Initial results and mixed directions for Research Methods Tutor [1]

Peter Wiemer-Hastings [a,2], Elizabeth Arnott [b] and David Allbritton [b]
[a] *DePaul University School of Computer Science, Telecommunications and Information Systems*
[b] *DePaul University Department of Psychology*

Abstract. RMT (Research Methods Tutor) is a dialog-based tutoring system currently being used in conjunction with university-level courses on research methods in psychology. RMT has a web-based interface and uses a talking head to to present its dialog acts to the student. The student types in unconstrained natural language responses. Although RMT currently focuses on natural language-based interaction with students, the use of the talking head brings in other modalities of interaction which must be integrated with the textual "message". This paper describes the basic architecture and approach of the RMT system, and its evaluation in the context of recent research methods classes. Our experiments compared the agent-based tutor to a text-only version of the system, and compared tutoring to a computer-aided instruction control. Due to technical difficulties with the agent software and some compliance issues with the students, we got no significant results that validated the usefulness of the system as a learning aid. We did however see consistent trends of additional learning in conjunction with the use of the tutor that warrant further investigation. This paper concludes by describing our current efforts for integrating graphical visual aids with dialog-based tutoring.

Keywords. Dialog-based tutoring, Interactive pedagogical agent / talking head, Graphics and text

1. Introduction

In the past decade, advances in natural language processing techniques have made it possible to create intelligent tutoring systems which interact with students more and more like human tutors do: via dialog about some content material. In some cases this interaction is appropriately entirely linguistic. In most cases, however, tutor-student interactions can and should include other types of shared information. This paper describes Research Methods Tutor (RMT), a dialog-based intelligent tutoring system for research methods in psychology. RMT is intended as an adjunct to a college-level course to be used by the student at their convenience to strengthen their understanding of the concepts discussed in class.

---

[1]This work was supported by a grant to the first author from DePaul University's Quality of Instruction Council.

[2]Correspondence to: Peter Wiemer-Hastings, School of Computer Science, Telecommunications and Information Systems, 243 South Wabash Avenue, Chicago, IL 60604, United States. Tel.: +1 312 362 5736; Fax: +1 312 362 6116; E-mail: peterwh@cs.depaul.edu.

RMT is embodied by an animated agent which uses text synthesis to present its questions, feedback, and other dialog moves. The tutor can also operate in text-only mode where its utterances are printed on the screen. The latter approach is much simpler from a technological viewpoint. On the other hand, there is evidence that students may pay scant attention to textual information in an ITS setting [6]. Furthermore, if graphics or other visual aids are presented simultaneously with explanatory text, the student's cognitive and perceptual resources for processing visual input may become overburdened [2]. In addition, the use of a "talking head" provides the possibility for utilizing another type of "mixed language", namely the intonation and facial gestures which human tutors often use to give graded feedback.

The RMT project has 3 major goals:

1. to explore the requirements for and benefits of a web-based tutoring system as an adjunct to classroom education,
2. to learn more about tutoring in an abstract, relatively informal domain as compared to science or math-oriented tutoring, and
3. to serve as a workbench for evaluating different aspects of dialog-based tutoring, for example how the use of a talking head might help students learn when compared to text alone.

This paper gives a high-level description of the RMT architecture and its approach to tutoring. Next, we present the results of our initial evaluations of the system with students in research methods courses at DePaul University. Finally, we discuss the directions which we are currently exploring for incorporating mixed-language explanations to help improve student learning.

2. RMT Basics

Psychology students at DePaul, as at most universities, are required to take one or more courses in research methods. These courses address the basic methodology required to do experimental psychology, including topics like variables, reliability and validity, different types of experimental designs, and ethics. Although the courses normally include many examples of aspects of experimental design, it is not the specific examples that the students must learn, but rather the processes for creating successful designs. This level of abstraction makes the courses especially difficult. When students interact with RMT, they have the opportunity to discuss additional examples and get direct feedback on their understanding and use of the relevant concepts. From a constructivist viewpoint, this active processing should increase the number of connections that they can make to related material and deepen their understanding.

RMT is a web-based system. Students log in to the system at their own convenience, typically during a window of time set by the instructor for each topic module, and typically from their own computers at home. The agent is implemented using Microsoft Agent software with its accompanying text-to-speech engine. The tutoring sessions start when the student has chosen one of the available topics. The tutor is normally "in control" of the dialog. The tutor asks questions, and the student types in his or her responses. The tutor evaluates each response using Latent Semantic Analysis [4,7] by comparing it to a set of one or more expected answers. The tutor then gives positive or negative feed-

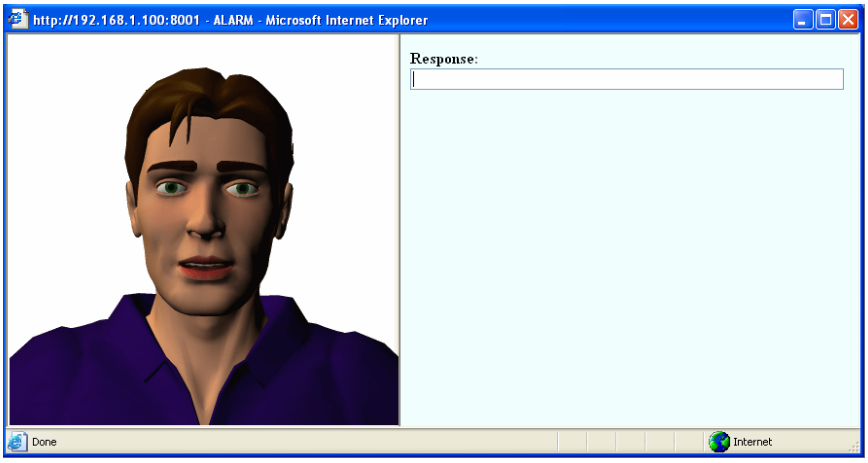

Figure 1. The basic RMT interface

back and either tries to get the student to add more information in response to the current question, or summarizes and moves on to a new question. RMT uses a transition network to control its dialog decisions, including responding to questions and certain requests or commands (e.g. asking the tutor to repeat its last utterance). All tutor and student utterances are logged in a database along with the tutor's evaluations of student response quality. Figure 1 shows a screenshot of the (agent-based) tutoring interface. The interface was intentionally kept as simple as possible, avoiding anything that might distract the student from the tutoring interaction.

Currently RMT uses entirely linguistic interactions with the student. For most tutoring items, the tutor simply asks a question, evaluates the the student's response, and follows with another question or prompt to keep the dialog moving. Even within this limited paradigm, there is an aspect of mixed language. As human tutors do [5], RMT avoids the use of direct negative feedback which might cause the student to become embarrassed or "lose face" and subsequently reduce their level of participation in the dialog. Instead RMT uses intonation and facial and hand gestures to suggest that something is not quite right without explicitly telling the student that they're wrong. Thus the use of an animated agent automatically brings in multiple modes of communication. In RMT, these modes are coordinated by its feedback mechanism which is invoked with an utterance and an affective stance like "positive" or "confused." The feedback mechanism will then choose an intonation pattern and animation to express this affect. A random component in the feedback mechanism keeps the agent from acting too "robot-like".

RMT has three different levels of tutorial material: conceptual, analytic, and synthetic [1] which roughly correspond to the student learning the basic concepts, how to apply them to analyze example experimental designs, and how to create new experimental designs. Many of the analytic and synthetic items include a fairly detailed description of a particular experimental scenario which is then referred to over the course of an extended discussion. It would be unreasonable to expect the students to maintain the details of these examples in their working memory while the discussion goes on. To provide an external memory device, RMT always presents these scenarios on the screen as shown in Figure 2 during the related dialog. The tutor refers to the scenario when asking the initial

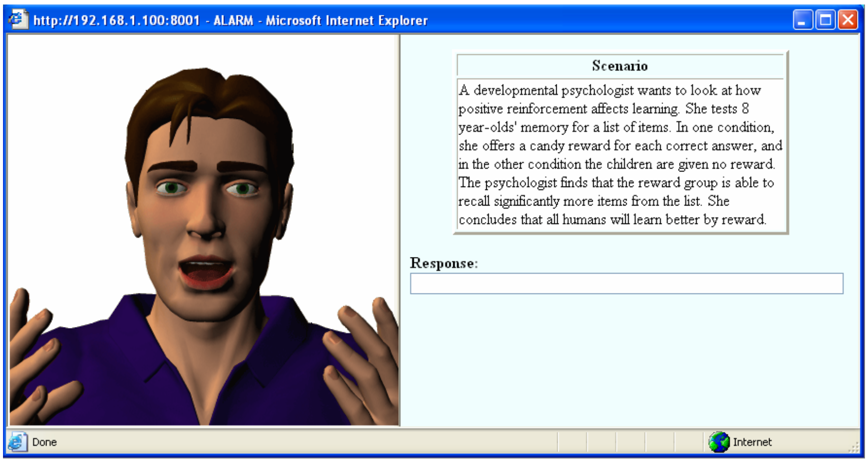

Figure 2. RMT with a scenario

question and gives the student the opportunity to read the scenario while coming up with their answer.

This relatively simple architecture allows RMT to engage students in extended conversations about a wide range of topics in research methods. We have currently developed modules for five topics: validity, reliability, variables, experimental design, and ethics. The next section discusses our initial evaluations of the RMT in conjunction with research methods courses at DePaul.

3. Evaluation

The motivation behind our goal of integrating RMT with classroom education was to enable us to fill a gap in the current understanding of dialog-based tutoring. Previously such systems have been evaluated in the context of laboratory experiments where participants take a pre-test, use the system intensively for a significant amount of time, then take a post-test [3, for example]. This protocol certainly highlights the potential effectiveness of these systems, but may be lacking in external validity. The fundamental question is whether such systems can help students learn on an ongoing basis. Our “real-world”, class-linked evaluation of an ITS is challenging since the students are using the tutor for significantly less time than they are sitting in class throughout the course of a term. Furthermore, as we describe below, other aspects of the real world have complicated our evaluations. Nevertheless, our results were encouraging if not conclusive.

Pilot data from two classroom studies of the current version of RMT supported the feasibility and usefulness of our approach. During the Winter and Spring quarters of 2004, 101 students from multiple sections of Research Methods courses at DePaul University volunteered to use the tutor throughout the quarter for extra credit. A few of them actually did use the tutor for one or two sessions. One of our main research questions focussed on whether the use of an animated agent helped or hurt student learning when compared to the text-only version of the tutoring system. Unfortunately, because the agent-based modules required the installation of additional software, very few students actually completed these modules. Those who were able to install the software were greeted by an agent with no animations (except for blinking and lip movements) and a rather high, squeaky voice. Thus, we were not especially surprised when the agent version failed to help students learn the material.

In the Fall quarter of 2004, 28 students in a single section of Research Methods II (the second of the two-course methods sequence taken by all psychology majors at DePaul) who were required to use the tutor as a course assignment consented for their data to be analyzed for research purposes as well. The architecture of the system had been extended to support automatic and dynamic assignment of students to different conditions for the different modules. This second experiment tested the following hypotheses: First, that the RMT system would lead to greater learning gains than a computer assisted instruction (CAI) control[1]; and second, that an animated agent would be superior to text-only presentation.

Across the three classes in our sample (one section of Research Methods I, and two sections of Research Methods II) a total of 43 students completed both the pre-test and post-test and were thus included in the analysis (18 from RM I, and 25 from RM II). Of these, 15 were unable to install the agent software at all, leaving a relatively small number of observations per condition.

The design was a 2x2 within-subjects factorial, with presentation (text-only vs. agent) and instruction type (RMT tutor vs. CAI) as the independent variables. For each student, one module was assigned to each of the four conditions. Because there were five modules and four conditions, one of the conditions was assigned to two modules for each subject, with the pairing of modules and conditions counterbalanced across groups of four subjects.

Unfortunately, many subjects failed to complete one or more modules, with the result that the design was no longer counterbalanced across modules. We therefore analyzed each module separately, with pre-test to post-test gain score for proportion correct on questions related to that module as the primary dependent measure, and pre-test score included as a covariate. The design of the analysis of each module was therefore a 2x2 ANCOVA, with both presentation and instruction type as between-subject factors. Only students who had installed the agent software and had completed at least one session on that module were included in each analysis ($N = 12, 12, 26, 25$, and 24 for the validity, reliability, variables, experimental design, and ethics modules respectively). There were no reliable main effects or interactions in any of these analyses, probably due to the low Ns and variability associated with technical problems with the agent software reported by students. The only significant effect in the analyses was that of the pre-test covariate — not surprisingly, lower pretest scores were generally associated with higher gain scores.

Although one of our aims had been to compare the effectiveness of the agent to text-only presentation, our data was uninformative on this point, revealing no significant main effect of presentation type in any analysis, and no consistent qualitative pattern (three modules had higher covariate-adjusted mean gain scores for text-only presentation and two for agent presentation). Because of the lack of presentation effects, and the fact that over a third of our sample failed to install the agent software at all, we then collapsed across presentation type and re-analyzed the data, focusing instead on instruction condition.

[1]In the CAI control condition, students were presented with a series of static texts (in either text-only form or “read” by the animated agent) which covered the same general content material as the tutoring sessions, and they were given a multiple choice test at the end of each module to ensure that they attended to the text.

	None	CAI	Tutor
Validity	.05	.08	.22
Reliability	.07	.05	.13
Variables	.13	.08	.12
ExptDes	.00	.09	.13
Ethics	.08	.16	.19

Table 1. Gain scores by Module and Condition

In the re-analysis, we focused on two questions: (1) Is there any evidence that RMT produces greater learning gains than the CAI, or that either produces greater gains than those found for students who did not complete that module at all (none). (2) Does aptitude (as measured by overall pre-test scores, not separated by module) interact with instruction type? Our intuition based on student comments was that high-aptitude students may learn more effectively from the CAI, which allows them to control the pacing and presentation themselves, while lower-aptitude students may learn better from the tutor, which provides more active, tutor-directed instruction.

We re-analyzed the gain scores from all 43 students for each module in separate one-way ANCOVAs, with instruction condition (CAI, RMT, or none) as the between-subjects factor and module pretest score as the covariate. There was no significant effect of instruction in any of the analyses; again the only significant effects were those of the pretest covariate. Qualitatively, however, a fairly consistent pattern emerged. The mean covariate-adjusted gain score for the RMT condition was higher than that of the CAI condition for all five modules ($p < .05$ by a sign test) and higher than the baseline "none" condition for four of the five modules. The CAI condition was not consistently above baseline, however, with only two modules having a higher mean gain score for CAI than "none." Table 1 lists the average gain scores for the five modules by instruction condition.

This is consistent with the qualitative pattern of the overall unadjusted mean gain scores. Averaging across all modules, the mean gain scores were .10 for CAI, .15 for RMT, and .06 for uncompleted modules. Thus we have a trend that suggests that the use of RMT increases learning, although no solid statistical evidence from our limited sample.

To test our hypothesis of an aptitude by instruction-type interaction, we entered the module gain scores into separate 2x2 ANCOVAs, with instruction condition (CAI vs RMT) and aptitude (low vs. high, as defined by a median split of overall pretest scores) as between-subjects factors, and module-specific pretest score as the covariate. (Uncompleted modules were excluded from these analyses). The predicted aptitude by instruction interaction was not confirmed. There were no significant main effects or interactions in any of the analyses, with the only significant effects being those of the module pretest covariates (significant in 4 of the 5 analyses). Qualitative examination of the mean gain scores revealed no consistent pattern of interaction either. To make sure the lack of interaction was not simply an artifact of reducing the variability in the aptitude measure by making it categorical, we also conducted regression analyses with aptitude (overall pretest score), instruction type, and the aptitude*instruction type interaction as predictors. No significant interactions were found in any of these analyses either. Thus we have no evidence at present to support our hypothesis that the RMT tutor is more effective for low-aptitude students, compared to a CAI.

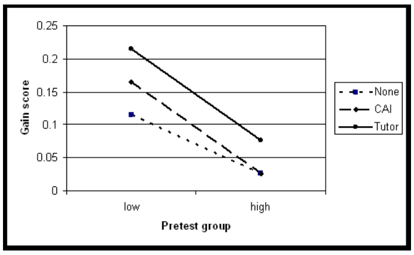

Figure 3. Average gain by instruction type and aptitude

Despite the fact that the differences were not significant, the data trends did come out in the right direction. Figure 3 shows the differences in gain scores for the low and high pretest groups for the different types of instruction. In this chart, the average gain scores for the modules which students did not complete (i.e. the learning gain due to the coursework alone) were also included for comparison and are marked as "None." Again, we make no claims based on this outcome other than that this question of an aptitude by treatment interaction in RMT warrants further study.

4. Mixed Language Explanations

Research on e-learning systems shows that including graphics with textual explanations can produce a cognitive synergy in the learner that leads to more effective processing of the material and deeper understanding [2]. Unfortunately in an abstract domain like research methods, it is relatively easy to create graphics that illustrate a particular example (e.g. participants in an experiment drinking coffee before taking a test), but such graphics will do little to improve the student's far transfer. It is much more difficult to create graphical materials that illustrate the important concepts related to the processes, and there is still the question of how much they will help the students.

Figure 4 shows one graphic that we developed to explain the concept of statistical validity, i.e. that the changes in the dependent variable in an experiment are caused by changes in the independent variable, and not due to chance. The graphic includes some aspects that should be familiar to the students, like a basic 2-dimensional graph and frequency distributions. We still don't know, however, if such a graphic is too abstract to be beneficial to the students. Another issue is the method of presentation of the graphic. Should we start with a blank slate and add components as the discussion progresses, or should we present the entire graphic and center the discussion around getting the student to understand it? A further complication is related to the use of talking head along with the graphics. The animated head might distract the student from attending to the graphic. A solution might be to either have the head disappear or wander off-screen, or to have it look at the graphic, "inviting" the student to do the same.

Another mixed-language explanatory technique could be used when the tutoring topic calls for the student to list a number of items, for example, possible confounds for an experiment. Using simple HTML commands, the tutor could jot the items down in a "notepad" area of the screen as the students names the items in order to reduce the student's working memory load.

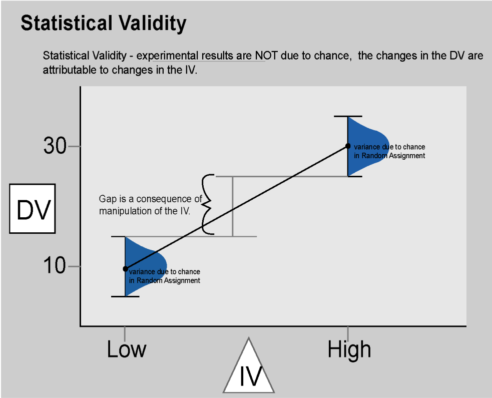

Figure 4. A graphic for explaining statistical validity

To evaluate RMT's graphical elements (including the talking head), we have started to use an eyetracker to analyze where and for how long a student's gaze is directed. We hope to correlate this data with eyetracker-based measures of the student's cognitive load [6] to identify aspects of the interaction that may be confusing to the student.

In this section we have presented a few ideas and a lot of questions. The use of mixed-language explanations has the potential to be extremely beneficial to students by helping them process and integrate new information, but significant research must be done to understand how it can best be deployed.

References

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

[2] R. Clark and R. Mayer. *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. Pfeiffer, 2002.

[3] A.C. Graesser, G.T. Jackson, E.C. Mathews, H.H. Mitchell, A. Olney, M. Ventura, P. Chipman, D. Franceschetti, X. Hu, M.M. Louwerse, N.K. Person, and TRG. 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, 2003. Erlbaum.

[4] T.K. Landauer and S.T. Dumais. A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997.

[5] N. K. Person. *An analysis of the examples that tutors generate during naturalistic one-to-one tutoring sessions*. PhD thesis, University of Memphis, Memphis, TN, 1994.

[6] D. Salvucci and J. R. Anderson. Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pages 923–928, Hillsdale, NJ, 1998. Lawrence Erlbaum Associates.

[7] P. Wiemer-Hastings, K. Wiemer-Hastings, and A. Graesser. Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis. In S. Lajoie and M. Vivet, editors, *Artificial Intelligence in Education*, pages 535–542, Amsterdam, 1999. IOS Press.