

Intelligent Tutoring in a Non-Traditional College Classroom Setting

Elizabeth Arnott-Hill^{1,*}, Peter Hastings², David Allbritton³

¹Department of Psychology, Chicago State University, Chicago, IL, USA

²Department of Psychology, DePaul University, Chicago, IL, USA

³Department of Computing and Digital Media, DePaul University, Chicago, IL, USA

Abstract Research Methods Tutor (RMT) is a web-based intelligent tutoring system designed for use in conjunction with introductory research methods courses. RMT has been shown to result in average learning gains of .75 SDs above classroom instruction in traditional college environments. However, a primary goal of the RMT project is to provide greater access to tutoring for students without access to traditional one-on-one human tutoring. Therefore, we further tested RMT's effectiveness at a university that enrolls primarily non-traditional students. Although we again found evidence of RMT's effectiveness, a few key outcome differences between traditional and non-traditional environments emerged, including perceptions of the pedagogical agent and access to the system.

Keywords Intelligent Tutoring, Technology and Education, Non-traditional Students

1. Introduction

Computing technologies, particularly those that allow students convenient access to learning resources, are infiltrating every aspect of the educational system. Textbooks routinely come with online student resources, and many instructors regularly assign online or computer - based homework to enhance the classroom learning experience. One of these computer-based technologies, intelligent tutoring systems (ITS), seeks to mimic the benefits of one-on-one human tutoring. Expert human tutoring is believed to be among the most beneficial learning techniques. Bloom, for example, found that a successful human tutoring situation can enhance learning by up to 2.3 standard deviations over classroom instruction alone[1]. In contrast to more didactic approaches, such as lecture, textbook learning, or websites that mimic a traditional textbook layout, tutoring allows students to engage in a dialog that helps them to assess their current levels of understanding and to work cooperatively to increase levels of knowledge[2-7]. The extent of the student's engagement in this dialog is correlated with the student's learning outcomes[8-10]. An ITS that uses the techniques employed in expert human tutoring, therefore, could provide access to this powerful learning technique without the drawbacks of human tutoring, which include the cost, the time involvement, and the inconvenience of coordinating schedules. Due to the challenges faced by many of

today's students, including childcare, work schedules, and transportation[11], ITS have the potential to make vast contributions to effective teaching and learning.

A number of successful ITS have been created and tested to date. For example, Koedinger and colleagues[12] conducted a large-scale study of the effectiveness of PAT (the PUMP Algebra Tutor) in conjunction with a high school algebra curriculum. They found that students who used the system scored 100% higher on basic skills tests than those who did not use the system. In laboratory-based tests, another ITS, AutoTutor, increased students' scores by one standard deviation above reading a textbook alone[13]. ANDES, an ITS used in undergraduate physics courses, has also been associated with significant learning gains for students who use the system in lieu of pencil-and-paper based homework[14].

Research Methods Tutor (RMT) is a dialog-based ITS created to increase student learning in introductory undergraduate psychology research methods courses. Most psychology programs require at least one course in research methods[15]. Research methods courses, however, tend to be difficult for psychology majors due to their applied, technical, and largely quantitative nature. In addition, research methods courses require students to engage in critical thinking, which is generally a skill that undergraduate college students are in the process of developing[16]. RMT is designed for use in conjunction with traditional coursework and is available online, making it convenient to use at any time of the day without cost. It consists of five topic modules that coincide with typical topics from psychology research methods - ethics, variables, reliability, validity, and experimental design. Each RMT topic module is assigned after the

* Corresponding author:

earnott@csu.edu (Elizabeth Arnott-Hill)

Published online at <http://journal.sapub.org/ijap>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

topic is covered in the classroom.

The RMT system has two presentation modes that allow the student to interact with the tutor in a specific way based on student preference or technical requirements. Students are presented the material either in text on the computer screen (the text-only mode) or via an on-screen interactive pedagogical agent (the agent mode). For the text-only mode, students see RMT's questions and feedback in text on the screen. In the agent mode, an interactive avatar "speaks" RMT's content and the student types his/her response in a

response box onscreen. RMT's pedagogical agent, Mr. Joshua, is shown in Figure 1. Mr. Joshua not only speaks, but also blinks, turns his head, and gestures with his hands. Due to the requirements of the agent software, the student must be able to download the required software (thus, students in a public computer lab cannot download the agent) and must have a PC with Internet Explorer. Students who meet these criteria can choose to download the software or can opt to use the text-only version of the system.

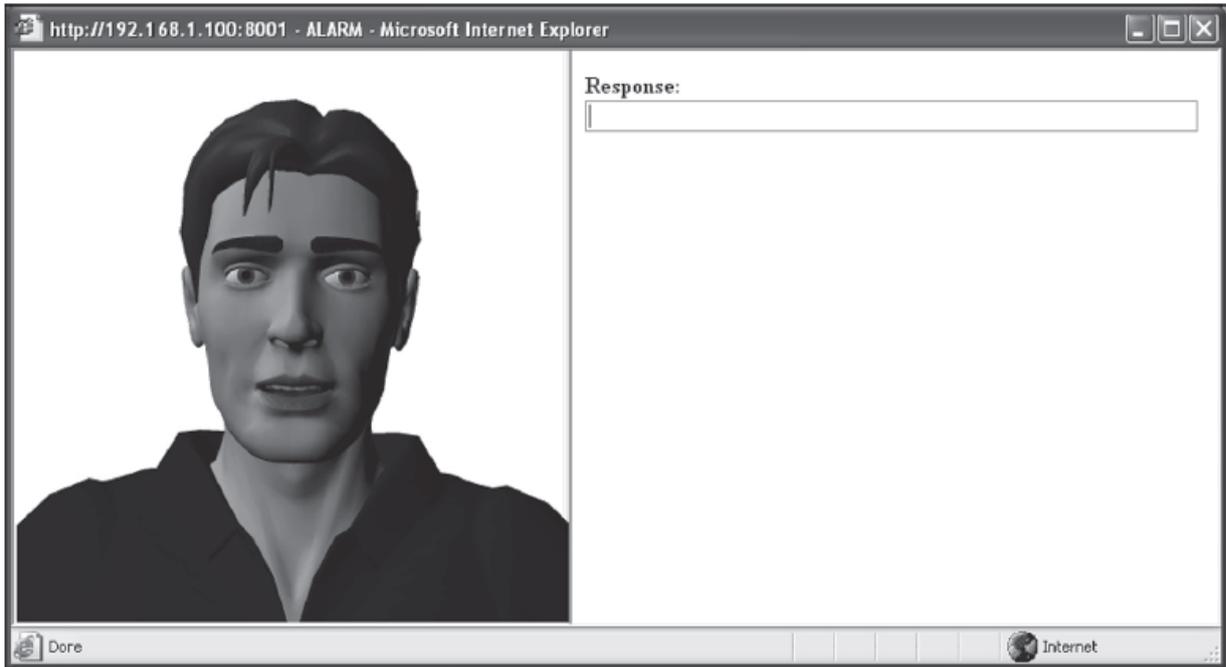


Figure 1. The RMT interface with the animated pedagogical agent, Mr. Joshua, used for the "agent mode"

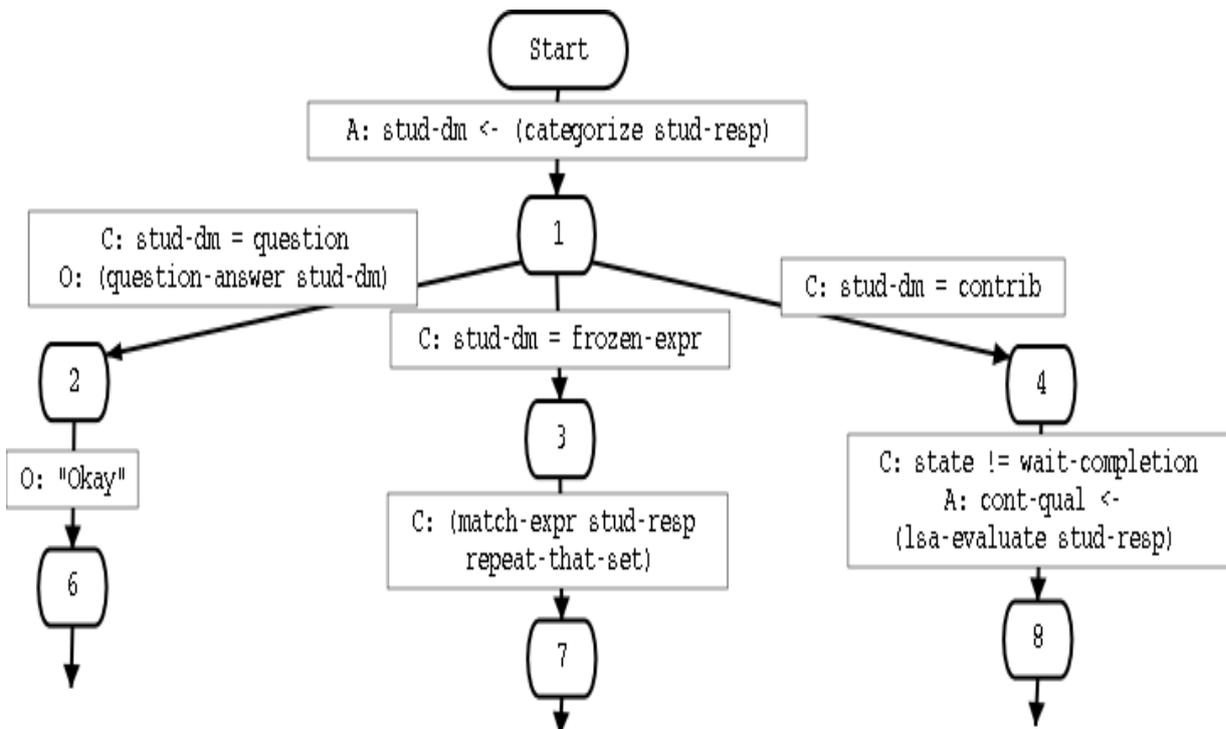


Figure 2. A Partial View of the Dialog Advancer Network (DAN)

RMT – modeled after AutoTutor[17,18] – involves a natural language dialog between RMT and the student. RMT poses a question to the student, and the student types his/her answer into a response box onscreen. If the student answers correctly, the tutor moves the dialog along. If the student does not answer correctly, the tutor provides prompts and hints. The tutor's behavior is controlled by a Dialog Advancer Network – or DAN – where each path categorizes student input as correct or incorrect, creates a response based on that categorization, and issues a follow-up utterance in the form of a prompt or hint if the student is incorrect or an additional question if the student is correct. This behavior allows feedback to be immediate, and forces a complete correct answer from the student before he/she moves on. A partial view of the Dialog Advancer Network is shown in Figure 2.

RMT evaluates the student's natural language responses using a three-part method. If a word is not located in RMT's lexicon, the system looks for possible misspellings using a spell checker. RMT also employs a keyword matcher that checks for literal similarity between words. Finally, RMT uses latent semantic analysis (LSA)[19] to compare the student's response to the expected answer by creating a high-dimensional vector representation of both the response and expected answer. The vectors are based on relationships found in a corpus of domain-relevant texts. The cosine of these vectors is used to represent the similarity of the response to the expected answer.

Following Bloom's taxonomy[20] the tutor asks conceptual ("What is an independent variable?"), analytic ("Given the following experimental scenario, can you find any potential problems with the reliability of the study?"), and synthetic questions ("Now, given what you know about informed consent, can you create an informed consent document for a study?") for each topic. At the end of each section the tutor provides a summary of the material that was covered.

Previous research has demonstrated the effectiveness of RMT in the traditional college classroom[21]. In this initial test, students who used RMT demonstrated learning gains of .75 standard deviations over students who did not use RMT. However, a primary goal of the RMT project was to provide greater access to tutoring for students who cannot easily obtain access to an expert tutor. Due to financial and time restrictions, this need is likely greatest among non-traditional student populations[11]. Since previous work has suggested that traditional and non-traditional students may respond differently to various aspects of the learning situation[22-24], we were interested in assessing RMT in a non-traditional student environment. Further, ITS research has suggested that students who use an ITS when they are beginning to learn a topic may benefit more than students who are more advanced in their understanding. In a review of results from a number of ITS, Van Lehn and colleagues[25] argued that the increased interaction that tutorial dialog provides is most beneficial for students who are learning material above their current knowledge level (novices

learning intermediate material, for example). This is commonly known as the *interaction hypothesis*. As we believed that a non-traditional sample might include students that were more varied in their preparedness level or familiarity with the material, the present assessment allowed us to investigate both non-traditional student learning outcomes and student perceptions of RMT, as well as to assess how student knowledge at the beginning of the semester might impact overall learning gains. Thus, our primary research questions were: 1) Does RMT result in increased learning gains when used with non-traditional students?, 2) Do non-traditional students perceive RMT differently than traditional students?, and 3) Do those who are less prepared at the beginning of a semester show greater gains when using RMT than those who are more prepared?

2. Method and Results

2.1. Participants

Our non-traditional student assessment was conducted at an urban university with a large non-traditional student population. For the purposes of the present study, we broadly defined "non-traditional student" as a student who was not of traditional college age (18-23), worked full-time, acted as a primary caregiver / parent, or belonged to a historically underrepresented group. Of the 87 students in the sample, 98% were African American, 65% were parents or primary caregivers of children, 67% held full-time jobs, and 100% lived off-campus. The average age of the students was 24. Table 1 compares the demographic attributes of students in this non-traditional student sample to the students in our traditional sample.

Students in RMT classes completed all RMT topic modules for course credit, as well as a 50-item pre-test and post-test, but were allowed to anonymously opt-out of data collection. Students in classes that did not use RMT completed the same 50-item pre-test and post-test for course credit, but, once again, they were allowed to anonymously opt-out of data collection.

2.2. Procedure

In order to assess the effectiveness of RMT, we selected two research methods courses to use the RMT system and two to serve as non-equivalent control groups. The data was collected over the course of two semesters. During the first semester of data collection the daytime section used RMT, and the evening section did not. During the second semester of data collection the daytime section did not use RMT, and the evening section used RMT. Students in all courses were given a 50-item multiple choice pre-test during the first week of the semester and the same 50-item post-test during the last week of regular classes. The test consisted of 10-items per topic module covered by RMT. At the end of the semester, an additional 25 question post-test (5 items per topic module) was given to test for transfer learning. This transfer learning test consisted of research scenario-based critical analysis

questions, which the students had not seen before. In addition, at the end of the semester, RMT students were given an 11-item questionnaire regarding their perceptions of the system. The questionnaire consisted of 8 questions that employed a 1 (Strongly Disagree) to 6 (Strongly Agree) scale and one yes/no question. Students were also encouraged to give general feedback in an open-ended question at the end of the questionnaire (“What, specifically, did you like/dislike about using the system? Do you have any other feedback for us?”).

Students in all courses used the same textbook, followed the same basic course outline, and were taught by the same instructor. The only major difference between the conditions was that the students in the RMT section were assigned to complete RMT topic modules after the topics were covered in the class. RMT students who could download software (had regular access to a PC with Internet Explorer) were asked to select the agent presentation condition, and students who could not were assigned by default to the text-only condition.

2.3. Results

Students who did not complete either the pre-test or post-test and RMT students who did not complete all of the RMT modules were removed from the analysis ($n=6$), leav-

ing 56 students in the RMT condition and 31 in the control group. To address our first research question (Does RMT result in increased learning gains when used with non-traditional students?), we conducted an ANCOVA with the pre-test score as the covariate, the post-test score as the dependent variable, and the classroom condition (RMT vs. control) as the independent variable. We found that RMT students performed significantly better than control students on the basic assessment measure (post-test only), $F(1, 84) = 54.78, p < .01$, with an NRP effect size[26] of 1.39 standard deviations. RMT students had an average score of 32.6 (out of 50 possible – 65%) with a standard error of 1.1, while the control students had an average score of 19.5 (39%) with a standard error of 1.4. When only the transfer task was examined, RMT students significantly outperformed control students, $F(1,84) = 14.05, p < .01$, NRP effect size = .71 standard deviations. RMT students scored an average of 13/25 with a standard error of .64, while control students scored an average of 8.9/25 with a standard error of .86. On the total assessment (post-test + transfer), RMT students also significantly outperformed the control, $F(1,84) = 42.99, p < .01$, NRP effect size = 1.21 standard deviations. A comparison of the measured learning outcomes of the non-traditional and traditional student samples is found in Table 2.

Table 1. Demographic Comparison of Traditional and Non-traditional Student Samples

	Student Sample Type	
	Traditional Sample	Non-traditional Sample
Student of Color	25%	98%
Student is a Parent/Guardian	8%	65%
Full-time Employee	33%	67%
Lives On-campus	12%	100%
Average Age	20	24

Table 2. Comparison of Non-traditional Student Sample and Traditional Student Sample

	Non-traditional Sample		Traditional Sample	
	RMT	Control	RMT	Control
N	56	31	73	85
% of Students Downloading Pedagogical Agent	31%	NA	79%	NA
Average Score at Pre-test	42.3% SD = 9.7	43.7% SD = 12.4	54.3% SD = 12.8	54% SD = 11.3
Average Score at Post-test	65% SD = 13.9	39% SD = 22.7	65.3% SD = 15.5	56% SD = 16.3
Average Learning Gain (Pre-test to Post-test)	22.7%	-4.1%	11%	2%

Table 3. Mean Ratings (and Standard Deviations) on the Student Questionnaire for the Non-traditional and Traditional Student Samples

Question	Non-traditional Sample	Traditional Sample
1. I enjoyed using the system.	M=5.2 SD = .65	M=5.5 SD = .41
2. RMT helped me to learn the material.	M=5.2 SD = .72	M=5.1 SD = .33
3. RMT helped by critical thinking.	M=5.2 SD = .99	M=5.0 SD = .35
4. I would choose to use the RMT system, even if it wasn't part of a course requirement.	M=5.1 SD = 1.1	M=5.2 SD = .87
5. I own a computer.	34% = yes 64.5% = no	85% = yes 15% = no
6. I am comfortable with technology	M=3.7 SD = .92	M= 5.6 SD = 1.2
7. The system is easy to use.	M=3.1 SD = .77	M=5.1 SD = .99
8. The system is easily accessible.	M=3.4 SD = .86	M=5.7 SD = .54
9. I like the overall appearance of the system	M=3.6 SD = 1.31	M=5.2 SD = .96

In order to assess our second research question (Do non-traditional students perceive RMT differently than traditional students?), we found mean ratings (with standard deviations) for each item on our student perception questionnaire. Of the students that used RMT, most gave the system high ratings in terms of overall learning and enjoyment (all averages over 5 on a 6-point scale; see Table 3 for means and standard deviations of all questions for both samples). However, students gave much lower ratings to the questions concerning accessibility ($M=3.4$, $SD=.86$), ease of use ($M=3.1$, $SD=.77$), and the overall look of the system ($M=3.6$, $SD=1.31$). Further, 64.5% of students indicated that they did not own a computer and on average students gave low ratings to their comfort with technology ($M=3.2$, $SD=.97$). In addition, when we compared the number of students in the pedagogical agent versus the text-only conditions, 69% of students were unable to download the animated pedagogical agent and were assigned by default to the text-only condition. Therefore, only a small percentage of students were able to interact with the agent (31%). This is particularly interesting when compared to the traditional sample, in which 79% of the students used the agent and only 21% of students were unable to download the required software.

Our last research question was intended to test the interaction hypothesis. That is, we wanted to know if students who are less prepared at the beginning of a semester show greater gains when using RMT than those who are more prepared. Although we found no significant relationship between low pre-test scores and overall learning gains within our non-traditional sample, our non-traditional sample did have lower pre-test scores (by approximately 10 percentage points) than our traditional sample (see Table 2). The samples had similar scores at post-test, which means that our non-traditional sample yielded learning gains that were twice those of the traditional sample.

3. Conclusions

Using a non-traditional student sample, we found additional evidence that RMT boosts student learning outcomes above classroom instruction alone. The observed RMT learning effect sizes of .71-1.39 SDs (based on the outcome measure used in the analysis) are impressive evidence of the system's effectiveness. The demonstrated learning gains are comparable with lab-tested ITS, such as AutoTutor[13], and with estimates of the learning gains from one-on-one human tutoring[1]. This is particularly impressive given the fact that, unlike researchers using lab-based evaluations of ITS, the "naturalistic" approach used in the present study gave the researchers little control over the amount of time and attention paid to each RMT topic module. In addition, RMT was only used for 3-5 total hours over the course of the 16-week semester.

When the results from the traditional and non-traditional samples were compared, a number of interesting differences emerged. First, although students in both traditional and

non-traditional samples indicated that they enjoyed using RMT and that they believed that it aided learning, major differences in student accessibility were observed. Approximately 64.5% of students in our non-traditional sample did not have access to a computer, compared with only 15% in our traditional sample. Since the animated pedagogical agent had to be downloaded, students using a lab or work computer could not use the agent. Not surprisingly, only 31% of students in the non-traditional sample downloaded the agent, while 79% of students in the traditional sample used the agent. Non-traditional students also gave lower ratings to the accessibility of the system and their own comfort with technology. This poses a significant problem, as a primary goal of the RMT system was to provide greater access to tutoring. The second major difference centered on student perception of the animated pedagogical agent. In the non-traditional sample, students gave lower ratings to the overall look of the system. When asked for specific likes and dislikes, a large number (55%) of the students who successfully downloaded and used the agent commented on the agent's appearance. The primary complaints were that the agent was difficult to understand and that the agent appeared to be a European American male. No student in the traditional sample voiced either of these concerns. Since previous research has suggested that various social cues, such as gender and perceived ethnicity[27], can affect the learning situation, it is possible that the largely female and African American non-traditional sample was not able to relate to the agent in a positive way. It is also possible that the agent distracted students from the learning situation, given the mixed evidence in support of animated pedagogical agents[28]. Future research would be needed to investigate this phenomenon, as the present study had a very low rate of agent usage, students self-selected into the agent condition, and there was no alternative agent for comparison.

We also found that learning effect sizes were higher with the non-traditional student sample than those found in the traditional sample. Students in our non-traditional sample performed worse at pre-test than those in the traditional sample. The scores at post-test, however, were equivalent, resulting in learning gains for the non-traditional sample that were double those of the traditional sample. This finding is consistent with the ITS literature concerning the benefits of tutorial dialog[25]. Tutorial dialog has been found to be more useful for those who are novices in a given area. If pre-test score can be used as a measure of preparedness or familiarity with the course material, then the non-traditional sample was clearly less familiar with the material than the traditional sample at the beginning of the course. A major limitation of direct score comparison between the two samples, however, is that the pre-test/post-test was altered significantly in order to increase reliability. The traditional sample was administered a 106-item pre-test / post-test with approximately 20 items per topic module, while the non-traditional sample was given a 50-item pre-test / post-test with exactly 10 items per topic module. In addition, the traditional sample did not take the 25-item transfer learning test. A follow-up study could

confirm and clarify these results by using the 50-item pre-test / post-test and transfer test with another traditional sample.

Taken together, the data provide encouraging evidence of effectiveness. Specifically, the large learning gains observed in our non-traditional sample suggest that ITS may be particularly appropriate in this setting. However, clear directions for improvement also emerged. If an ITS is to be useful it should reach people that do not have access to human tutoring. The large number of student access difficulties, together with the dislike of the animated agent suggests that a new approach be taken in the design of the system. Thus, we plan to not only change the type of animated agent used, but also to investigate new techniques for using an agent that does not require a software download. We hope that these changes will make the system easier to use – especially for those uncomfortable with technology – and more widely accessible to students.

ACKNOWLEDGEMENTS

The authors wish to acknowledge their research assistant, Michelle Rogers, for her help with data entry and analysis.

REFERENCES

- [1] B. S. Bloom, 1984, The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring, *Educational Researcher*, 13, 4-16.
- [2] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pletier, 1995, Cognitive tutors: Lessons learned, *Journal of the Learning Sciences*, 4, 167-207.
- [3] A. C. Graesser, N. K. Person, and J. P. Magliano, 1995, Collaborative dialogue patterns in naturalistic one-to-one tutoring, *Applied Cognitive Psychology*, 9, 495-522.
- [4] H. C. Lane, and K. VanLehn, 2005, Teaching the tacit knowledge of programming to novices with natural language tutoring, *Computer Science Education*, 15, 183-201.
- [5] A. Lesgold, S. Lajoie, M. Bunzo, and G. Eggan, 1992, Sherlock: A coached practice environment for an electronics troubleshooting job, In J. H. Larkin and R. W. Chabay, Eds., *Computer assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches*, 201-238, Hillsdale, NJ: Erlbaum.
- [6] 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, USA, 1994.
- [7] N. K. Person, A. C. Graesser, J. P. Magliano, and R. J. Kreuz, 1994, Inferring what the student knows in one-to-one tutoring: The role of student questions and answers, *Learning and Individual Differences*, 6, 205-229.
- [8] M. T. H. Chi, S. A. Siler, H. Jeong, T. Yamauchi, and R. G. Hausmann, 2001, Learning from human tutoring, *Cognitive Science*, 25, 471-533.
- [9] D. Wood and D. Middleton, 1975, A study of assisted problem-solving, *British Journal of Psychology*, 66, 181-191.
- [10] D. Wood, H. Wood, and D. Middleton, 1978, An experimental evaluation of four face-to-face teaching strategies, *International Journal of Behavioral Development*, 1, 131-147.
- [11] C. Leathwood, and J. Dalgety, 2002, 'It's a struggle': The construction of the 'new student' in higher education, *Journal of Education Policy*, 18(6), 597-615.
- [12] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark, 1997, Intelligent tutoring goes to school in the big city, *International Journal of Artificial Intelligence in Education*, 8, 30-43.
- [13] 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 the Tutoring Research Group, 2003, Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog., Proc., 25th Annual Conference of the Cognitive Science Society, Mahwah, NJ.
- [14] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill, 2005, The Andes Physics Tutoring System: Lessons Learned, *International Journal of Artificial Intelligence in Education*, 15, 3.
- [15] B. Perlman & L. I. McCann, 1999, The structure of the psychology undergraduate curriculum, *Teaching of Psychology*, 26, 171-176.
- [16] D. R. Lehman and R. E. Nisbett, R. E., 1990, A longitudinal study of the effects of undergraduate training on reasoning, *Developmental Psychology*, 26, 431-442.
- [17] A. Graesser, N. Person, D. Harter, and the Tutoring Research Group, 2001, Teaching tactics and dialog in AutoTutor, *International Journal of Artificial Intelligence in Education*, 12, 23-39.
- [18] P. Wiemer-Hastings, A. Graesser, D. Harter, and the Tutoring Research Group, 1998, The foundations and architecture of AutoTutor, Proc., 4th International Conference on Intelligent Tutoring Systems, Berlin: Springer, 334-343.
- [19] T. K. Landauer, P. W. Foltz, and D. Laham, 1998, An introduction to latent semantic analysis, *Discourse Processes*, 25, 259-284.
- [20] B. S. Bloom, 1956, *Taxonomy of educational objectives: The classification of educational goals*, Essex, England: Longman Group Limited.
- [21] E. Arnott, P. Hastings, and D. Allbritton, 2008, Research Methods Tutor: Evaluation of a dialog-based tutoring system in the classroom. *Behavior Research Methods*, 40(3), 694-698.
- [22] M. Hoyert, and C. D. O'Dell, 2009, Goal orientation and academic failure in traditional and nontraditional aged college students, *College Student Journal*, 43(4, PtA), 1052-1061.
- [23] A. Strage, 2008, Traditional and non-traditional college students' descriptions of the 'ideal' professor and the 'ideal' course and perceived strengths and limitations, *College Student Journal*, 42(1), 225-231.
- [24] J. Taylor, and B. House, 2010, An exploration of identity,

motivations and concerns of non-traditional students at different stages of higher education. *Psychology Teaching Review*, 16(1), 46-57.

- [25] K. VanLehn, A. C. Graesser, G. T. Jackson, P. Jordan, A. Olney, and C. P. Rosé, 2007, When are tutorial dialogues more effective than reading?, *Cognitive Science*, 31, 3-62.
- [26] Report of the 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, Washington, DC: National Institute of Child Health & Human Development.
- [27] R. Moreno, and T. Flowerday, 2006, Students' choice of animated pedagogical agents in science learning: A test of the similarity attraction hypothesis on gender and ethnicity, *Contemporary Educational Psychology*, 31, 186-207.
- [28] R. Moreno, 2004, Animated pedagogical agents in educational technology, *Educational Technology*, 44, 23-30.