learning gains, we found evidence that the use of the dialogue-based version of the RMT system resulted in higher learning gains than the CAI version of the system. Topic modules in which students used the dialogue-based tutor had significantly higher learning gains than topics in which students used the CAI version of the system. Since the tutoring condition was much more interactive than the CAI

condition, the results align with previous research suggesting that the increased interaction in the tutoring condition facilitates learning (Wood, Wood, & Middleton, 1978; Graesser et al., 2003; Lane & VanLehn, 2005). Studies of this interaction hypothesis, however, are not conclusive, and recent research (VanLehn, Graesser, Jackson, Jordan, Olney, & Rosé, 2007) has suggested that interaction is most effective when students are learning material that is above their current preparation level (as when novices are learning material written for intermediates). It seems likely, then, that the dialogue-based tutor version of the RMT system is best suited for students who have little or no previous experience with research methods content. Students who are more advanced may benefit equally from use of the system and more traditional modes of learning.

In addition to the overall evaluation of the RMT system, we were able to investigate another aspect of the learning situation – the effect of an animated pedagogical agent. Our results indicated that students who used the agent yielded marginally significant higher scores than those who used the text-only version. Although this evidence should be interpreted with caution given that students self-selected into presentation modes (those who could not install the software were assigned to text-only), it is interesting in light of the mixed evidence in support of pedagogical agents (Moreno, 2004). Future studies will be necessary to clarify the role of the pedagogical agent in the RMT system.

As we continue to develop the RMT system, we plan to add elements to the existing modules and expand the current number of topics. The primary topical additions will involve the integration of research design and statistics. At most universities, these courses are taught separately, and many students find it difficult to associate research

design information with the appropriate statistical test. We are currently developing conceptual statistics modules that will address the application of statistical methods to research design, including data description, graphical representation of data, and various types of inferential statistical tests. We are also developing a module that addresses more complex experimental research designs.

In addition to integrating statistics and research design in the next generation of RMT, we plan to incorporate various tutoring styles. The current system uses a dialogue-based approach. We plan to supplement the dialogue-based approach with tabular presentation of problems which will require the student to solve a particular design problem in steps. As the student answers each question, he/she will begin "filling out" the table and can see his/her progress through the problem.

The initial classroom results from the investigation of the effectiveness of the RMT system have been encouraging. We believe that RMT has the potential to serve as an effective platform for the study of various issues in intelligent tutoring, while also supporting learning for students as they navigate more traditionally difficult subject matter in psychology.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 2, 167-207.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals.* Essex, England: Longman Group Limited.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 4-16.

Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25,* 471-533.

Clark, R., & Meyer, R. (2002). *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning.* Pfeiffer.

Corbett, A. T., & Anderson, J. R. (1991). Feedback control and learning to program with the CMU LISP tutor. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Graesser, A.C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology, 9*, 359–387.

Graesser, A. C., Jackson, G.T., Mathews, E.C., Mitchell, H.H., Olney, A., Ventura, M., Chipman, P., Franceschetti, D., Hu, X., Louwerse, M. M., Person, N. K., & the Tutoring Research Group. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Mahwah, NJ.

Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8,* 30-43.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25,* 259-284.

Lane, H. C., VanLehn, K. (2005). Teaching the tacit knowledge of programming to novices with natural language tutoring. *Computer Science Education, 15*(3), 183-201.

Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1992). Sherlock: A coached practice environment for an electronics troubleshooting job. In J. Larkin and R. Chabay, (Eds.), *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches,* 201-238. Hillsdale: Lawrence Erlbaum Associates.

Moreno, R. (2004). Animated pedagogical agents in educational technology. *Educational Technology, 44*, 6, 23-30.

Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction, Reports of the Subgroups. (NIH Publication No. 00-4754). Washington, DC: National Institute of Child Health & Human Development.

Salvucci, D., & Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* 923-928. Hillsdale: Lawrence Erlbaum Associates.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*, 3-62.

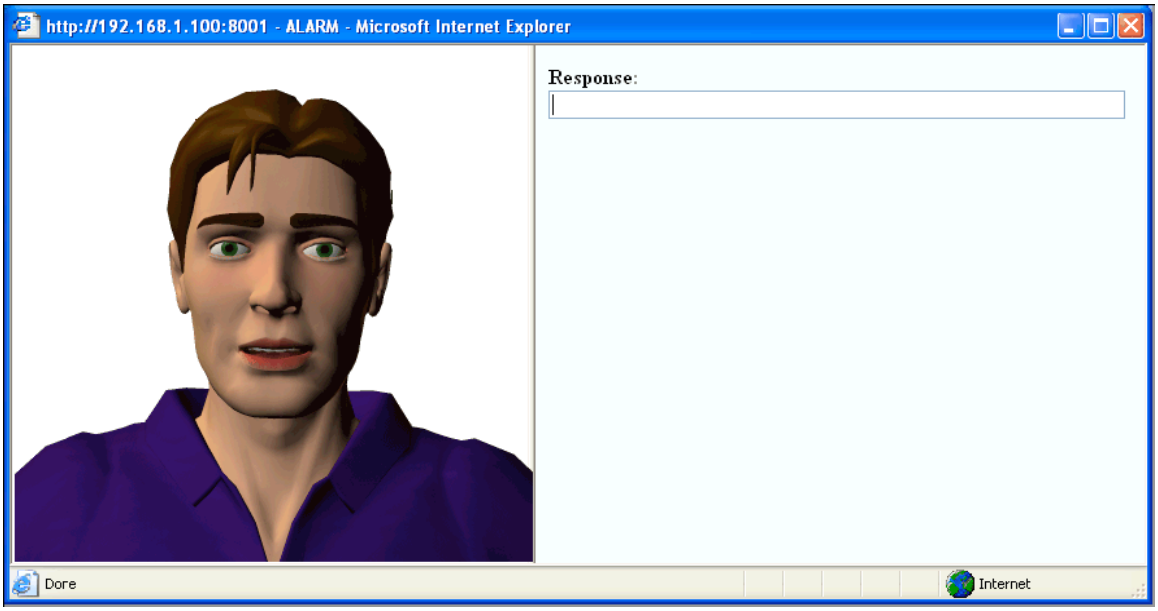
Wiemer-Hastings, P., Graesser, A., Harter, D., & the Tutoring Research Group. (1998). The foundations and architecture of AutoTutor. *Proceedings of the 4th International Conference on Intelligent Tutoring Systems*, 334–343. Berlin: Spring.

Wood, D. J., & Middleton, D. (1975). A study of assisted problem-solving. *British Journal of Psychology, 66,* 181-191.

Wood, D. J., Wood, H., & Middleton, D. (1978). An experimental evaluation of four face-to-face teaching strategies. *International Journal of Behavioral Development, 1,* 131-147.

Figure Captions

Figure 1. The animated pedagogical agent, Mr. Joshua.


http://192.168.1.100:8001 - ALARM - Microsoft Internet Explorer
Response:
Dore
Internet