Running head: COMPUTER LEARNING

Computer Learning Environments with Agents that Support Deep Comprehension and Collaborative Reasoning

Arthur C. Graesser, David Lin, and Sidney D'Mello

University of Memphis

Send correspondence to:

Art Graesser
Department of Psychology & Institute for Intelligent Systems
202 Psychology Building
University of Memphis
Memphis, TN 38152-3230
901-678-2146
901-678-2579 (fax)
a-graesser@memphis.edu

To appear in a book edited by M.T. Banich and D. Caccamise, *Generalization of Knowledge*. Mahwah, NJ : Erlbaum.

Abstract

During the last decade, learning scientists have developed technologies with animated pedagogical agents that interact with the student in natural language and other communication channels, such as facial expressions and gestures. These pedagogical agents model good learning strategies and coach the students in actively applying their knowledge. This chapter focuses on agent-based learning environments that attempt to facilitate deep comprehension (e.g., causal explanations, plans, logical justifications), reasoning in natural language, and inquiry (i.e., question asking, question answering, hypothesis testing). These agent-based learning environments have targeted high school and college students who learn about topics in science and technology. Tests of these systems have exhibited both successes and failures with respect to learning and generalization. One of these projects on AutoTutor has analyzed transfer at both course-grain and fine-grain levels. The course-gain assessments have examined whether working on physics problems facilitate the solutions for similar physics problems with different surface characteristics. The fine-grain assessments have tracked the mastery and application of particular principles (e.g., net force equals mass times acceleration) throughout the history of pretest, training, and posttest. In all of these projects with agents, learning and generalization were assessed with multiple tests, tasks, and criteria, as opposed to relying on a single measure or goal standard.

Computer Learning Environments with Agents that Support Deep Comprehension and Collaborative Reasoning

Animated conversational agents play a central role in some visions of advanced learning environments. Imagine a virtual world that has a close correspondence with everyday scenarios and problems, with human-like cyber agents that interact with students and help them learn by holding a conversation in natural language. The cyber agents may take on different roles: mentors, tutors, peers, players in multiparty games, or avatars in the virtual worlds. The students communicate with the agents through speech, keyboard, gesture, touch panel screen, or conventional input channels. In turn, the agents express themselves with speech, facial expression, gesture, posture, and other embodied actions. In essence, students and agents have face-to-face conversations in the context of authentic situations and problems. The students are highly engaged in their interactions with the agents in the virtual worlds because the system dynamically adapts to their cognitive, emotional, and motivational states. The entire system is also embedded in a game environment in which the students score points to the extent that their speech and actions reflect mastery of the material. These are serious games -- games that help the students acquire important academic and technical content.

This vision of virtual worlds with agents is not science fiction. Outstanding examples of virtual environments with agents are those developed at University of Southern California, funded by the U.S. Army, National Science Foundation, and other agencies. The *Mission Rehearsal* system (Gratch et al., 2001) has dozens of agents in a combat scenario. The user interacts with a commander at a war scenario in spoken language and tries to help the entire company solve a crisis. In *Tactical Iraqi* (Johnson & Beal, 2005), the

student speaks to soldiers and civilians in an Iraqi village that has been attacked and needs assistance. The student learns both the language and the culture during dozens of hours of training. They learn by holding spoken dialogues in natural language and by scoring points in a serious game with authentic situations and problems. These award-winning virtual environments are major milestones that could only be achieved by interdisciplinary teams of researchers working on projects over a sustained period of time.

The Mission Rehearsal and Tactical Iraqi systems have not yet been sufficiently tested on learning gains, transfer, and generalization. However, there are theoretical and empirical reasons for being optimistic that they will be successful. There are at least four major reasons:

(1) *Similarity helps transfer*. The higher the similarity between the training context and the test situation, the better the transfer of performance, representations, and strategies (Anderson, Reder, Simon, 1996; Bransford, Brown, & Cocking, 2000). The virtual realities are very rich perceptually and socially, and the training scenarios are carefully selected to have a close correspondence with real world problems.

(2) *Tutoring helps learning*. One-on-one human tutoring is among the most effective methods of helping students learn (Graesser & Person, 1994). Meta-analyses show learning gains of .42 sigma (effect size in standard deviation units) compared to classroom controls and other suitable controls (Cohen, Kulik & Kulik, 1982). There are many potential reasons for the effectiveness of one-on-one tutoring (Graesser, Person, & Magliano, 1994), most notably that the tutor adapts to the learner's cognitive states (Van Lehn, et al., in press) and emotions (Lepper & Henderlong, 2000).

(3) *Agents model good learning strategies.* Students learn from observing others who model good behavior (Bandura, 1986). Students rarely have the opportunity to observe other students exhibiting good learning strategies in the classroom and other typical settings in school systems. Agents not only enact these strategies but can also think aloud while they do so (McNamara, Levinstein, & Boonthum, 2004).

(4) *Games are engaging.* Games are extremely engaging so there will be an increase in training time if serious content is woven into the games (Gee, 2003; Johnson & Beal, 2005; Malone & Lepper, 1987). Time on task is of course predictive of learning gains and presumably transfer (Taraban, Rynearson, & Stalcup, 2001).

It is important to acknowledge that many forms of computer based training have been shown to facilitate learning. This gives us additional reasons to be optimistic about the effectiveness of virtual environments with agents. The outcome variables have varied widely in existing assessments of learning technologies. The assessments have included tests of retention for shallow knowledge, answers to questions that tap deep knowledge (e.g., causal explanations, justifications of claims), problem solving performance, and transfer of knowledge/skill to different but related contexts. Meta-analyses have revealed that computerized learning environments fare well compared to classroom instruction (Dodds & Fletcher, 2004; Wisher & Fletcher, 2004); the effect sizes (i.e., sigma, comparing treatment to control conditions) are .39 for conventional computer-based training, .50 for multimedia, and 1.08 for intelligent tutoring systems. At this point in the science, there is precious little data on learning gains from learning environments with virtual reality and serious games, so research is needed in these arenas. Learning gains are routinely reported in published

studies, but there often are incomplete data on usage (attrition), engagement (including how much the learners like the system), study time, system development time, and development costs. But just as important, from the standpoint of this edited volume, there has been insufficient attention to segregating test items that involve retention of explicit training material versus transfer of knowledge, skills, and strategies to new contexts. Therefore, the matter of generalization needs more focused analyses in all classes of learning technologies.

This chapter describes some of the agent-based learning technologies that we have developed in the Institute for Intelligent Systems at the University of Memphis. All of these projects have assessed these technologies with respect to learning gains and most of them have assessed transfer of knowledge. We will take a very close look at transfer in AutoTutor, an intelligent tutoring system that helps college students learn technical topics (such as physics) by holding a conversation in natural language. As we describe these projects, we will offer some conclusions about the status of transfer and generalization in learning technologies with agents.

Agent-based Learning Technologies

Embodied animated conversational agents have become increasingly popular in learning technologies (Atkinson, 2002; Baylor & Kim, 2005; Cole et al., 2003; Graesser, Jackson, & McDaniel, in press; Johnson, Rickel, & Lester, 2000; McNamara, Levinstein, & Boonthum, 2004; Moreno & Mayer, 2004; Reeves & Nass, 1996). These agents speak, point, gesture, walk, and exhibit facial expressions. Some are built in the image of humans, whereas others are animals or cartoon characters. The potential power of these agents, from the standpoint of LEs, is that they can mimic face-to-face communication with human

tutors, instructors, mentors, peers, or people who serve other roles. Single agents can model individuals with different knowledge, personalities, physical features, and styles. Ensembles of agents can model social interaction. Both single agents and ensembles of agents can be carefully choreographed to mimic virtually any activity or social situation: curiosity, inquiry learning, negotiation, interrogation, arguments, empathetic support, helping, and so on. Researchers can have precise control over what the agents say, how they say it, and what conditions trigger specific actions. Therefore, agent technologies are having a revolutionary impact on social science research.

Animated conversational agents could conceivably have a negative impact on learning. For example, the agent might create cognitive overload, a split attention effect, or a distraction from other information on the display that has higher importance (Moreno & Mayer, 2004). An agent might be so realistic that the student has too high of expectations on its intelligence (Norman, 1994; Shneiderman & Pleasant, 2005). Such concerns can only be mitigated by systematic empirical research that tests theoretical models that insect social psychology and the learning sciences. For example, we might explore what impact the personality and attractiveness of the agent has on learning and transfer. We can explore whether students learn best from agents that are similar to them. The results of the studies are sometimes counterintuitive. Available research suggests that it is the content of what is expressed, rather than the aesthetic quality of the speech or face, that is most important in predicting learning (Graesser, Moreno et al., 2003). Research suggests that it is possible to create social presence from simple facial icons with expressions (☺) -- a minimalist form of the persona effect (Reeves & Nass, 1996).

Researchers in the Institute for Intelligent Systems at the University of Memphis have developed several learning environments with animated conversational agents. The remainder of this chapter describes some of these systems, their goals, and learning outcomes in the empirical studies that have been conducted. We will hereafter refer to these agents as *animated pedagogical agents* because they were designed for the purpose of improving learning. Most of these agents have a direct pedagogical role as a tutor or instructor. Others are members of a social ensemble of agents, taking on the role as a student or conversational partner with a tutor/instructor. In either case, researchers designed the agents in a fashion that was theoretically motivated by pedagogical theories.

AutoTutor

AutoTutor was the first pedagogical agent developed and tested at the University of Memphis (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Weimer-Hastings et al., 1999). AutoTutor is an intelligent tutoring system that helps students learn by holding a conversation in natural language. AutoTutor's dialogues are organized around difficult questions that require reasoning and explanations. The primary method of scaffolding good student answers is through *expectation and misconception tailored dialogue*. Both AutoTutor and human tutors (Graesser et al., 1995) typically have a list of anticipated good answers (called *expectations,* e.g., force equals mass times acceleration) and a list of anticipated *misconceptions* associated with each main question. AutoTutor guides the student in articulating the expectations through a number of dialogue moves: *pumps* (what else?), *hints*, and *prompts* for specific information. As the learner expresses information over many turns, the list of expectations is eventually covered and the main question is

scored as answered. Another conversation goal is to correct the misconceptions that are manifested in the student's talk. When the student articulates a misconception, AutoTutor acknowledges the error and corrects it. Another conversational goal is to be adaptive to what the student says. AutoTutor adaptively responds to the student by giving short *feedback* on the quality of student contributions (positive, negative or neutral) and by *answering* the student's questions. The answers to the questions are retrieved from glossaries or from paragraphs in textbooks via intelligent information retrieval.

It is beyond the scope of this chapter to describe the mechanisms of AutoTutor that drive the conversation. It suffices to say that AutoTutor attempts to hold a mixed initiative dialogue that mimics the conversational patterns of human tutors. This is now possible by virtue of recent advances in computational linguistics (Jurafsky & Martin, 2000), statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007), and discourse processes (Graesser, Gernsbacher, & Goldman, 2003).

The pedagogical framework of AutoTutor was inspired by three bodies of theoretical, empirical, and applied research. The first is explanation-based constructivist theories of learning (Aleven & Koedinger, 2002; Chi, deLeeuw, Chiu, LaVancher, 1994; McNamara, 2004). These theories postulate that learning is more effective and deeper when the learner must actively generate explanations, justifications, and functional procedures than when they are merely given information to study. The second is intelligent tutoring systems that adaptively respond to student knowledge at a fine-grained level (Anderson et al., 1995; VanLehn, Lynch, et al., 2002). These tutors give immediate feedback to learner's actions and guide the learner on what to do next in a fashion that is sensitive to what the system

believes the learner knows. The third is empirical research that has documented the collaborative constructive activities that routinely occur during human tutoring (Chi et al., 2001, 2004; Fox, 1993; Graesser et al., 1995). The patterns of discourse uncovered in naturalistic tutoring were directly imported into the dialogue management facilities of AutoTutor.

One version of AutoTutor on introductory computer literacy covers the topics of hardware, the operating system, and the internet. Each of these topics has 6 challenging questions that required about a paragraph of information (3-7 sentences) in an ideal answer. The questions required answers that involved inferences and deep reasoning, such as *why, how, what-if, what if not, how is X similar to Y?*. An example question about the operating system is "When you turn on the computer, how is the operating system first activated and loaded into RAM?" A typical exchange has 50-100 turns to answer a single challenging question. In this version of AutoTutor, the students give spoken input whereas students typed their contributions into the keyboard in earlier versions. We use the commercially available Dragon Naturally Speaking ™ (version 6) speech recognition system for speech-to-text translation.

The AutoTutor interface has 3 major windows, as shown in Figure 1, in the version with speech recognition. Window 1 (top of screen) is the main question that stays on the computer screen throughout the conversation with the question. Window 2 (left middle) is the animated conversational agent that speaks the content of AutoTutor's turns. Window 3 (right middle) is either blank or has auxiliary diagrams. In addition to these interface

components, there are 2 buttons on the keyboard that the learner presses to start speaking and stop speaking.

**** INSERT FIGURE 1 ABOUT HERE ****

Each turn of AutoTutor in the conversational dialogue has three information slots (i.e., units, constituents). The first slot of most turns is feedback on the quality of the learner's last turn. This feedback is either positive (*very good, yeah*), neutral (*uh huh, I see*), or negative (*not quite, not really*). The second slot advances the interaction with either prompts for specific information, hints, assertions with correct information, corrections of misconceptions, or answers to student questions. The third slot is a cue for the floor to shift from AutoTutor as the speaker to the learner. For example, AutoTutor ends each turn with a question or a gesture to cue the learner to do the talking. Discourse markers (*and also, okay, well*) connect the utterances of these three slots of information.

The conversations managed by AutoTutor are sufficiently smooth that students can get through the session with minimal difficulties (Person & Graesser, 2002). In fact, the dialogue is sufficiently tuned so that a bystander who observes tutorial dialogue in print cannot tell whether a particular turn was generated by AutoTutor or by an expert human tutor of computer literacy (Person & Graesser, 2002). A series of studies were conducted that randomly sampled AutoTutor's turns. Half of the turns were generated by AutoTutor and half were substituted by a human expert tutor on the basis of the dialogue history. Bystander participants were presented these tutoring moves and asked to decide whether each was generated by a computer or a human. Signal detection analyses revealed that the bystanders had zero *d'* scores in making these discriminations. In this sense, AutoTutor

passed the bystander Turing test for individual tutoring turns. Of course, a bystander can eventually tell whether a sequence of turns was part of a dialogue with AutoTutor versus a human tutor. The dialogue is far from perfect because AutoTutor does not have the depth of language comprehensions that humans do. But AutoTutor is surprisingly close.

Versions of AutoTutor

Several versions of AutoTutor have been developed since 1997, when the initial NSF project was funded. Most versions of AutoTutor have animated conversational agents with facial expressions, synthesized speech, and gestures. These full versions have been compared with alternative versions with voice only, text only, and various combinations of modalities in presenting AutoTutor's dialogue messages (Graesser, Moreno et al., 2003). The full animated conversational agent has shown advantages in promoting learning over alternative modalities under some conditions, particularly for deeper levels of learning (Atkinson, 2002; Moreno, Mayer, Spires, & Lester, 2001). However, available research on AutoTutor suggests that it is the verbal content of the tutor's messages that most robustly explains learning gains (Graesser, Moreno et al., 2003).

A version of AutoTutor called *AutoTutor-3D* guides learners on using interactive simulations of physics microworlds (Graesser, Chipman et al., 2005; Jackson et al., 2006). For example, an example question in conceptual physics is "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?" For each of the physics problems, we developed an interactive simulation world in *3-d Studio Max*. The world included the people, objects, and spatial setting associated with the problem. The student can

manipulate parameters of the situation (e.g., mass of objects, speed of objects, distance between objects) and then ask the system to simulate what will happen. They can cognitively compare their expected simulated outcome with the actual outcome after the simulation is completed. Moreover, they are prompted to describe what they see. Their actions and descriptions are evaluated with respect to covering the expected principles in an ideal answer. In order to manage the interactive simulation, AutoTutor gives hints and suggestions, once again scaffolding the learning process with dialogue. Thus, AutoTutor combines interactive simulation with mixed-initiative dialog.

We are currently working on a version of AutoTutor that is sensitive to the student's emotions. AutoTutor is augmented with sensing devices and signal processing algorithms that classify affective states of learners. Emotions are classified on the basis of dialog patterns during tutoring, the content covered, facial expressions, body posture, and speech intonation (D'Mello et al., 2005, 2006). The primary emotions that occur during learning with AutoTutor are frustration, confusion, boredom, and flow (engagement), whereas surprise and delight occasionally occur (Graesser et al., in press). We plan on investigating whether learning gains and learner's impressions of AutoTutor are influenced by dialogue moves of AutoTutor that are sensitive to the learner's emotions. For example, if the student is extremely frustrated, then AutoTutor presumably should give a good hint or prompt that directs the student in a more positive learning trajectory. If the student is bored, AutoTutor should give more engaging, challenging, and motivating problems. If the student is very absorbed and satisfied, then AutoTutor should be minimally directive.

Learning Gains with AutoTutor

The learning gains of AutoTutor have been evaluated in 15 experiments conducted during the last 9 years. Assessments of AutoTutor on learning gains have shown effect sizes of approximately .8 standard deviation units in the areas of computer literacy (Graesser et al., 2004) and Newtonian physics (VanLehn, Graesser et al., in press). These evaluations place previous versions AutoTutor somewhere between an untrained human tutor (Cohen et al., 1982) and an intelligent tutoring system (Corbett, 2001). The assessments of learning gains from AutoTutor have varied between 0 and 2.1 sigma (a mean of .8), depending on the learning performance measure, the comparison condition, the subject matter, and the version of AutoTutor. Approximately a dozen measures of learning have been collected in these assessments on the topics of computer literacy and physics, including: (1) multiple choice questions on shallow knowledge that tap definitions, facts and properties of concepts, (2) multiple choice questions on deep knowledge that taps causal reasoning, justifications of claims, and functional underpinnings of procedures, (3) essay quality when students attempt to answer challenging problems, (4) a cloze task that has subjects fill in missing words of texts that articulate explanatory reasoning on the subject matter, and (5) performance on problems that require problem solving.

Assessments of learning uncovered a number of findings that were either provocative or very illuminating (see Graesser, Lu et al., 2004; VanLehn et al., in press).

(1) *AutoTutor versus reading a textbook*. Learning gains with AutoTutor are superior to reading from a textbook on the same topics for an equivalent amount of time. However, this gap gets smaller to the extent that the textbook content is restricted to

the content that directly corresponds to the problems covered by AutoTutor. Therefore, it is important to guide the learner's attention to the most relevant text when they read.

(2) *Reading a textbook versus doing nothing*. Learning gains are zero in both of these conditions when the tests tap deeper levels of comprehension. This provocative result is compatible with the results of comprehension calibration studies (Maki, 1998) that report a very low correlation (r = .27) between college students' perceptions of how well they are comprehending and their actual comprehension measured by objective tests. Readers need difficult problems that challenge their *illusions of comprehension* when they read at shallow levels; challenging problems encourage them to have deeper standards of comprehension.

(3) *AutoTutor versus expert human tutors*. One recent evaluation of physics tutoring compared learning gains of AutoTutor with the gains of accomplished human tutors via computer mediated communication. These learning gains were equivalent for students with a moderate degree of physics knowledge. In contrast the expert human tutors prevailed when the students had low physics knowledge and the dialogue was spoken.

(4) *Deep versus shallow tests of knowledge*. The largest learning gains from AutoTutor have been on deep reasoning measures rather than measures of shallow knowledge (e.g., definitions of terms, lists of entities, properties of entities, recognition of explicit content).

(5) *Zone of proximate development*. AutoTutor is most effective when there is an

intermediate gap between the learner's prior knowledge and the ideal answers of AutoTutor. AutoTutor is not particularly effective in facilitating learning in students with high domain knowledge and when the material is too much over the learner's head.

One way of analyzing the learning gains is to compare the normal conversational AutoTutor with different comparison conditions. We computed mean effect sizes for these contrasts on multiple choice questions that tapped deep reasoning. An example deep reasoning question in physics is presented below.

As a truck moves along the highway at constant speed, a nut falls from a tree and smashes into the truck's windshield. If the truck exerts a 1,000 N force on the nut, what is the magnitude of the force that the nut exerts on the truck?

a) 1,000 N

b) less than 1,000 N

c) N (the nut does not exert a force on the truck)

d) greater than 1,000 N (because the nut hit the truck, it exerts a greater force on the truck than the truck exerts on the nut)

The conversational AutoTutor has (a) a .80 effect size (sigma) compared with pretests, reading a textbook, or doing nothing, (b) a .22 sigma compared with reading text book segments directly relevant to the AutoTutor problems, (c) a .07 sigma compared with reading a script that answers the questions posed by AutoTutor, (d) a .13 sigma compared with AutoTutor presenting speech acts in print instead of the talking head, (e) a .08 sigma compared with expert human tutors in computer-mediated conversation, and (f) a -.20

sigma compared with a version of AutoTutor that is enhanced with interactive 3D simulations (i.e., the interactive simulations are better).

Transfer and generalization

Our definition of generalization consists of the application of principles, procedures, solutions to problems, and conceptual structures to relevant new situations. In essence, there is transfer from one set of learning activities to appropriate new activities. It is widely acknowledged in cognitive science that transfer and generalization can be very difficult or nearly impossible when the surface characteristics are different between training and transfer problems and when the correspondences are not highlighted (Forbus, Gentner & Law, 1995; Gick & Holyoak, 1980; Hayes & Simon, 1977). For example, Hayes and Simon's classical study showed that college students experienced zero transfer between successive problems that were solved when the problems were structurally isomorphic but varied in surface features. Gick and Holyoak (1980) reported that students needed instruction to focus on comparisons between Dunker's radiation problem and a story analogue before the story facilitated solving the problem. Gentner's research (this volume) has emphasized the importance of making explicit comparisons between problems before transfer is maximized.

We have conducted detailed analyses of transfer and generalization when students learn physics with AutoTutor (Van Lehn et al., in press). Our analyses have contrasted course-grain and fine-grain analyses. In course grain-analyses, there is a comparison between training problems and transfer problems that vary in surface similarity with the training problem. For illustration, Table 1 shows a comparison between a near transfer

problem and a far-transfer problem. The training problem would be predicted to facilitate the near-transfer problem more than the far-transfer problem, even though they have the same set of principles (expectations) and the same structure.

*** INSERT TABLE 1 ABOUT HERE ***

Our assessments of AutoTutor did not show facilitation for far transfer problems (Jackson, Ventura, Chewle, & Graesser, 2004; Van Lehn et al., in press) when compared with a *scripted minilesson* condition in which they read solutions to the same problems that are solved with AutoTutor. We also failed to show better facilitation in near-transfer than far-transfer problems. This appears to be bad news from the standpoint of having a learning environment that promotes far transfer. However, this conclusion appears to be limited to course-grain analyses, a comparatively insensitive method of assessing transfer. Differences emerged only when we performed fine-grained analyses.

The fine-grained analyses are involve partistic scoring of solutions rather than wholistic performance on the problem as a whole. Each problem has a set of expectations and misconceptions, as discussed earlier. Once the problems are decomposed in this fashion, we can pitch these expectations (and misconceptions) in a more generic form that allows comparisons between problems. There are more abstract principles that correspond to the expectations in the training, near-transfer, and far-transfer problems. For example, the following principles underlie the three problems in Table 1.

1. The magnitudes of the forces exerted by A and B on each other are equal.
2. If A exerts a force on B, then B exerts a force on A in the opposite direction.

3. The same force will produce a larger acceleration in a less massive object than a more massive object.

Similarly, there are the following abstract forms of the misconceptions.

1. A lighter/smaller object exert no force on a heavier/larger object.
2. A lighter/smaller object exerts less force on other objects than a heavier/larger object.
3. The force acting on a body is dependent on the mass of the body.
4. Heavier objects accelerate faster for the same force than lighter objects.
5. Action and reaction forces do not have the same magnitude.

After we decompose the problems into abstract principles and misconceptions, we can analyses how accurately the student performs on each of these decomposed units throughout the course of pretest, training, and posttest. We have indeed tracked 60 principles and misconceptions across the following events: 4 pretest essay questions, 26 multiple choice questions, 10 training problems (AutoTutor versus comparison conditions), 4 posttest essay questions, 4 far-transfer essays, and 26 multiple choice questions. Any given principle P_i is relevant to some but not all of these 74 events. Similarly, any given misconception M_j is relevant to some but not all of these events. We observe how well the individual student performs on each principle (or misconception) over the course of relevant events. For example, suppose there are 4 relevant events in each of the following 5 phases of testing and training: pretest essay, pretest MC, training, posttest essay, and posttest MC. The student receives a 1 if the behavior is correct in an event and a 0 if not correct. A student with the following history would exhibit all-or-none learning, with learning emerging in the third training problem:

All-or-none learning = [(0000)(0000)(0011)(1111)(1111)]

The following history would exhibit no learning because the likelihood of a correct response is only 25% in each of the 5 stages:

No learning = [(0010)(1000)(0100)(0100)(0010)]

The following history consists of variable learning because the learner does better, on the average, in the posttests than the pretests, but never ends up being consistently perfect towards the end.

Variable learning = [(0010)(1000)(0101)(1110)(1011)]

There is also refresher learning, when the learner gets reminded of the correct answers during pretests, as illustrated below.

Refresher learning = [(0001)(1111)(1111)(1111)(1111)]

It should be apparent that this partistic, fine-grained analysis offers a more precise framework for investigating learning processes and products.

VanLehn et al (in press) conducted a fine-grained analysis on the physics principles and misconceptions. They computed the proportion of relevant events in which correct performance was manifested for the specific principles and misconceptions. The good news was that (a) these fine-grained measures showed significant learning gains from the pretests to the posttests and (b) the AutoTutor condition had better scores at posttest than the control conditions when we performed a sign test on the means for multiple dependent measures. For example, the far-transfer principle proportions were .275 and .213 for AutoTutor and control conditions, respectively, whereas the corresponding far-transfer misconception proportions were .078 versus .095. The disappointing news, however, is that only a few of

the differences were statistically significant when we examined each dependent variable separately.

We are currently conducting additional fine-grained analyses on the particular principles and misconceptions. We are interested in the sequential patterns and consistency of performance over time and over contextual fluctuations. We are interested in what training or testing conditions influence particular principles and misconceptions. It may be that some classes of principles are amenable to learning by a conversational AutoTutor and some by 3D simulation, whereas other classes of principles are best acquired by reading or didactic instruction. We believe these detailed analyses have merit, but the proof of the pudding awaits further research.

The SEEK Web Tutor

Current standards for science education assume that *critical thinking* is an essential component for understanding in science (AAAS, 1993). Critical thinking about science requires learners to actively evaluate the truth and relevance of information, to think about the quality of information sources, to trace the likely implications of evidence and claims, and to ask how the information is linked to the learner's goals and larger conceptual frameworks (Halpern, 2002; Linn, Davis, & Bell, 2004). Critical thinking is needed to achieve deeper levels of learning that involve causal reasoning, integration of the components in complex systems, and logical justifications of claims. A *critical stance* presupposes that the quality of the information is potentially suspect and requires close scrutiny with respect to its truth, relevance, and other dimensions of quality. A critical stance toward scientific information is especially important in the internet age. The internet furnishes millions of web pages on any

topic imaginable, yet there is no control over the quality of the scientific information presented over the internet. Learners need a critical stance in their arsenal of self-regulated learning strategies (Azevedo & Crowley, 2004) in this age of information pollution from the public and media.

We developed a web tutor to scaffold the acquisition of a critical stance to science learning. The web tutor is called SEEK, an acronym for Source, Evidence, Explanation, and Knowledge (Graesser, Wiley et al., 2006; Wiley, 2001). The SEEK Tutor was designed to improve college students' critical stance while they search for information on the Internet. The learners search through web pages on the topic of plate tectonics. Some of the web sites are reliable information sources on the topic whereas others have erroneous accounts of earthquakes and volcanoes. The goal assigned to the students in our experiments was to search the web for the purpose of writing an essay on what caused the eruption of Mt. St. Helens volcano.

The SEEK Tutor fosters critical stance with three main facilities. First, there is a *Hint* button on the Google search engine page which contains suggestions on how to effectively guide students' search. This page was a mock Google page with titles and URL's for web sites. Half of the sites are reliable, including sites from NASA (National Aeronautics and Space Administration), PBS (Public Broadcasting Station), and Scientific American. Half are unreliable sites that explain volcanoes and earthquakes by appealing to the stars, the moon, and oil drilling. Whenever the learner clicks on the Hint button, there are spoken messages that give reminders of the goal of the task (i.e., writing an essay on the causes of the Mt. St. Helens) and suggestions on what to do next (i.e., reading web sites with reliable information).

One version of the SEEK Tutor has a talking head, but the studies we conducted were on a version that had voice only.

The second facility to foster critical stance are *Pop-up Ratings* that ask students to evaluate the expected reliability of the information in a site. Students provide a rating and a rationale for their rating. The Pop-up Rating and justification appear after the students first view a particular website for 20 seconds. The third facility consists of a *Pop-up Journal* that has 5 questions about the reliability of the site that the learner just visits. These questions were designed to address some of the core aspects of critical stance: *Who authored this site? How trustworthy is it? What explanation do they offer for the cause of volcanic eruptions? What support do they offer for this explanation? Is this information useful to you, and if so, how will you use it?*. Each question has a Hint button that can be pressed to evoke spoken hints (at least 20 auditory statements per question) to guide the learners on answering each question. The Pop-up Journal is launched whenever the learner exits one of the web sites. It forces the learner to think about each of the 5 core aspects of critical stance and also to articulate verbally the reasons for their ratings.

We conducted two experiments that evaluated the impact of the SEEK Tutor in acquiring a critical stance (Graesser, Wiley et al., 2006). College students explored the web sites for 50 minutes with the goal of writing an essay on the causes of the eruption of Mt. St. Helens. In the first experiment, the participants were randomly assigned to either the *SEEK Tutor* condition or to a *Navigation* condition that had no training on critical stance. We expected that the 50-minute training of the SEEK Tutor would be effective in enhancing a critical stance, influencing the learner's exploration of the web sites, evaluating the quality of

the web sites, learning the content of plate tectonics, and articulating the causes of the volcano in the essay. To the extent that the Tutor is effective, there should be better performance in the SEEK Tutor condition than the Navigation condition. On the other hand, it is also conceivable that much more training is needed before students can effectively plan, monitor, and strategically apply a critical stance to science learning. In a second experiment, we augmented these two conditions with a set of instructions and example web sites that more thoroughly described and illustrated critical stance in the context of an Atkins diet. In essence, our goal was to pack in as much training on critical stance as we could in approximately 70 minutes and compare it to a condition in which there was no training on critical stance.

We were surprised to learn that 70 minutes of intense training on critical stance had very little impact on college students, even when we assessed the impact of the SEEK Tutor (with instructions) on dozens of measures of study processes and learning. The SEEK Tutor did not improve learners' ability to detect reliable information sources, as manifested by their ratings and rank orderings of web sites on reliability. The Tutor had no impact on the amount of study time they allocated to reliable versus unreliable sites. The SEEK Tutor had no significant impact on a statement verification task in which they rated the truth/falsity of 30 statements about plate tectonics, including: true statements, false plausible statements, misconceptions, and ridiculous distracters. The essays college students wrote on the causes of the Mt. St. Helen's eruption did not have more core ideas and fewer misconceptions if they had the SEEK Tutor with Instructions than if they had no training at all on critical stance. Indeed, after assessing dozens of measures, there was only one measure that showed a benefit of the SEEK Tutor: Students had more expressions in the essay with language about causal

explanations (such as "cause" and "explanation") compared to controls. The Tutor+Instructions did affect the language in their essays, which is a reassuring manipulation check, but had virtually no influence on the learning processes and results.

It appears that there will need to be much more training and scaffolding from the SEEK Tutor before robust effects emerge on the application of critical stance to web learning. Simply put, very little can be accomplished in one hour of on-line training. One wonders whether 20 hours of the SEEK Tutor on multiple topics and problems would produce deep learners of science who have a more penetrating critical stance.

Human Use Regulatory Affairs Advisor (HURA Advisor)

HURAA is a comprehensive learning environment on the web with didactic lessons, a document repository, hypertext, multimedia (including an engaging video), lessons with concrete scenarios to assess case-based reasoning, query-based information retrieval, and an animated agent that serves as a navigational guide (Graesser, Hu, Person, Jackson, & Toth, 2004; Hu & Graesser, 2004). Trainees learn the U.S. federal policies and regulations on the ethical use of human subjects in research. The goals of HURAA were to train high-ranking military personnel on research ethics in a small amount of time (less than an hour) and to provide a repository of up-to-date information on research ethics that can be retrieved by learner questions.

The animated conversational agent, appearing in the upper left of the web page, serves as a navigational guide to the trainee. It makes suggestions on what to do next and answers the trainee's questions. Below the agent are the major learning modules. *Introduction* and *Historical Overview* provide didactic instruction, including an engaging

video. The *Lessons* module presents trainees case scenarios and they are to evaluate whether the cases violate one of 7 critical ethical issues (e.g., informed consent, favorable risk-benefit ratio, independent review). Thus, the trainees actively apply the didactic knowledge (which is potentially inert, Bransford et al., 2000) to case-based reasoning. *Explore Issues* and *Explore Cases* allow further training on particular cases and issues, whereas *Decision Making* provides further testing. *Search IRB Documents* allows the learner to search for information in a large repository of documents through natural language queries. Generally, speaking, the recommendation is for the trainees to proceed in the sequential order of these modules. However, the navigational guide can recommend skipping modules, depending on the trainee's cognitive profile. The more active learners can immediately access any of the modules at any point in the learning session. There also is a repository of information they can access in the lower left of the screenshot: *Glossary, Archives & Links, Bibliography*, and so on. All of these learning modules can be expanded as information accumulates over time.

Case-based scenario training is one of the foundational principles of learning in cognitive science and education (Ashley, 1990; Kolodner, 1993; Sweller & Cooper, 1985). Each scenario in HURAA provides a text description of a particular experiment that was problematic with respect to 1 or more of 7 ethical issues. Thus, there is a case scenario (S) by critical ethical issue (I) matrix that specifies whether each scenario S_i did or did not have a problem with ethical issue I_j. The trainees are asked to rate potential problems with these scenarios. The trainee's ratings are continuously being compared with the ideal SI matrix. Discrepancies between trainee decisions/actions and the SI matrix are used to guide

feedback to the learner (i.e., when there were false alarms and misses) and also to select the next scenario for training. Scenarios are dynamically selected in a fashion that optimizes the repair of their false alarms and misses.

HURAA was evaluated in experiments that contrasted it with conventional computer-based instruction containing the same content. There were two pieces of good news in evaluations of the system on over a dozen measures of retention, reasoning, and inquiry. First, memory for core concepts was enhanced by HURAA compared to the conventional web software; the effect sizes varied between .56 and 1.19 sigma (mean = .78)(Hu & Graesser, 2004). Second, HURAA's answers to learner questions in the information retrieval facilities were impressive; 95% of the answers were judged as relevant by the learner and 50% were judged as being informative (Graesser, Hu et al., 2004). However, HURAA had no significant increment for many measures compared with the control condition. In particular, there was no improvement in case-based reasoning, as measured by the accuracy in identifying ethical problems in new cases and in the response time in making these judgments. There was no improvement in the speed of accessing information when trainees were given difficult questions that required information search. There was no improvement in the trainees' perceptions of the system with respect to interest, enjoyment, amount learned, and ease of learning.

Our evaluation of the HURA Advisor leads to some similar conclusions as our evaluation of the SEEK Tutor. One hour of training is not sufficient to train adults on reasoning strategies and applying their knowledge to new cases. It clearly takes substantially more training than one hour to actively transfer one's knowledge to new

situations. In contrast, one hour of training did show improvements in the articulation of causal language in the SEEK Tutor and the recall of key concepts in the HURA Advisor. These results underscore the importance of assessing learning, retention, and transfer with multiple measures. The results also suggest that extensive training is needed before we can expect changes in reasoning, strategies, and application of knowledge to new situations.

Ensembles of Agents: iSTART and iDRIVE

Ensembles of agents can be designed to exhibit good learning strategies and social interactions. It is very difficult, if not impossible, to train teachers and tutors to apply specific pedagogical techniques. This is because years of normal social interaction and conversation are often incompatible with ideal pedagogical methods (Person, Kreuz, Zwaan, & Graesser, 1995). However, it is possible to design pedagogical agents to have such precise forms of interaction. And of course, the agents do not get weary, irritated, and snippy. Researchers at the University of Memphis have designed two systems in which students learn by observing and interacting with ensembles of agents: iSTART and iDRIVE.

iSTART (Interactive Strategy Trainer for Active Reading and Thinking) is an automated strategy trainer that helps students become better readers by constructing self-explanations of the text (McNamara et al., 2004). It uses groups of animated conversational agents to scaffold these strategies. The primary goal of iSTART is to help adolescent and college students learn meta-comprehension strategies that support deeper comprehension while they read. It combines the power of self-explanation in facilitating deep learning (Chi et al., 1994) with content-sensitive, interactive strategy training (Palincsar & Brown, 1984). iSTART interventions teach readers to self-explain using five

reading strategies: *monitoring comprehension* (i.e., recognizing comprehension failures and the need for remedial strategies), *paraphrasing* explicit text, making *bridging inferences* between the current sentence and prior text, making *predictions* about the subsequent text, and *elaborating* the text with links to what the reader already knows.

The animated agents of iSTART provide three phases of training. The *Introduction Module* provides instruction on self-explanation and reading strategies. There is a trio of animated agents (an instructor and two students) who cooperate with each other, provide information, pose questions, and provide explanations of the reading strategies. After the presentation of each strategy, the trainees complete brief multiple-choice quizzes to assess their learning. Next comes the *Demonstration Module*, where two Microsoft Agent characters (Merlin and Genie) demonstrate the use of self-explanation in the context of a science passage and the trainee identifies the strategies being used. The final phase is *Practice*, where Merlin coaches and provides feedback to the trainee while the trainee practices self-explanation using the repertoire of reading strategies. For each sentence in a text, Merlin reads the sentence and asks the trainee to self-explain it by typing a self-explanation. The system interprets the trainee's contributions using recent advances in computational linguistics. Merlin gives feedback and sometimes asks the trainee to modify unsatisfactory self-explanations.

Studies have evaluated the impact of iSTART on both reading strategies and comprehension for thousands of students in K12 and college (McNamara, Best, O'Reilly, & Ozuru, 2006). The three-phase iSTART training (approximately 3 hours) has been compared with a control condition that didactically trains students on self-explanation, but

without any vicarious modeling and any feedback via the agents. After training, the participants are asked to self-explain a transfer text (e.g., on heart disease) and are subsequently given a comprehension test in the form of open-ended questions, short answer questions, and multiple choice tests. The results have revealed that strategies and comprehension are facilitated by iSTART, with impressive effect sizes (1.0 sigma or higher) for strategy use and for comprehension. Therefore, after approximately 3 hours of training, we do begin to see some impact on the mastery and application of comprehension strategies.

The results have revealed that the facilitation by iSTART depends on world knowledge and general reading ability. For example, readers with low prior knowledge of reading strategies benefit primarily at the level of the explicit textbase, whereas those with high prior knowledge of reading strategies benefit primarily on tests of bridging inferences. These findings are in line with Vygotsky's theory of zone of proximal development (Vygotsky, 1978), as we discovered in our research with AutoTutor. iSTART can help students to achieve a level of comprehension that is closest to their proximal level of development, or the highest level they can achieve with appropriate scaffolding.

The iDRIVE (Instruction with Deep-level Reasoning questions In Vicarious Environments) system has duets of animated agents train students to learn science content by modeling deep level reasoning questions in question-answer dialogues. A student agent asks a series of deep questions about the science content and the teacher agent immediately answers each question. There is evidence that learning improves when learners have the mindset of asking deep questions (*why, how, what-if, what-if-not*) that tap causal structures,

complex systems, and logical justifications (Craig, Gholson, Ventura, & Graesser, 2000; Driscoll et al., 2003; King, 1994; Mayer, 2005; Rosenshine, Meister, & Chapman, 1996). However, the asking of deep questions does not come naturally (Graesser, McNamara, & VanLehn, 2005; Graesser & Olde, 2003) so the process needs to be modeled by agents or humans. The iDRIVE system models the asking of deep questions with dialogues between animated conversational agents. Learning gains on the effectiveness of iDRIVE on question asking, recall of text, and multiple-choice questions have shown effect sizes that range from .56 to 1.77 compared to a condition in which students listen to the monologue on the same content without questions.

Closing Comments

Animated pedagogical agents are destined to have a major impact on learning environments of the future. Researchers have only begun to scratch the surface on their potential. Individual agents can have an endless number of dialogue styles, strategies, personalities, and physical features. They can be matched to the cognitive, personality, emotional, and social profiles of individual learners in an endless number of ways. The agents can exhibit the activities of good learners in addition to the activities of good teachers. There are also an endless number of agent ensembles that can be choreographed to implement promising theories of social interaction. The agents can tirelessly train learners for hundreds of hours on many topics and in many contexts. This is apparently necessary, according to the research presented in this chapter, for an adequate transfer and generalization of knowledge and strategies. Very little is accomplished in a 1-hour training session.

The role of technology in education and training has had its critics. In particular, Cuban (1986, 2001) documented that technology has historically had a negligible impact on improvements in education. He pointed out that radio and television did not had much of an impact on education, even after initial hopes and promises that they would revolutionize the educational landscape. Cuban has similarly argued that computers have had a negligible impact on education, in part because teachers have not adequately integrated them into the curriculum. It is our contention, however, that agent-based learning technologies will ultimately prevail and revolutionize the educational enterprise. Agents are like humans, and humans provide the best education that we know. Available empirical research supports the claim that some agents are almost as good as accomplished human tutors. And who knows what the future may bring as the science of pedagogical agents evolves. The agents just might end up being better.

References

AAAS (American Association for the Advancement of Science) (1993). *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147-179.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167-207.

Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25*, 5-11.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*, 416-427.

Ashley, K.D. (1990). *Modeling legal argument: Reasoning with cases and hypotheticals*. Cambridge, MA: MIT Press.

Azevedo, R., & Cromley, J.G. (2004). Does training on self-regulated learning faciliate students' learning with hypermedia. *Journal of Educational Psychology,* 96, 523-535.

Bandura, A. (1986). Social foundations of thought and action: A social cognition theory. Engelwood Cliffs, NJ: Prentice-Hall.

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*, 95-115.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn*. Washington, D.C.: National Academy Press.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Chi, M. T. H., Siler, S. A., Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction, 22*, 363-387.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25,* 471-533.

Cohen, P. A., Kulik, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19,* 237-248.

Cole, R. van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D. Ward, W., & Yan, J. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human computer interaction. *Proceedings of the IEEE*, *91*, 1391-1405.

Corbett, A.T. (2001). Cognitive computer tutors: Solving the two-sigma problem. *User Modeling: Proceedings of the Eighth International Conference, UM 2001, 137-147.*

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., & the Tutoring Research Group (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education, 11*, 242-253.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

D'Mello, S.K., Craig, S.D., & Graesser, A.C. (2006). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education, 16,* 3-28.

Dodds, P., & Fletcher, J. (2004). Opportunities for new "smart" learning environments enabled by next-generation web capabilities. *Journal of Educational Multimedia and Hypermedia, 13*, 391-404.

Driscoll, D.M.,Craig, S.D., Gholson, B., Ventura, M.,Hu, X., & Graesser, A.C. (2003). Vicarious learning: Effects of overhearing dialog and monolog-like discourse in a virtual tutoring session. *Journal of Educational Computing Research, 29*, 431-450.

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205.

Fox, B. (1993). *The human tutorial dialogue project*. Hillsdale: Erlbaum.

Gee, J. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gick, M.L., & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

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., D'Mello, S.K., Craig, S.D., Witherspoon, A., Sullins, J., McDaniel, B., & Gholson, B. (in press). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research*.

Graesser, A.C., Gernsbacher, M.A., & Goldman, S. (2003)(Eds.). *Handbook of discourse processes*.

Mahwah, NJ: Erlbaum.

Graesser, A.C., Hu, X., Person, P.,Jackson, T., and Toth, J (2004). Modules and information retrieval facilities of the Human Use Regulatory Affairs Advisor (HURAA). *International Journal on eLearning, 3*, 29-39.

Graesser, A.C., Jackson, G.T., & McDaniel, B. (in press). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology*.

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., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225-234.

Graesser, A.C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves deep learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, and J. Kay (Eds.), *Proceedings of Artificial Intelligence in Education*. (pp, 47-.54). Amsterdam: IOS Press.

Graesser, A.C., & Olde, B.A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology,* 95, 524-536.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104-137.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*, 359.1-28.

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

Graesser, A.C., Wiley, J., Goldman, S.R., O'Reilly, T., Jeon, M., & McDaniel, B. (in press). SEEK Web Tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning*.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems, 17*, 54-63.

Halpern, D.F. (2002). *An introduction to critical thinking* (4th edition). Mahwah, NJ: Erlbaum.

Hayes, J.R., & Simon, H.A. (1977). Psychological differences among problem isomorphs. In J. Castellan, D.B. Pisoni, & G. Potts (Eds.), *Cognitive theory, vol. 2*. Hillsdale, NJ: Erlbaum.

Hu, X., & Graesser, A.C. (2004). Human Use Regulatory Affairs Advisor (HURAA): Learning about research ethics with intelligent learning modules. *Behavioral Research Methods, Instruments, and Computers, 36*, 241-249.

Jackson, G.T., Olney, A., Graesser, A.C., Kim, H.J. (2006). AutoTutor 3-D Simulations: Analyzing user's actions and learning trends. In R. Son (Ed.), *Proceedings of the 28th Annual Meetings of the Cognitive Science Society*. (pp. 1557-1562). Mahwah, NJ: Erlbaum.

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.

Johnson, W.L., & Beal, C. (2005). Iterative evaluation of a large-scale intelligent game for language learning. In C. Looi, G. McCalla, B. Bredeweg, and J. Breuker (Eds.), *Artificial Intelligence in Education: Supporting learning through intelligent and socially informed technology* (pp. 290-297). Amsterdam: IOS Press.

Johnson, W. L., Rickel, J., and Lester, J. (2000). Animated Pedagogical Agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*, 47-78.

Jurafsky, D., & Martin, J.H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

King A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31, 338-368.

Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufman, San Mateo, CA

Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.) (2007). *Handbook on Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257-307). San Diego, CA: Academic Press.

Linn, M. C., Davis, E. A. & Bell, P. (Eds.)(2004). *Internet environments for science education*. Mahwah, NJ: Erlbaum.

Maki, R.H. (1998). Test predictions over text material. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 117-144), Mahwah, NJ: Erlbaum.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Cognitive Science, 5*, 333-369.

Cognitive Science, 5(4), 333-369. Mayer, R.E. (2005). *Multimedia Learning*. Cambridge, MA: Cambridge University Press.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1-30.

McNamara, D.S., Levinstein, I.B. & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers, 36*, 222-233.

McNamara, D. S., O'Reilly, T., Best, R. & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research. 34*, 147-171.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165-173.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, 177-213.

Norman, D. A. (1994). How might people interact with agents? *Communication of the ACM, 37*(7), 68-71.

Palincsar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, *1*, 117-175.

Person, N.K., Graesser, A.C., & the Tutoring Research Group (2002). Human or computer?: AutoTutor in a bystander Turing test. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems 2002* (pp. 821-830). Berlin, Germany: Springer.

Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. Cognition and Instruction, 13, 161-188.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, televisions, and new media like real people and places*. Cambridge, U.K.: Cambridge University Press.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66,* 181-221.

Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: Strategies for effective human-computer interaction* (Ed. 4). Reading, MA: Addison-Wesley.

Sweller, J., & Cooper, M. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Taraban R., Rynearson, K., & Stalcup, K. A. (2001). Time as a variable in learning on the world-wide web. *Behavior Research Methods, 33*(2), 217-225.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (in press). When are tutorial dialogues more effective than reading? *Cognitive Science*.

VanLehn, K., Lynch, C., Taylor, L., Weinstein, A., Shelby, R. H., Schulze, K. G., et al. (2002). Minimally invasive tutoring of complex physics problem solving. In S. A. Cerri, G. Gouarderes & F. Paraguacu (Eds.), *Intelligent Tutoring Systems, 2002, 6th International Conference* (pp. 367-376). Berlin: Springer.

Vygotsky, L.S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

White, B. Y & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction, 16(1)*, 1998, 3-118.

Wiley, J. (2001). Supporting understanding through task and browser design. *Proceedings of the Twenty-third annual Conference of the Cognitive Science Society*, (pp. 1136-1143). Hillsdale, NJ: Erlbaum.

Wisher, R.A., & Fletcher, J.D. (2004). The case for advanced distributed learning. *Information & security, 14*, 17-25.

Author Notes

The research on AutoTutor was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918), the Institute of Education Sciences (R305H050169), 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 NSF, IES, DoD, or ONR.

Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230, a-graesser@memphis.edu.

Table 1. *Training, near-transfer, and far-tranfer physics problems*.

Training problem

If a lightweight car and a massive truck have a head-on collision, upon which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion? Defend your answer.

Near transfer problem

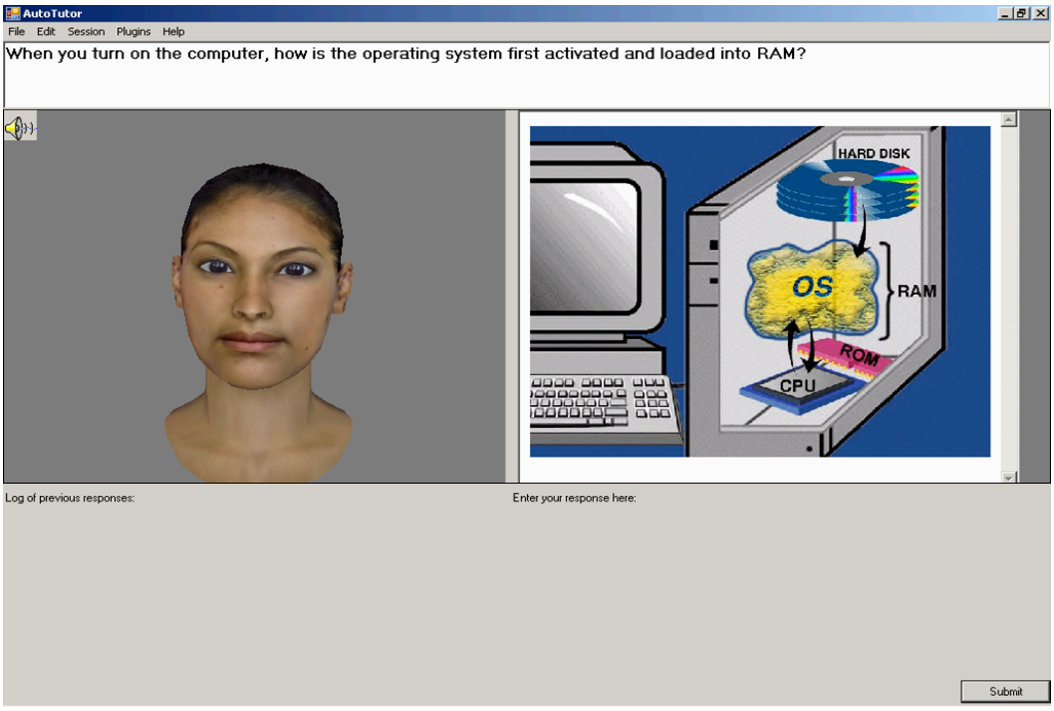
A huge oceanliner traveling due east collides with a small yacht, traveling due west. During the collision, the front end of the yacht is smashed in (causing the yacht to sink and the passengers to evacuate to their lifeboat). The oceanliner merely suffered a dent. What is true of the relationship between the force of the oceanliner on the yacht and the force of the yacht on the oceanliner?

Far-transfer problem

A 30-kg child receives her first "A+" on a spelling test and, overcome with joy, jumps up and down on her 200-kg desk. This desk is very strong and does not move while the child jumps on it. Does the child exert a force on the desk? Does the desk exert a force on the child? Justify both your answers.

Figure Caption Page

Figure 1: Interface of AutoTutor.




Form1
Google
Searched the web for causes volcanic eruptions
Results 1 - 7 of about 17,900 Search took 0.20 seconds
Volcanic Eruptions and T
Danger Days
Savage Earth
thirteen
Rate this site
Please rate this site on a scale of 1 to 6. Where 6 is most reliable:
Please give your reasons for this rating?
Submit
Savage Earth
Who authored this site?
How trustworthy is it?
What explanation do they offer for the cause of volcanic eruptions?
What support do they offer for this explanation?
Is this information useful to you? If so, how will you use it?
HINT
Complete
Reliability: 5
Reasoning: I think this site is fairly reliable because it seems more factual based