

AutoTutor 3-D Simulations: Analyzing Users' Actions and Learning Trends

G. Tanner Jackson (gtjacksn@memphis.edu)

Department of Psychology, University of Memphis
Memphis, TN 38152 USA

Andrew Olney (aolney@memphis.edu)

Department of Computer Science, University of Memphis
Memphis, TN 38152 USA

Arthur C. Graesser (a-graesser@memphis.edu)

Department of Psychology, University of Memphis
Memphis, TN 38152 USA

Hyun-Jeong Joyce Kim (kimh@rhodes.edu)

Department of Psychology, Rhodes College
Memphis, TN 38112 USA

Abstract

Research in simulations and their educational efficacy has had mixed success in the educational and cognitive sciences. There are challenges in understanding the complex nature of simulations, critical points of learner interactivity with computer simulations, and appropriate methods for testing and discovering potential benefits. The current research investigated learning from AutoTutor, an intelligent tutoring system, that interacts with learners in natural language and that launches embedded interactive 3-D simulations for tutoring conceptual physics. An experiment on college students shed light on conditions that promote learning, with results that hopefully will scale up in diverse educational settings with learning technologies.

Keywords: intelligent tutoring systems, simulations

Introduction

Researchers have recently developed interactive simulation environments with the hopes that they provide benefits similar to real world experiences and that they allow for exploration in a broad landscape of hypothetical situations. Although simulations might have an intuitive appeal, research has provided conflicting results regarding their pedagogical effectiveness. Some studies have shown that simulations are an effective means of teaching (Brant, Hooper, & Sugrue, 1991; Carlsen & Andre, 1992; Goldstone & Son, 2005; Kinzie, Strauss, & Foss, 1993; Stockburger, 1982) while others have shown little or no positive results (Rieber & Wayne, 1992; Schlechter, Bessemer, & Kolosh, 1992). These conflicting results may in part be due to various methodological flaws: lack of appropriate control conditions, difficulty of subject matter, poor experimental design, and varying levels of user control.

Thomas and Hooper (1991) reported that the effects of simulations are typically not revealed through direct tests of

knowledge, but rather they can be found through tests of transfer and application. This may help to explain some of the large discrepancies between simulation studies. There are currently no standard methods for testing simulation effectiveness, which causes problems when trying to generalize or interpret results.

Some lines of research have focused on systematic differences between simulations. Goldstone and Son (2005) reported that the content representation within simulations had a significant impact on simulation effectiveness for learning science principles. They manipulated the level of content abstraction, as well as the order of presentation, while students attempted to learn the concept of competitive specialization. Ultimately they found that simulations were most effective when they started as relatively concrete representations of real world scenarios and faded into more abstract/idealized simulations where principles and concepts were emphasized. This progression led to the best performance during training, as well as the most robust transfer of the underlying scientific principles.

Many of the simulation environments incorporate some form of practice, where the users can take control and change objects in the simulations however they see fit. This approach allows users to regulate their own learning and adapt their actions to their unique conceptual understanding. One down side to this is that research on self-regulated learning (Azevedo & Cromley, 2004) has shown that students do not spontaneously engage in appropriate meta-cognitive strategies to track their progress during learning. However, when students are introduced to these effective learning strategies, their subsequent performance increases. This may explain why several simulation environments have failed: students do not know the proper learning strategies.

Some researchers have tried to counteract this naivety of student learning by incorporating forms of learning guides into their technology. Rivers and Vockell (1987) compared

a variety of simulations which allowed users to openly practice with the environments, but differed on the level of guidance provided. This research consisted of fifteen different simulations which were categorized into two groups: guided and unguided. The researchers found an overall positive effect for those students who received guidance during the simulation environments. These results, along with the research in self-regulated learning, seem to suggest that some form of guidance may be necessary to help students utilize simulations effectively.

The present research was conducted on interactive simulations embedded in an intelligent tutoring system called AutoTutor, which will be described below. AutoTutor helps students learn by holding a conversation in natural language. The subject matter was conceptual physics on Newton's laws of motion. AutoTutor launches interactive simulations whenever students express misconceptions or miss critical physics principles while solving conceptual physics problems (example problem: "Suppose a runner is running in a straight line at constant speed. He throws a pumpkin straight up. Where will it land? Explain.") These simulation environments allow users to practice openly on 3-D micro-worlds of entities in motion, but they also include dialog scaffolding with natural language that captures intelligent pedagogy. This initiative was designed to synthesize previous research from simulation environments with self-regulated learning strategies. The simulations, as well as the dialog, include misconception identification and remediation techniques. The purpose of the current analysis was to examine student performance within the new environments and to explore possible relations between users' actions and learning outcomes.

AutoTutor

AutoTutor is a natural language Intelligent Tutoring System (ITS) that has proven to be effective at producing learning gains (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; Jackson, Ventura, Chewle, Graesser, & TRG, 2004). The computational underpinnings of the AutoTutor system have been previously reported in a variety of outlets (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Lu et al., 2004; Graesser, VanLehn, Rose, Jordan, & Harter, 2001; Graesser, Wiemer-Hastings, Weimer-Hastings, Kreuz, & TRG, 1999), so only a brief overview of relevant components is presented.

The pedagogical strategies implemented within AutoTutor are based on several decades of research on human-to-human tutoring and reflect the pedagogical strategies of scaffolding for student-centered constructivist learning. At the beginning of each problem, AutoTutor poses a seed question to the student, the student provides an answer (about a paragraph is required for a complete answer), and AutoTutor pumps the student to supply more information. For the remainder of the problem, AutoTutor engages the student in natural language conversation and attempts to cover all the missing "problem-specific principles" (we call them *expectations*) from the student's original answer.

Once a problem has all expectations covered, AutoTutor moves on and presents the next problem seed question (or exits the session). While covering the expectations, AutoTutor also attempts to recognize student misconceptions and corrects these when they are manifested.

AutoTutor is personified as an animated agent with speech synthesis. Each conversational turn for AutoTutor is spoken, via the speech engine, as well as posted as text in a dialog history box. The animated agent makes use of lip synchronization, along with conversationally appropriate pedagogical gestures which range from focusing attention to various parts of the screen to incorporating emotions for feedback purposes. Below is an excerpt from an actual student dialog, with inserted dialog move identifiers in parentheses and brackets.

AutoTutor (SEED QUESTION): When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation?

Student: The passenger experiences a backward acceleration, while the car moves forward.

AutoTutor (PUMP): And can you add to that?

Student: This is because the passenger experiences a force opposite in direction to the force on the car.

AutoTutor (POSITIVE FEEDBACK): Um, yes, correct!

AutoTutor (HINT): When the neck provides a large force, what within the neck would get damaged?

Student: the spine

AutoTutor (NEUTRAL NEGATIVE FEEDBACK): umm

AutoTutor (PROMPT): I think you can get this. Damage within the neck will occur in the ligaments and the ? [*spoken with rising intonation and accompanied with a prompting hand gesture*]

Student: muscles

AutoTutor (POSITIVE FEEDBACK): Brilliant!

AutoTutor (PROMPT RESPONSE): ligaments and the muscles.

AutoTutor (DIALOG ADVANCER): Let's keep going.

Why/AutoTutor uses LSA as the backbone for representing world knowledge about conceptual physics, or any other subject matter that is tutored (Graesser, P. Wiemer-Hastings, K. Wiemer-Hastings, Harter, Person, & TRG, 2000). LSA is a high-dimensional, statistical technique that, among other things, measures the conceptual similarity of any two pieces of text, such as a word, sentence, paragraph, or lengthier document (Landauer, Foltz, & Laham, 1998). Why/AutoTutor uses LSA to perform conceptual pattern matching operations when we compare student contributions to expected good answers and to anticipated misconceptions. An expectation is considered covered if the student's contributions end up matching the expectation by some LSA threshold of overlap. Similarly, a misconception is considered present if the student's input matches the misconception by some LSA threshold.

AutoTutor 3-D

AutoTutor 3-D is the newest version of the AutoTutor system. It has the same pedagogical algorithms as previous versions, but the current version adds 3-D micro-worlds with interactive simulation. Previous research has already shown that the AutoTutor architecture is an effective learning environment when compared to ecological controls, such as reading a textbook for an equivalent amount of time (Graesser, Lu et al., 2004; Jackson et al., 2004). The current research examines the role of a new component, interactive simulation, in promoting learning of physics concepts.

Simulations The simulations contained in AutoTutor 3-D are aligned with each of the problems covered during the session. When a student is struggling through a problem, AutoTutor may decide to launch one of the relevant simulations (sometimes triggered after a series of incorrect statements made by the student and sometimes triggered by missing important principles). When a simulation is launched, the animated agent moves to the top-left corner of the interface and several windows fade into view (see Figure 1 for a screen shot).



Figure 1: Screen shot of an AutoTutor 3-D Simulation.

The first time a simulation loads, AutoTutor provides the student with an introduction to the new interface, along with a very brief description and question concerning the simulation scenario. On subsequent launches, AutoTutor only provides the brief statement and question of what may be involved in the current simulation scenario. Multiple simulations may be launched within a problem, each of which may focus on a particular physics principle or misconception. All principles and misconceptions have been inserted into AutoTutor using conversational phrasing. For example, one of the physics principles contained in AutoTutor is stated as, “After an object is dropped or thrown, the only force acting on it is gravity”, while an example misconception would be, “Heavier objects fall faster”.

After a simulation has loaded, AutoTutor asks a question and poses a challenge to the user. The challenges are designed to require the user to manipulate variables in order to either confirm or falsify a hypothesis that they create. The user may make any changes they desire, and then click the start button to set the effect of the parameter combinations on the 3-D micro-world in motion.

Within each problem, the simulation environment appears to remain the same each time it is loaded (the interface and parameter controls are constant), although the corresponding simulation dialog adapts to emphasize the current principle being discussed. Between the different physics problems, the simulation parameter components change to fit the scenario. For example, consider our example problem mentioned before where a runner in motion throws a pumpkin straight up and asks where it will land. In this simulation we include a runner and the pumpkin while we provide parameters that change the horizontal velocity of the runner, the mass of the pumpkin, the magnitude of gravity, etc. Another simulation involves a plane trying to drop a packet onto a target and asks if the packet will hit the target. The simulation for this scenario includes the plane, the packet, and the target, while providing parameters that change the horizontal velocity of the plane, the magnitude of gravity, the location of the target, mass of the packet, air resistance, etc. Between these problems the interface layout and design remains consistent, but the parameter labels change appropriately.

Each simulation has a set of parameters, each of which were judged (by physics experts) as being relevant or irrelevant. Relevant parameters are those which directly relate to the key concepts and principles for the problem. One example may be manipulating the parameter for horizontal velocity of a runner before he throws an object up in the air (showing that horizontal and vertical velocities are independent). Irrelevant parameters do not directly pertain to the principles, but may relate to a common misconception. For example, an irrelevant parameter, in a falling body problem, would be mass (as it relates to a misconception that heavier objects fall faster).

Overall, the simulation environments were designed to foster self-regulated and discovery learning. During active use of the simulations, AutoTutor takes a relatively hands-off approach, allowing the user to take most of the control. Users are allowed to practice with the simulations as many times as desired, and at times AutoTutor may make a suggestion or ask if the user would like to continue with the tutoring session. For each simulation AutoTutor holds a short dialog with the users and ultimately asks for them to verbally explain what actually occurred (using physics terms). The dialog during simulations was designed to scaffold the student’s conceptual knowledge of the simulation, and progress through varying levels of specificity and deeper levels of processing. At the culmination of each simulation dialog, AutoTutor provides an ideal answer/summary of the simulation topic. Not all

principles are explicitly addressed by AutoTutor, but all principles are at least indirectly covered during the session.

Current Experiment

We conducted an experiment that included 42 undergraduate students interacting with AutoTutor 3-D at the University of Memphis or Rhodes College. Some of the previous studies with AutoTutor involved students enrolled in physics classes. However this study, and other recent studies have incorporated physics novices. So, participants were recruited from subject pools at the various institutions and questioned about previous physics courses. All participants completed pretest, training, and posttest phases.

Pretests consisted of 26 multiple choice questions on conceptual physics that were pulled from, or similar to, the Force Concept Inventory (FCI). The FCI is a well-known and accepted test for assessing conceptual physics knowledge (Hestenes, Wells, & Swackhamer, 1992). These questions provide very short scenarios and require the student to apply their knowledge and select the correct solution.

The training phase consisted of working with AutoTutor 3-D through 4 conceptual physics problems. Training sessions typically lasted from an hour to an hour-and-a-half.

Immediately following the training sessions participants completed the posttest. The posttest also consisted of a 26 item multiple choice test similar to the FCI, which was counterbalanced with the pretest. Both the pretests and the posttests were administered independent from the AutoTutor program.

Results and Discussion

This was the first study conducted with AutoTutor 3-D, so we were primarily interested in analyzing the actions taken by the users along with any corresponding learning. Each manipulation of a simulation was logged into a database and later extracted for analysis. Analyses have been performed at the manipulation and at the problem units of analysis, however this study focuses on specific data patterns of students as a unit of analysis.

We selected two primary measures which were predicted to positively correlate with student learning: total number of simulations and the relevance of manipulations. It was expected that practice with more simulations would lead to higher learning outcomes. Therefore, the total number of simulations was selected for comparison with the learning indices. It was also hypothesized that those students who manipulated more relevant parameters would have a better understanding of the underlying conceptual principles and would therefore exhibit higher learning outcomes. Thus, the average manipulated parameter relevance was computed for each student and was compared to the learning measures. Although these analyses are correlational, they are relevant at this stage of simulation research which has produced a large number of null effects. If there are significant correlations, then we can turn to dissecting the precise causes of learning.

Four learning indices were included in the current study. Both the pretest and posttest proportions were included as

independent variables (i.e., proportion of questions that were answered correctly). A learning gains measure was computed by subtracting the pretest proportions from the posttest proportions. This is typically seen as a simple difference score that is often biased towards low ability students, as they have more ground to cover and consequently have higher gains scores. A fourth variable was computed which tries to account for this bias by computing a proportional gain. This particular variable is referred to as a proportional learning gain score and is computed with the following equation $[(\text{posttest proportion} - \text{pretest proportion}) / (1 - \text{pretest proportion})]$. The denominator calculates the amount of gain required to make a perfect score, while the numerator calculates the amount of actual gain for each student. The computed fraction then represents the proportional gain attained with less bias towards low pretest participants.

Correlation Results

Initial inspection of the data revealed that several participants did not receive any simulations and that some participants who actually received simulations did not interact with the system much, making very few manipulations. These two groups of students (no simulations and few manipulations) did not truly participate with the simulations, and therefore would not be expected to glean any significant learning from the experience. To account for this, the current analysis included only those participants who made at least ten total manipulations (across all four problems). Those students who made less than ten manipulations and who were excluded from these results did not significantly differ on pretest scores from those who remained in the following analyses.

Table 1 shows the results from the correlational analysis.

Table 1: Correlations with outcome measures for students with more than 10 manipulations (n=25).

Simulation Parameter	Pretest	Posttest	Learning Gains	Proportional gain Scores
Total number of simulations launched	-.047 (.82)	.299 (.15)	.420* (.04)	.508* (.01)
Parameter relevance	.145 (.49)	.530* (.01)	.440* (.03)	.466* (.02)

* = significant for two-tailed test

Table 1 shows that participants' prior knowledge did not significantly relate to the number of simulations or their selection of relevant parameters. So it appears that students who already possessed a conceptual understanding did not necessarily make better selections for relevant parameters. The students who ended up manipulating relevant parameters and those who received more practice with the simulations benefited the most from the experience. This is evidenced by the significant correlations between the

simulation variables and the posttest proportions, learning gains, and proportional gain scores.

Regression Results

We further examined those students who participated with the simulations by computing regression equations that predict posttest performance using the various simulation parameters.

Regression analyses were conducted to determine if the total number of simulations and the relevance of the manipulations could significantly predict posttest performance. Previous studies with AutoTutor in conceptual physics have found that pretest scores were often the largest significant predictor of posttest performance, so this variable was inserted first into the regression equation. We wanted to see if the two simulation variables would add any significant predictive power above and beyond the pretest. A series of analyses revealed that the pretest proportion and the mean parameter relevance were the only two consistently significant predictors in the regression equations, $r^2=.665$, $p=.002$. Table 2 shows the corresponding coefficients from the regression equation.

Table 2: Regression coefficients for posttest performance

Predictor Variable	Standardized Beta	Sig.
Pretest Proportion	.647	.00
Parameter Relevance	.366	.01
Total number of Simulations	.205	.12

As with previous studies, the pretest scores account for a large portion of the posttest variance. In this case it also appears that the relevance of manipulations can help to predict how well students will perform at posttest.

Conclusions and Future Directions

Results from the previous section have led us to believe that not all students utilize simulations equally. As with any self-regulated learning experiences, the degree of learning primarily lies within the hands of the student. In this case, the level of interaction was left up to the students, and the results indicate that those who utilized the situation effectively experienced more gains. This means that future work will need to be more active in helping engage the students with the simulations, and may need to include a brief introduction to the appropriate learning strategies (Azevedo & Cromley, 2004).

It appears that the mere presence of simulations (i.e., grounding the situation) does not help students' conceptual knowledge, and therefore simulation conditions as a whole should not necessarily be treated equally. The actions of each user should be taken into account, as they allow for a better representation of what the simulation environments can provide.

Unfortunately, no cause and effect relations can be directly attributed here, but this analysis does provide another

important step toward determining what factors are important when interacting with simulations. Hopefully work with other systems will begin to provide similar in-depth analyses which dissect simulation interactions, and allow for specific comparisons between systems. Future work with AutoTutor 3-D will continue to explore users' actions within simulation environments, and will be likely to incorporate previous results into new system designs and experimental manipulations. This precursory research helps to further define the complex landscape of simulation environments and helps to integrate a new line of research into an established technology.

Acknowledgments

The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of approximately 35 researchers from psychology, computer science, physics, and education (visit <http://www.autotutor.org>). The 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. 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

- Atkinson, M & Burton, J. S. (1991). Measuring the effectiveness of a microcomputer simulation. *Journal of Computer-Based Instruction*, 18(2), 63-65.
- Azevedo, R., Cromley, J.G. (2004) Does training on self-regulated learning facilitate students learning with hypermedia? *Journal of Educational Psychology*, 96(3). 523-535.
- Brant, G., Hooper, E. & Sugrue, B. (1991). Which comes first the simulation or the lecture?. *Journal of Educational Computing Research*, 7(4), 469-481.
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, 14(1), 69-110.
- Graesser, A.C., Chipman, P., Haynes, B.C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education*, 48, 612-618.
- 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 Computers*, 36, 180-193.
- Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22, 39-51.
- Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the TRG 1999. Auto Tutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, 1, 35-51.

- Graesser, A.C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N., and the TRG (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments*, 8, 129-148
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher* 30, 141-158.
- Hewitt, P. G. 1987. *Conceptual Physics*. Menlo Park, CA: Addison-Wesley.
- Jackson, G.T., Ventura, M.J., Chewle, P., Graesser, A.C., and the Tutoring Research Group (2004). The Impact of Why/AutoTutor on Learning and Retention of Conceptual Physics. . In J.C. Lester, R.M. Vicari, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems 2004* (pp. 501-510). Berlin, Germany: Springer.
- Jong, T. D. (1991). Learning and instruction with computer simulations. *Education & Computing*, 6, 215-227.
- Kinzie, M. B., Strauss, R., & Foss, J. (1993). The effects of an interactive dissection simulation on the performance and achievement of high school biology students. *Journal of Research in Science Teaching*, 30, 989-1000.
- Kulick, C., Kulick, J., and Cohen, P. (1980). Effects of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research*, 50, 525-544.
- Landauer, T.K., Foltz, P.W., Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.
- Lee, June. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International journal of instructional media*, 26(1), 71-85.
- Merrill, M. D. (1994). *Instructional design theory*. Englewood Cliffs, New Jersey: Educational Technology Publications.
- Rieber, L. P. & Kini, A. (1995). Using computer simulations in inductive learning strategies with children in science. *International Journal of Instructional Media*, 22(2), 135-144.
- Rivers, R- H. & Vockell, E. (1987). Computer simulations to stimulate scientific problem solving. *Journal of Research in Science Teaching*, 24(5), 403-416.
- Shlechter, T. M., Bessemer, D. W., & Kolosh, K. P. (1992). Computer-based simulation systems and role-playing: An effective combination for fostering conditional knowledge. *Journal of Computer-Based Instruction*, 19(4), 110-114.
- Stockburger, D. W. (1982). Evaluation of three simulation exercises in an introductory statistics course. *Contemporary Educational Psychology*, 7(4), 365-370.
- Thomas, R. & Hooper, E. (1991). Simulations: An opportunity we are missing. *Journal of Research on Computing in Education*. 23(4), 497-513.