The Effectiveness of Computer Simulations in a Computer-based Learning Environment

Hyun-Jeong Joyce Kim[1], Art Graesser[2], Tanner Jackson[2],
Andrew Olney[2], and Patrick Chipman[2]
[1]*Department of Psychology, Rhodes College, 2000 N. Parkway, Memphis, TN 38112,*
[2]*Department of Psychology, University of Memphis, Memphis, TN 38152*
kimh@rhodes.edu, {a-graesser|gtjacksn|aolney|pchipman}@memphis.edu

Abstract: We explored the effectiveness of interactive simulations on an intelligent tutoring system, called AutoTutor. College students (N=132) were randomly assigned to one of three conditions: (1) AutoTutor with simulations, (2) AutoTutor without simulations, and (3) Monte Carlo AutoTutor. A pretest-posttest design was used to measure learning gains. The findings indicate that all versions of AutoTutor were successful in promoting learning. Although differences were not significant among all of the tutor conditions, they were in the predicted directions (1>2>3) and condition 1 was significantly better than 3. Improved simulation dialogues and faster display of simulations are expected to enhance learning in future versions of AutoTutor.

Background

Learning physics in high school and college is a great challenge for students and sometimes requires advanced cognitive abilities. Physics contains complex topics which require an understanding of three-dimensional dynamics of objects, force vectors, counterintuitive laws, and sometimes theoretical entities that are difficult to visualize (diSessa, 1993; Ploetzner & VanLehn, 1997). These challenges make it difficult to construct meaningful mental representations that are coherently integrated. The difficulty of performing these cognitive activities requires cognitive effort and imposes a large cognitive load on verbal and visual memory systems (Baddeley, 1992).

Advanced learning technologies offer possible methods of promoting active constructive learning, deeper understanding, and reductions in cognitive load. One notable class of learning technologies is computer simulation. Simulations have been considered a valuable tool because they present a visual representation and an animation of the mechanisms. Meaningful learning can occur when learners view concrete simulations of abstract properties of physics (e.g., acceleration, gravity, etc.) and can allocate their mental resources to other mental activities, such as building coherent connections between components of the system and more accurate mental models. Simulations are potentially more valuable when learners can interact with the system, as opposed to simply observing the simulations in action. Interactive simulations enable students to manipulate parameters of the world, test their hypotheses, interpret simulated output, and identify their misconceptions. Therefore, interactive computer simulations should help students develop deep reasoning and problem solving skills.

A number of studies have investigated the effectiveness of simulations during the last two decades, but the results from these studies have had conflicting outcomes (Bourque & Carlson, 1987; Choi & Gennaro, 1987; Gorsky & Finegold, 1992; Lee, 1999; Michael, 2001; Roberts & Blakeslee, 1996; Zhang, Chen, & Reid, 2000). In a meta-analysis of 93 studies 25 years ago, Dekker and Donatti (1981) concluded the evidence of simulation efficacy was inconclusive. However, that study is now dated so the question arises whether more contemporary systems are more effective. There are notable successes, of course. Some researchers have reported that computer simulations can enhance students' understanding of scientific concepts and problem solving skills under some conditions (Rivers & Vockell, 1987; Woodward, Carnine, & Gernsten, 1988).

Rieber (1993) has advocated the use of appropriately-designed simulations in the teaching of natural sciences and troublesome topics. If the material is simple and easy to comprehend, there is little or no incremental enhancement of learning through interactive simulation. Strangman, Hall, and Meyer (2003) emphasized three aspects of student learning when considering the usefulness of computer simulations: conceptual change, skill development, and content area knowledge. According to their claims, computer simulations were most effective in math and science domains. Simulations were also effective at enabling students to identify and correct misconceptions with immediate feedback and outcomes, at developing more accurate scientific and mathematical models, at enhancing basic cognitive skills (e.g., reading, problem solving, abstract thinking), and at developing complex content area knowledge (e.g., a lake's food chain, chemical molecules). Although Strangeman et al. (2003) acknowledged computer simulations were a potentially powerful learning technology, they drew conclusions with caution because there were serious limitations of most studies and pervasive problems with research methodologies. Some of the limitations included failure to randomize group assignment, absence of control groups, the presence of confounding variables, and no pretest to allow comparisons between pre-test and post-test. Nevertheless, when the poorly designed studies were excluded, computer simulations were shown to offer an advantage over more traditional instructional experiences.

The current study attempted to examine the effectiveness of simulations while interacting with AutoTutor, an intelligent tutoring system with interactive simulation. Given that the literature shows mixed results on the value of interactive simulation in promoting learning, it was important to document whether AutoTutor's interactive simulation was effective. One salient advantage of AutoTutor is that it helps learning by holding a conversation in natural language. It is conceivable that this adaptive conversational facility of AutoTutor will help guide the learner in using interactive simulation. In essence, the combination of adaptive natural language dialogue and interactive simulation might be a perfect marriage to promote substantial learning gains.

AutoTutor

AutoTutor is a learning environment on the Internet that holds a mixed-initiative dialogue with students in natural language, with dialogue moves and strategies that human tutors typically use. Current versions of AutoTutor teach conceptual physics, computer literacy and critical thinking to students (Graesser et al., in press; Graesser et al., 2004; Graesser et al., 1999; Kim et al., 2005; Person et al., 2001). AutoTutor has an animated talking head with a synthesized speech, facial expressions, and gestures. The agent presents difficult problems that tap into deep knowledge, engages the students in complex problem solving through conversations, and helps them actively construct knowledge. AutoTutor evaluation studies for the past five years have demonstrated that AutoTutor is successful in enhancing learning with sigma between .4 and 1.5 (Graesser et al., 2004; Person et al., 2001; VanLehn, Graesser, et al., 2002).

It is beyond the scope of this article to describe the computational mechanisms of AutoTutor, but a few highlights should convey salient features of the system. AutoTutor uses Latent Semantic Analysis (LSA) (Foltz, Kintsch, & Landauer, 1998; Kintsch 1998; Landauer & Dumais 1997) in its semantic analyzer that helps interpret what the student expresses and that guides the construction of dialogue moves to orchestrate a relatively smooth conversation. One of AutoTutor's dialogue mechanisms is called *expectation and misconception dialogue* (Graesser, Hu, & McNamara, 2005). For a particular problem, ideal answers (i.e., expectations) and wrong answers (i.e., misconceptions) are stored in AutoTutor's curriculum script that is generated by domain experts. LSA is used to compare student input to every expectation and misconception in the curriculum script, to determine if student input covers the list of expectations and misconceptions, and then to select what AutoTutor will say next in response to the student input.

Recently AutoTutor incorporated interactive simulations into its learning environments (Graesser et al., 2004; Kim et al., 2005). Students are presented with 3-D simulations in which the students are allowed to manipulate a set of parameters (e.g., speed and mass of objects) to alter the simulations, to test their hypotheses, and to find solutions to a given problem. The simulations are presented to the student only when he/she does not understand the problem well. Moreover, while interacting with simulations, the student's verbal expressions are carefully monitored and its corresponding tutorial dialogue is composed in a fashion that scaffolds the learning process. In other words, AutoTutor generates dialogue moves that are adaptive to the student's manipulations within the simulations.

The Present Study

The present study focused on the effectiveness of simulations and adaptive dialogues in AutoTutor. The version of AutoTutor we used in the study taught students Newtonian physics. There were three tutor conditions: (1) AutoTutor with interactive simulations, (2) the normal AutoTutor with conversation that are sensitive to LSA modules, and (3) AutoTutor with conversation that randomly generates dialogue movers that are relevant to a particular expectation without consideration of LSA (i.e., Monte Carlo AutoTutor). AutoTutor with conversation generated by LSA (*Normal AutoTutor*) implemented the LSA component that senses and adapts to the knowledge and contributions of the students. AutoTutor with interactive simulations (*Simulation AutoTutor*) also used LSA to adapt the conversation to individual students, but also included interactive simulations and associated dialogues. *Monte Carlo AutoTutor* was the same as the Normal AutoTutor except that the selection of the dialogue moves was not sensitive to the learners' knowledge. Instead, thousands of dialogue patterns were collected and analyzed from previous AutoTutor evaluation studies in which AutoTutor used LSA in the normal way. The Monte Carlo approximation of LSA used the statistical distribution of these dialogue patterns to select randomly the response to student input. In essence, the same distribution of dialogue moves was presented to the group of subjects in both the Normal and Monte Carlo conditions, except that learners in the Monte Carlo condition did not receive dialogue moves that were sensitive to their knowledge states.

The present study compared the three AutoTutor conditions in order to examine the effect of LSA-based user modeling and simulations on learning. If Normal AutoTutor produces significantly larger learning gains than Monte Carlo AutoTutor, then we can conclude that user modeling is an important component of AutoTutor's one-on-one conversational interaction. If Simulation AutoTutor produces larger learning gains than Normal AutoTutor, then interactive simulations are useful educational techniques for teaching physics in computer-based learning environments.

Methods

There were two experiments. In Experiment 1, 132 students at Rhodes College were paid to participate. There were three phases: a pre-testing phase, a learning phase, and a post-testing phase. During the pre-test, the participants were instructed to complete 26 multiple choice questions developed by VanLehn et al. (2005) and a demographics survey. During the learning phase, the participants were randomly assigned to one of the three versions of AutoTutor, interacted with one of them and attempted to solve four problems in Newtonian physics. During the post-test, the participants filled out a different set of 26 multiple-choice questions and a user perception survey. There were two versions of the test, such that the post-test was counterbalanced with pretest. The participants spent approximately two hours to complete the experiment. In Experiment 2, 95 students at the University of Memphis and Rhodes College participated in the study. The procedure of study 2 was identical to that of study 1, except that there was no Monte Carlo condition. We did not include Monte Carlo condition in Experiment 2 because in the previous experiment, the Monte Carlo version did not produce as good performance scores as Normal AutoTutor and Simulation AutoTutor.

Results and Discussion

The three different tutors (Normal, Simulation, vs. Monte Carlo) were compared on four outcome measures: pre-test, post-test, learning gains, and effect size. The pre- and post-test scores were proportion scores, consisting of the proportion of items answered correctly out of 26 questions. We calculated the learning gains by subtracting the pre-test proportion from the post-test proportion. Effect size was computed by subtracting a pretest mean score from a posttest mean score and dividing it by the standard deviation of the pretest score.

When we examined the pre-test scores, there were no significant differences among the tutoring conditions. That is, all the participants across the conditions started on an even playing field. When we examined learning gains, we found that all the versions of AutoTutor produced significant learning gains; posttest scores (M =.60, SD =.18) were significantly better than pretest scores (M = .46, SD = .19), $F(1, 129) = 141.17$, $p < .001$. Although there was no overall significant effect of tutoring conditions on learning gains, the data trend did support the predictions, with learning gains (post – pre) of .17, .15, and .12 in the Simulation, Normal, and Monte Carlo versions, respectively. The effect sizes also reflected the predicted trend: Simulation (d =.94) > Normal (d =.75) > Monte Carlo (d =.60).

We further analyzed the data by examining the middle 50% of the students whose pretest scores were greater than .3 and less than .55. We wanted to examine AutoTutor effects for these middle participants because they may have been the most malleable to effects of learning methods. There was indeed a trend showed that the middle participants performed significantly better with Simulation than with Monte Carlo and Normal AutoTutor. The Simulation condition therefore was the most effective of all for the middle participants.

We divided the participants into high and low prior knowledge using a median split (.42) on the pretest scores. A 3 (three tutor conditions) x 2 (low vs. high knowledge) ANOVA was conducted on learning gains.

Although a significant main effect was found on domain knowledge, with high-ability students scoring better than low-ability students on simple learning gains [$F(1, 126) = 17.42$, $p < .001$], no other significant effects were found.

In study 2, we compared the two different AutoTutor conditions on learning gains. Overall, we found that both AutoTutor versions produced significant learning gains. The posttest scores (M =.56, SD =.19) were significantly greater than pretest scores (M = .39, SD = .20), $F(1, 93) = 87.73$, $p < .001$. However, there were no significant simulation effects. Effect sizes revealed that Normal AutoTutor produced higher performance improvement (d = 1.19) than Simulation AutoTutor (d = .75). However, because the students who dropped out of the Simulation condition were low-ability students, students in the Simulation condition started out with higher pretest scores. So the trend is not reliable and these results are difficult to interpret.

Conclusions

It is noteworthy that all of the three versions of AutoTutor produced significant learning gains. However, we did not consistently find overall significant effects with the presence of computer simulations. This result is not surprising because mixed results for simulation efficacy have been a common finding throughout simulation literature, as reported in the Introduction.

We have speculated why we did not find a simulation effect. One potential problem might be that during the experiment, some students who learned using the simulations showed frustration and fatigue at a sometimes slow display of simulations and the continuous questions generated by the tutor. Learning gains were significant with these simulations, in spite of these problems. Moreover, the Simulation versions were better than the Monte Carlo AutoTutor that did not have LSA-based student modeling (at least for students with intermediate level of knowledge).

We are currently in the process of revising the simulation dialogues and improving the simulation environments. In the revised version, simulation dialogues are much more succinct and context-sensitive. The simulation introductions have been shortened and redundancy has been reduced. Improved simulation dialogues and a faster display of simulations will hopefully help AutoTutor produce deeper learning of abstract physics concepts. If successful, AutoTutor's interactive simulations will serve as a new medium for creating an immersive environment with visual embodiment of physics principles.

Acknowledgements

Research on AutoTutor was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428) and the DoD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600 (visit http://www.autotutor.org). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DoD, ONR, or NSF.

References

Baddelely, A. D. (1992). Working memory. *Science, 255*, 556-559.

Bourque, D.R., & Carlson, G. R. (1987). Hands-on versus computer simulation methods in chemistry. *Journal of Chemical Education, 64(3),* 232-234.

Choi, B., & Gennaro, E. (1987). The effectiveness of using computer simulated experiments on junior high students' understanding of the volume displacement concept. *Journal of Research in Science Teaching, 24(6),* 539-552.

Dekker, J., & Donatti, S. (1981). The integration of research studies on the use of simulation as an instructional strategy. *Journal of Educational Research, 74(6),* 424-427.

di Sessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10(2 & 3), 105-225.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual Coherence with Latent Semantic Analysis. *Discourse Processes*, 25, 285-307.

Gorsky, F., & Finegold, M. (1992). Effects of computer simulations and problem-solving approaches on high school students. *Journal of Computers in Mathematics and Science Teaching, 11,* 163-178.

Graesser, A.C., Chipman, P., Haynes, B., & Olney, A. (in press). AutoTutor: An intelligent-tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education*.

Graesser, A.C., Hu, X., & McNamara, D.S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A.F. Healy (Ed.), *Experimental cognitive psychology and its applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (pp. 183-194). Washington, D.C.: American Psychological Association.

Graesser, A.C., Person, N., Harter, D., & TRG (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial intelligence in Education, 12,*257-259.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Compute36*, 180-193.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialog. *AI Magazine, 22,* 39-51.

Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & Tutoring Research Group (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research, 1,* 35-51.

Hupper, J., Lomast, S.M., & Lazarowitz, R. (2002). Computer simulations in the high school:students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education, 24(8),* 803-821.

Kim, H. J., Graesser, A., Jackson, T. G., Pool, L. D., Sadkin, J, Ray, A., & Haynes, B. (Jan, 2005). Incorporating Interactive Simulations and Dialogues into Intelligent Tutoring Systems. Presented at the annual meeting of Winter Text and Cognition, Jackson, Wyoming.

Kintsch, W. (1998). *Comprehension; A paradigm for cognition*. New York: Cambridge University Press.

Landaur, T., & Dumais, S. (1997). An answer to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 2111-2400.

Landaur, T., & Dumais, S. (1997). An answer to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 2111-2400.

Lee. J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media,26(1),*71-85.

Michael, K. Y. (2001). The effect of a computer simulation activity versus a hands-on activity on product creativity in technology education. *Journal of Technology Education, 13(1),* 21-43.

Person, N. K., Graesser, A.C., Bautista, L., Mattews, E.C., & the Tutoring Research Group (2001). Evaluating student learning gains in two versions of AutoTutor. In J.D. Moore, C.L. Redfield, & W.L. Johnson (Eds.) *Artificial intelligence in education: AI-ED in the wired and wireless future* (pp. 286-293). Amsterdam, IOS Press.

Ploetzner, R., & VanLehn, K. (1997). The acquisition of informal physics knowledge during formal physics training. *Cognition and Instruction, 15*(2), 169-206.

Rieber, L. P. (1993). A pragmatic view of instructional technology. In K. Tobin (Ed.). *The practice of constructivism in science education* (pp. 193-212). Hillsdale, NJ: Lawrence Erlbaum.

Rivers, R. H., & Vockell, E. (1987). Computer simulations to stimulate scientific problem solving. *Journal of Research in Science Teaching, 24(5),* 403-415.

Roberts, N., & Blakeslee, G. (1996). The dynamics of learning in a computer simulation environment. *Journal of Scicne Teacher Education, 7(1),* 41-58.

Strangman, N., Hall, T., Meyer, A. (2003). Virtual reality and computer simulations and the implications for UDL implementation: Curriculum Enhancements Report. http://www.cast.org/ncac/downloads/VR_URL.pdf.

VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., & Rose, C.P. (2005). When is reading just as effective as one-on-one interactive tutoring? . In B. Bara (Ed.), *Proceedings of the 27th Annual Meetings of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press

Woodward, J., Carnine, D., & Gersten, R. (1988). Teaching problem solving through computer simulations. *American Educational Research Journal, 25(1),* 72-86.

Zhang, J., Chen, Q., & Reid, D. (2000). http://www.ifip.or.at/con2000/iceut2000/iceut10--05.pdf.