Research Methods Tutor: Evaluation of a dialogue-based tutoring system in the classroom

ELIZABETH ARNOTT
Chicago State University, Chicago, Illinois

AND

PETER HASTINGS AND DAVID ALLBRITTON
DePaul University, Chicago, Illinois

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 nonequivalent control group design. Students in three classes ($n = 83$) used RMT, and students in two classes ($n = 53$) did not use RMT. Results indicated that the use of RMT yielded strong learning gains of 0.75 standard deviations above classroom instruction alone. Further, the dialogue-based tutoring condition of the system resulted in higher gains than did the textbook-style condition (CAI version) of the system. Future directions for RMT include the addition of new topics and tutoring elements.

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 discipline, many undergraduates struggle with research methods courses. Such courses tend to be more technical, quantitative, and applied than do other types of psychology courses. As is true of most college-level courses, time spent in such a class is rarely enough to provide students with sufficient practice, but unlike with other courses, research methods is not a subject students can learn without practice in applying their knowledge to research scenarios. Because the students are unlikely to encounter research scenarios in their everyday lives, they often lack the ability to sufficiently practice this skill. This article describes the evaluation of Research Methods Tutor (RMT), an intelligent tutoring system (ITS) 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 (*SD*s) 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 (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Wood & Middleton, 1975). Tutorial dialogues allow interaction between the tutor and student and, therefore, can yield a number of advantages over more traditional learning methods. Tutorial dialogue involves cooperation to solve a wide variety of problems (Graesser, Person, & Magliano, 1995). This cooperation can allow the tutor to assess the student's current level of knowledge and react appropriately 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 nontraditional students. Using an ITS avoids the practical disadvantages of one-on-one human tutoring. An ITS can provide some of the learning benefits of one-on-one human tutoring with little or no cost to the student, and it 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 those of a comparison class that did not use the tutor (Koedinger, Anderson, Hadley, & Mark, 1997). Dialogue-based ITSs support natural language interaction with students and can allow students to experience collaborative

E. Arnott, earnott@csu.edu

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 1 *SD* above those from reading a textbook alone (Graesser et al., 2003).

Description of the System

RMT is a dialogue-based ITS designed for use in conjunction with 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 compares responses and expected answers using latent semantic analysis (LSA) (Landauer, Foltz, & Laham, 1998), which creates a high-dimensional vector representation of both the expected answer and the student's response on the basis of 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 have a single correct answer ("What is an independent variable?"). Analytic questions require a student to not only know about concepts but 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 ("Design a study to test this hypothesis. What is the independent variable?").

RMT includes two instructional conditions for 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 . . ."), hints ("Internal validity is about finding what type of relationship between your independent variable and dependent variable?"), and 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 it presents the information as a monologue (similar to the experience of listening to someone read a textbook aloud) 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 via synthesized speech and a number of hand and facial gestures, including nodding and turning his head, blinking his eyes, and moving his hands.

The text-only presentation mode has no agent. The questions and responses of the tutor simply appear on-screen as text. Although the text-only version is much simpler technologically, it has been shown that in some situations, learners pay little attention to text presented on ITS screens (Salvucci & Anderson, 1998). In addition, when combined with additional figures, textual displays may visually overload the student and "short-circuit" visual processing (Clark & Mayer, 2003). Thus, a secondary goal of our research was to determine whether 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 hypothe-
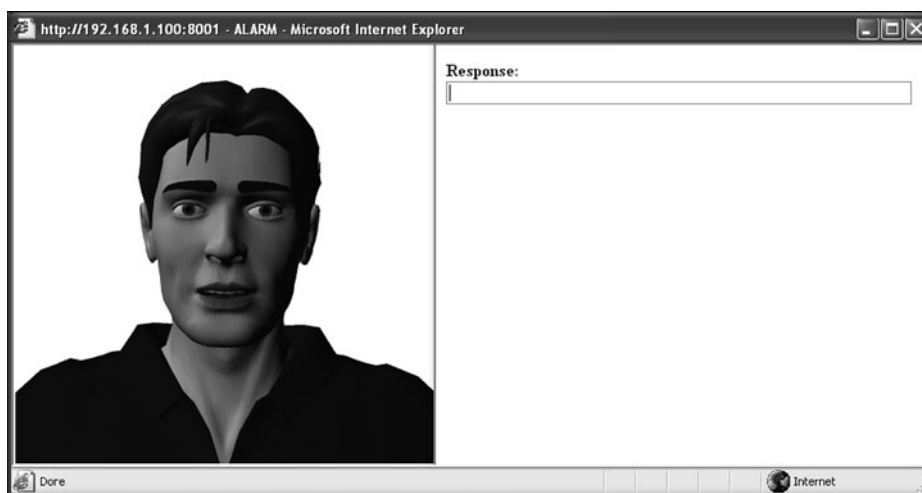

Figure 1. The animated pedagogical agent, Mr. Joshua.

sized that the classes that used RMT would exhibit greater learning than would classes that did not. We also assumed that greater interaction between the system and the student would result in increased learning, and, thus, we predicted that students would show greater evidence of learning in the tutoring condition than in the CAI condition. Finally, on the basis of 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, 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 nonequivalent control group. Four of the five courses (two RMT and two 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 1st day of class, and the same test was administered during the last class period. Students were given 1 h 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 1st 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 corresponding 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). There were equal numbers of students who used the tutoring and CAI conditions for each topic module.

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 to 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 not random; students self-selected into a presentation mode. Students in non-RMT sections stopped using the system after registration and installation.

During the spring of 2007, additional control data were collected. These students were also enrolled in introductory research methods courses at DePaul University 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 greater evidence of learn-ing than do students who do not? In addition, we were interested in two secondary questions: (1) Do learning differences exist between those using the tutoring and those using the CAI conditions? (2) Do learning differences exist between those using the agent and those using the text-only presentation modes?

Before the analysis, we excluded the data from any student who could not complete both the pretest and posttest. Ten students were eliminated from the RMT classes and 1 was eliminated from the control classes, leaving 73 students in the RMT condition and 52 students in the control condition (before the additional control data were added). To investigate our hypothesis that students who use RMT would have higher posttest scores than would those who do not, we conducted an ANCOVA with the posttest score as the dependent variable, the pretest score as the covariate, and the classroom condition (RMT vs. control) as the independent variable. We found that RMT classes scored significantly higher on posttest than did control classes [$F(1,122) = 17.24$, $p < .01$; $\eta^2 = .12$]. RMT classes had an average posttest score of 65.3% ($SD = 15.5$), whereas non-RMT classes had an average posttest score of 56.42% ($SD = 13.3$). The learning gain score (posttest score − pretest score) for the RMT classes was .109 (10.9% gain from pretest to posttest), with an SD of .118, whereas the control classes yielded an average gain of .03 (3% gain), with an SD of .094. The NRP (National Reading Panel, 2000) effect size corresponding to the posttest difference was 0.61 SDs, and the NRP effect size corresponding to the learning gain difference was 0.75 SDs. This difference remained statistically significant when only the four sections taught by the same instructor were analyzed [$F(1,94) = 5.99$, $p = .016$; $\eta^2 = .06$]. The NRP effect size for the posttest difference was 0.375, and the NRP effect size for the learning gain difference was 0.49.

When the control data from the spring of 2007 were added, we again compared RMT classrooms ($n = 73$) and control classrooms ($n = 85$). Using an ANCOVA, we found additional support that RMT classes have significantly higher posttest scores than do non-RMT classes [$F(1,155) = 23.21$, $p < .01$; $\eta^2 = .13$]. RMT classes had an average posttest score of 65.3% ($SD = 15.5$) and a mean learning gain of 0.109 (10.9% gain from pretest to posttest), whereas control classes had an average posttest score of 56% ($SD = 16.3$) and a mean gain of 0.02 (2% gain). The NRP effect size corresponding to the posttest score difference was 0.58, and the NRP effect size for the learning gain difference was 0.76 SDs.

The effectiveness of the tutoring condition (in comparison with the CAI condition) was evaluated by conducting a within-subjects comparison of scores in tutoring condition modules and CAI modules. There was a significant overall difference in average gain score between tutoring and CAI versions of the tutor [$F(1,72) = 8.178$, $p = .006$; $\eta^2 = .102$]. Students had average gains of .135 (13.5% gain from pretest to posttest), for modules in the tutoring condition, and .088 (8.8% gain), for modules in the CAI condition.

Finally, we examined the difference 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 [$F(1,70) = 2.84$, $p = .096$; $\eta^2 = .039$]. Those who used the agent had a mean posttest score of 65.9% ($SD = 15.6$) and an average learning gain of 0.123 (12.3% gain from pretest to posttest). Those who used the text-only condition had a mean posttest score of 63% ($SD = 15.5$) and an average learning gain of 0.062 (6.2% gain).

DISCUSSION

RMT was assessed using five sections of introductory research methods. We found that the use of RMT resulted in greater evidence of learning than did classroom instruction alone, with an overall NRP learning gain effect size of 0.75 SDs. This effect size is less than the 2.3 SDs maximum that Bloom (1984) reported for human tutors, but it ranks among the best results for ITSs. Although not as large as the 1 SD learning gain increase reported for Auto-Tutor (Graesser et al., 2003), it is impressive evidence for the system's effectiveness, especially given the differences between the RMT and AutoTutor assessments. AutoTutor was evaluated in a laboratory setting, with the pretest, two 2-h tutoring sessions, and the posttest all conducted in a 1-week period. Our study differed in three key ways: (1) Students in both the RMT and non-RMT conditions were studying the subject matter in a regular course during a 10-week quarter, (2) RMT participants used the system for only 3–5 h over the course of that quarter, and (3) students used RMT in their "natural environment," where they may have been distracted or may not have given RMT the full attention that they would have given it in a controlled lab setting.

In addition to the overall learning differences observed, we found evidence that the use of the dialogue-based version of the RMT system resulted in greater learning than did the CAI version of the system. Topic modules in which students used the dialogue-based tutor yielded significantly improved scores from pretest to posttest over those topics in which students used the CAI version. Because 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 (Graesser et al., 2003; Lane & VanLehn, 2005; Wood, Wood, & Middleton, 1978). Studies of this interaction hypothesis, however, are not conclusive, and recent research (VanLehn et al., 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. More advanced students who have previous course experience or are enrolled in an intermediate or advanced research methods course may benefit equally from use of the system and from 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 indicate that students who used the agent yielded marginally significant higher scores than did 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 (see Moreno, 2004, for a review). The majority of pedagogical agent studies, however, involve minimal agent exposure and rather simplistic laboratory settings. Thus, results from a semester of RMT usage may add richness to the existing agent data. 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 to expand the 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 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, which we plan to supplement with tabular presentation of problems that 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 will be able to 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 matters in psychology.

REFERENCES

ANDERSON, J. R., CORBETT, A. T., KOEDINGER, K. R., & PELLETIER, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, **4**, 167-207.

BLOOM, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Essex, U.K.: Longman.

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. A., JEONG, H., YAMAUCHI, T., & HAUSMANN, R. G. (2001). Learning from human tutoring. *Cognitive Science*, **25**, 471-533.

CLARK, R. C., & MAYER, R. E. (2003). *e-Learning and the science of*

instruction: Proven guidelines for consumers and designers of multimedia learning. San Francisco: Pfeiffer.

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

GRAESSER, A. C., JACKSON, G. T., MATHEWS, E. C., MITCHELL, H. H., OLNEY, A., VENTURA, M., ET AL. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In R. Alterman & D. Kirsh (Eds.), Proceedings of the 25th Annual Meeting of the Cognitive Science Society (pp. 474-479). Mahwah, NJ: Erlbaum.

GRAESSER, A. C., PERSON, N. K., & MAGLIANO, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied Cognitive Psychology, 9, 495-522.

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, 183-201.

LESGOLD, A., LAJOIE, S., BUNZO, M., & EGGAN, G. (1992). Sherlock: A coached practice environment for an electronics troubleshooting job. In J. H. Larkin & R. W. Chabay (Eds.), Computer assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches (pp. 201-238). Hillsdale, NJ: Erlbaum.

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

NATIONAL READING PANEL (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4754). Washington, DC: National Institute of Child Health & Human Development.

SALVUCCI, D. D., & ANDERSON, J. R. (1998). Tracing eye movement protocols with cognitive process models. In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 923-928). Hillsdale, NJ: Erlbaum.

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. C., HARTER, D., & THE TUTORING RESEARCH GROUP (1998). The foundations and architecture of AutoTutor. In Proceedings of the 4th International Conference on Intelligent Tutoring Systems (pp. 334-343). Berlin: Springer.

WOOD, D., & MIDDLETON, D. (1975). A study of assisted problem-solving. British Journal of Psychology, 66, 181-191.

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