RMT in the Classroom

Elizabeth Arnott
Department of Psychology
Chicago State University
earnott@csu.edu

Peter Hastings
School of Computer Science, Telecommunications, and Information Systems
DePaul University
peterwh@cs.depaul.edu

David Allbritton
Department of Psychology
DePaul University
dallbrit@depaul.edu

Abstract

Research Methods Tutor (RMT) is a dialog-based intelligent tutoring system designed for use in conjunction with courses in psychology research methods. The current RMT system includes five topic sections: ethics, variables, reliability, validity, and experimental design. The tutor can be used in an "agent mode," which utilizes synthesized speech and an interactive pedagogical agent, or in a "text-only mode," which presents the tutor content in written text on the screen. The tutor was used in three psychology research methods courses during the winter and spring quarters of 2006 at DePaul University. These three sections were evaluated against two (non-equivalent) control sections that did not use the RMT system. Pretest and posttest scores on a research methods knowledge test were used to assess learning in each class. Results indicated that, compared with the two control sections, the classes that used RMT showed significantly higher learning gains. In addition, those that used the agent version of the tutor showed significantly higher learning gains than those who used the text-only version of the tutor. Future directions will focus on expanding the current RMT content to include conceptual statistics and more complex research designs and to identify subgroups of students for which RMT may be particularly useful.

Introduction

A course in research methods, a requirement for psychology majors at most universities, tends to be difficult for students to navigate, both due to its technical, "hands-on" nature and its marked differences from other types of psychology courses. Like most courses, time spent in class is rarely enough to provide the students with sufficient practice, but unlike other psychology 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 life, they often lack the ability to sufficiently practice this skill. This paper evaluates Research Methods Tutor (RMT), an intelligent tutoring system that engages students in one-on-one discussions on a range of current topics from their research methods course.

There is considerable evidence for the effectiveness of one-on-one tutoring. For example, studies of tutored students have shown that they can achieve learning gains up to 2.3 standard deviations above classroom education alone (Bloom, 1984). Why is tutoring so effective? Tutoring can provide a much richer learning environment than a classroom experience. Effective tutors can continuously assess student progress (Anderson, Corbett, Koedinger, & Pelletier, 1995), react to that changing level of knowledge, and model appropriate problem solving strategies when the student cannot generate them on his or her own (Lesgold, Lajoie, Bunzo, & Eggan, 1992). A number of studies have identified strategies that effective human tutors may use. Graesser, Person, and Magliano (1995) asserted that the key to human tutoring success is the considerable amount of time spent cooperating to solve a wide range of problems. Tutoring sessions often consist of tutors modeling worked examples for the students. Given a choice, learners opt for these types of worked examples in lieu of verbal descriptions (Anderson, Farrell, & Sauers, 1984; LeFevre & Dixon, 1986). Human tutors also give feedback that allows the student to assess his or her progress. This feedback is immediate, which leads to decreased learning time necessary for concept mastery (Corbett & Anderson, 1991).

Although tutoring has marked advantages, it is impractical for many students due to its cost and potential inconvenience, especially at institutions which attract non-traditional students. Intelligent Tutoring Systems (ITSs) 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, &

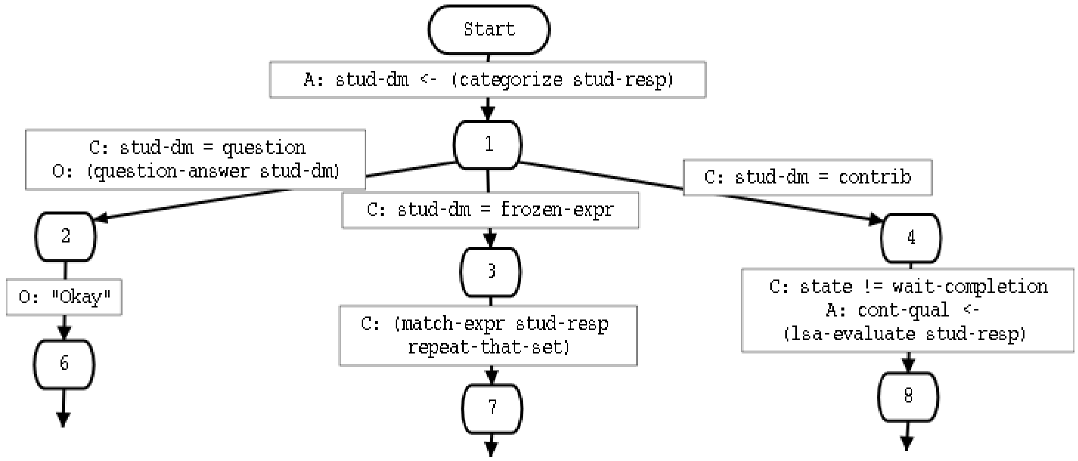

Figure 1. A Partial Dialog Advancer Network

Mark, 1997). Dialog-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. One dialog-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, & TRG, 2003).

Description of the RMT System

Research Methods Tutor (RMT) is a dialog-based tutor that aids students in learning concepts from psychology research methods. With a similar tutoring style to that of AutoTutor, RMT engages students in a question-and-answer dialog and evaluates answers by comparing them to a set of expected answers (Wiemer-Hastings, Graesser, Harter, & the Tutoring Research Group, 1998). RMT is a web-based system, so students can use the tutor when it is most convenient for them.

The tutor's behavior is controlled by a Dialog Advancer Network (DAN). The DAN is a transition network where each arc is associated with conditions, actions, and outputs. A partial DAN is shown in Figure 2. The DAN allows RMT to respond appropriately to many different types of student inputs. Each path through the DAN starts by categorizing the current student input, creates a response to it, and then creates a follow-up utterance to keep the dialog moving. For example, if a student asks the tutor to repeat what he just said, the DAN recognizes the request, outputs "Once again," and copies the previous tutor utterance. If the student has just provided an answer to a question, the DAN will evaluate the answer, provide positive or negative feedback as appropriate, and then present a follow-up prompt or hint, or, if the current question has been thoroughly answered, a summary of the complete response. The DAN mechanism provides a measure of flexibility to the RMT system. The behavior of the tutor can be altered by simply modifying the network.

As mentioned above, RMT evaluates student answers by comparing them to a set of expected answers. Three major components are responsible for the comparison. The first is an automatic spelling correction module. If any word that the student enters is not in RMT's lexicon, then aspell is called to provide a set of possibilities. RMT selects the most likely respelling from its lexicon. The second component is a keyword matcher, which checks for literal similarity of strings. This is especially useful for the shorter answers. The third component is Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997). LSA creates a high-dimensional vector representation for each content word from a corpus of domain-related texts. The vector for a student answer is created by combining the LSA vectors of the words in the answer. Vectors for expected answers are created in the same way. The cosine function measures the similarity of the vectors.

Currently RMT has five topic modules that correspond to traditional topics in introductory undergraduate research methods courses – ethics, variables, reliability, validity, and experimental design. Following Bloom's taxonomy (1956), each topic contains a mix of conceptual, analytic, and synthetic questions. Conceptual questions are traditional textbook questions with a single correct answer (for example, "What is the difference between validity and reliability?"). Analytic questions require the students to apply their conceptual knowledge to a particular scenario (for example, "What threats to validity may be a problem for this study?"). Synthetic questions require a higher level of understanding of the material that allows the students to construct an entire solution for a scenario (for example, "I want to know if frustration causes aggression, can you design an experiment to address this issue?").

In order to make comparisons between dialog-based tutoring and a more textbook-style approach, the system includes two instruction conditions: a computer-aided

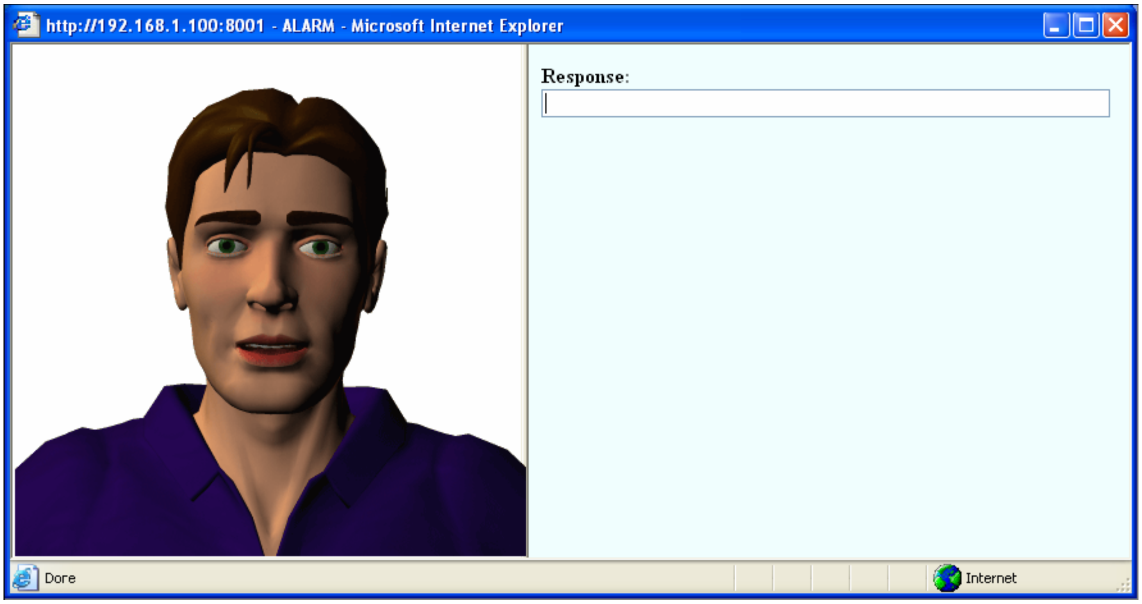

Figure 2. Mr. Joshua and the Basic RMT Interface (Tutor/Agent Mode).

instruction (CAI) condition and a tutor condition. The CAI condition consists of textbook-style passages of information that are presented sequentially to the student. To ensure that the topic material was read, the student answers brief multiple choice questions about the topic. In the tutoring condition, RMT engages in a dialog with the student about the topic. The tutor asks the student questions, and the student types answers in a response box. RMT uses latent semantic analysis (LSA) to process the student answers and respond appropriately. If the student answers incorrectly, RMT avoids giving negative feedback (as expert human tutors often do – Person, 1994) and engages in a series of prompts and hints to help the student arrive at the correct answer. Prompts are sentence-completion items that can be answered with a short phrase ("Informed consent means that you obtain the participant's consent without…."). Hints are questions or statements that help the learner arrive at the correct answer by soliciting a sentential answer to a smaller sub-question. After each question has been successfully navigated, RMT summarizes the key elements of the problem and moves on to the next question.

The tutor also has two presentation modes: a text-only mode and an animated agent mode. The text-only mode consists of a textual display of all of the questions, feedback, prompts, and hints, which the students read on screen. The animated agent mode, in contrast, features a male "talking head" named "Mr. Joshua" (see Figure 1). At this time Mr. Joshua gestures in a number of human-like ways, including turning, nodding, and shaking his head (to indicate agreement or disagreement), gesturing with his hands, raising his eyebrows, and blinking his eyes. The agent is implemented using Microsoft Agent software and "speaks" using a text-to-speech engine.

Although the agent makes the system more visually appealing, the purpose of its inclusion is to assess any potential increase in learning that an agent may provide. Some evidence suggests that an automated speaking agent may aid learning by keeping the visual channels free for assessing other content. Consistent with Mayer's (2001) cognitive theory of multimedia learning, Clark and Mayer, (2002) found that textual displays combined with additional figures may visually overload the student and "short circuit" visual processing. In addition, Salvucci & Anderson (1998) suggested that learners pay little attention to text presented on screen in intelligent tutoring settings. However, there is also evidence to suggest that an nimated, speaking agent may have negative affects on learning. When an animated agent is presented, the visual stimulus of the agent itself may distract the learner and be disadvantageous to learning (Moreno, 2004). This effect is especially problematic if the student is required to evaluate a visual stimulus onscreen while the agent is present. Although the current materials do not rely heavily on additional graphics, a secondary goal of our research is to determine whether the presence of an agent aids in learning in this particular intelligent tutoring situation.

Method: Evaluating RMT in the Classroom

The primary goal of the present study was to investigate the effectiveness of RMT in the "real world," i.e. not in a laboratory setting, but in combination with an actual course. We hypothesized that the classes that used RMT in conjunction with their traditional curriculum would show larger learning increases than classes that did not use RMT. In addition, we investigated learning differences between the dialog-based tutoring and textbook-based (CAI) conditions, as well as differences in agent and text presentation modes.

Participants

During the winter and spring of 2006, 136 participants were included in the evaluation of RMT in the classroom. All participants were students enrolled in an introductory undergraduate research methods course in psychology at DePaul University. Three of the course sections used RMT as part of the course requirements, and two of the sections acted as non-equivalent control groups. Four of the five courses (two control and two RMT) were taught by the same instructor, and the courses were evenly split between daytime and evening sections. In total, there were 83 participants in the RMT sections and 53 participants in the control sections.

Materials and Design

A 106-item multiple choice test was created for use as a pre- and post-test. The test took approximately 1 hour to complete. Items were categorized according to the topic to which they applied (ethics, variables, reliability, validity, or experimental design), with two questions corresponding to more than one topic. There were 20 ethics questions, 25 variables questions, 20 reliability questions, 23 validity questions, and 20 experimental design questions.

The students were assigned according to their enrollment in the RMT or non-RMT sections. Thus assignment to the RMT vs. baseline control condition was between-subjects. In addition, students in the RMT courses completed their RMT modules in one of two instruction conditions: the tutor condition or the CAI control condition. Each RMT student saw two of the topics in one condition and three of the topic in the other condition. This allowed for comparison between not only the RMT and control participants across classes, but also between the tutor and CAI conditions.

Finally, students in all courses were asked to install the relevant software on their personal computers. RMT students who could not successfully install the software on a computer were assigned to the text-only presentation mode. All other RMT students were assigned to the agent presentation mode. Thus, RMT students self-selected into the agent or text-only presentation modes.

Procedure

On the first day of class, students in all evaluated sections were given the pretest. In order to ensure that the students did not differ markedly in their ability to use computer technology or their access to computers, all students were given RMT registration and installation instructions and were asked to install the RMT software on their personal computers. Most students did so successfully (109 of the 136 total students). As mentioned above, students in the RMT courses who could not install the software were assigned to the text-only presentation mode. Throughout the quarter, the RMT students were assigned to complete the topic modules as the topics were covered in class. On the last day of classes students in all sections were asked to complete the 106-item posttest.

Results: Evaluating RMT in the Classroom

Three primary questions were investigated: 1) Do classes that use RMT show higher learning gains than non-equivalent control classes? 2) Are there differences between instruction conditions (tutor vs. CAI) for the individual topic modules? 3) Are there differences between the agent and text-only presentation modes?

If a student was unable to complete the posttest or pretest, his/her data was excluded from the analysis. Six students were excluded from an RMT section because of an incomplete pretest or posttest, and one student was excluded from a control section. After excluding these participants, there were 77 participants in the RMT sections and 52 participants in the control sections.

Our initial hypothesis was that the learning gains for classes that used RMT would be higher overall than learning gains for classes that did not use RMT. This hypothesis was confirmed using an ANCOVA with overall gain scores (percent correct on posttest minus percent correct on pretest) as the dependent variable, the pretest score as the covariate, and class condition (RMT class versus control class) as the independent variable. As predicted, we found a significant difference between RMT courses and non-RMT courses, $F(1, 126) = 15.154$, $p < .01$. The effect size corresponding to this difference was .71 standard deviations, which was calculated using the National Reading Panel (NRP, 2000, p. 15) standardized mean difference formula: (treatmentMean - controlMean) / (0.5 * (treatmentStdDev + controlStdDev)). The eta-squared was $\eta^2 = .11$. Although we expected all classes to show some gains (since they had been enrolled in the relevant course for a quarter), those who used the tutor had

	Ethics	Variables	Reliability	Validity	Experimental Design
Tutor	.106	.201*	.107*	.126*	.102*
CAI	.09	.151#	.077	.168#	.079#
Control	.071	.077*#	.021*	.044*#	-.038*#

**Indicates a significant difference at the .05 level.*
#Indicates a significant difference at the .05 level.

Table 1. Mean Gain Scores by Instruction Condition for each Tutor Topic Module.

an average gain of .105 (10.5 percentage points from pretest to posttest), and those who did not use the tutor had an average gain of .03 (3 percentage points from pretest to posttest). This difference remained statistically significant when only the four sections taught by the same instructor were analyzed, F (1, 98) = 4.99, $p = .028$, NRP effect size = .46, $\eta^2 = .05$.

When we broke the data down by individual class we found that two of the three classroom sections that used RMT had higher overall gain scores than the control sections. Using an ANCOVA with gain score as the dependent variable, pretest score as the covariate, and section as the independent variable, there was a significant difference overall among courses, F (4, 123) = 10.852, $p < .01$, $\eta^2 = .26$. Two of three sections using RMT had significantly higher gain scores than the sections that did not use RMT. These two sections did not differ significantly from one another, but did differ significantly from the third RMT section. Similarly, the non-RMT classes did not differ significantly from one another. Gain scores for RMT classes were .1263, .1550, and .0117. Gain scores for non-RMT classes were .0142 and .0476. Because the pretest/posttest was broken down into questions for each individual topic module, we were also able to investigate differences among instruction conditions for each topic. For each of the five RMT modules, student gain scores were analyzed in an ANCOVA with instructional condition (control class versus CAI mode versus tutor mode) as the between-subjects factor and pretest score as the covariate. (Although each RMT student completed some modules in tutor mode and some in CAI mode, instructional condition was manipulated between-subjects when considering each module separately.) If a student in the control group had completed any topics in a module, then that student's data were excluded from the analysis of that module. Likewise, any RMT student who did not complete any topics for that module was also excluded.

There was a significant overall difference among instruction conditions for all of the modules except the ethics module. For the other modules there was a significant difference between tutor and control, and for variables, validity, and experimental design there was also a significant difference between CAI and control. Although the means were in the predicted direction for all topic modules but validity (tutor followed by CAI followed by control), there was not a significant difference between tutor and CAI for any of the modules. Means by condition for each module are found in Table 1.

Lastly, we tested the hypothesis that the agent presentation mode would produce higher learning gains than the text-only presentation mode or the control condition. Using an ANCOVA with gain score as the dependent variable, pretest score as the covariate, and presentation mode (agent versus text-only versus control) as the between-subjects factor, this hypothesis was supported, F (2, 125) = 9.924, $p < .01$, $\eta^2 = .14$. The agent condition had a mean gain of .119, the text condition had a mean gain of .06, and the control condition had a mean gain of .03. Using LSD paired comparisons, the agent mode was associated with significantly higher gain scores than both text-only (MD = .06, SE = .029) and control conditions (MD = .09, SE = .02). The text condition was not significantly different from the control condition. This effect remained when only the four instructor-consistent sections were analyzed, F (2, 97) = 4.543, $p = .013$, $\eta^2 = .09$. (For these sections, the agent condition had a mean gain of .094, the text condition had a mean gain of .028, and the control condition had a mean gain of .03.)

Implications and Future Directions for RMT

Five courses at DePaul University were evaluated during the 2005-2006 academic year. Of these five courses, three used the RMT system and two acted as non-equivalent control groups. It was hypothesized that the use of RMT would result in higher learning gains on the pretest/posttest measure. This hypothesis was confirmed, with RMT classes achieving an average gain of .71 standard deviations over the control classes. This result is close to the 1 standard deviation gain obtained by students using the AutoTutor system during intensive lab-based tutoring sessions (Graesser, Jackson, Mathews, Mitchell, Olney, Ventura, Chipman, Franceschetti, Hu, Louwerse, Person, & TRG, 2003). As RMT was used in a naturalistic environment with students who all interacted with the material to some extent (all students – even those in the control group – were enrolled in a research methods course), we believe that this gain is impressive evidence of the effectiveness of the system. Although the evidence for the advantage of the system's tutor version over the CAI version was weaker, we believe that the overall evidence supports the effectiveness of the system.

In addition to evaluating the system itself, the results obtained from using RMT in the classroom have allowed us to investigate other issues in intelligent tutoring. Although students in this particular study self-selected into a presentation mode condition, we believe that the advantages seen in the agent condition provide evidence that the presence of a tutoring agent may aid in the learning process in our particular situation. Given the mixed nature of the evidence for the effectiveness of animated pedagogical agents (Morena, 2004), this initial finding provides an interesting issue for further study, particularly as we add more visual elements to the system. It is possible that the advantages seen in the present study were affected by the presence of synthesized speech and few additional (non-agent) visual stimuli. This issue will be an avenue for future study.

During the next phase of RMT development, we plan to add topic modules that will aid students as they attempt to integrate research methods and statistics. At most universities, these courses are taught separately, and many students find it difficult to understand the close connection between them. We are currently developing a conceptual statistics module that will address the application of statistical methods to particular types of research design. We are also developing a module that addresses more complex research designs. In addition, we plan to add graphical elements to RMT by creating graphics for our current content and forming a data description and graphing module, which will help students to display the results of their studies visually and understand the types of graphs that are appropriate under given conditions.

In addition to integrating statistics and research design in the next generation of RMT, we also plan to add elements which will incorporate various tutoring styles. We plan to augment our current dialog-based tutor with table-style problems which 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.

As RMT continues to develop, we are encouraged by our classroom results, and believe that RMT can be of value, not only in investigating the effectiveness of intelligent tutoring strategies and design features, but also in aiding students as they navigate more difficult research design and statistical issues.

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.

Anderson, J. R., Farrell, R., & Sauers, R. (1984). How subject matter knowledge affects recall and interest. *American Educational Research Journal, 31,* 313-337.

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.

Clark, R., & Mayer, 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., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

LeFevre, J. A., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction, 3,* 1-30.

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.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, England: Cambridge University Press.

Moreno, R. (2004). Animated pedagogical agents in educational technology. *Educational Technology, 44*, 6, 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, Reports of the Subgroups. (NIH Publication No. 00-4754). Washington, DC: National Institute of Child Health & Human Development.

Person, N. K. (1994). An analysis of the examples that tutors generate during naturalistic one-to-one tutoring sessions. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 55*, 10-B, 4628.

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.

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