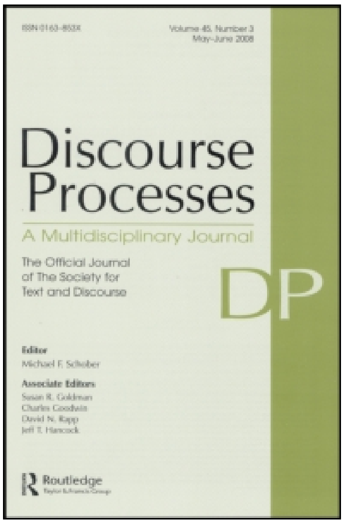

Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653637

Agent Technologies Designed to Facilitate Interactive Knowledge Construction

Arthur C. Graesser [a]; Moongee Jeon [a]; David Dufty [a]
[a] Department of Psychology, University of Memphis,

Online Publication Date: 01 July 2008

To cite this Article Graesser, Arthur C., Jeon, Moongee and Dufty, David(2008)'Agent Technologies Designed to Facilitate Interactive Knowledge Construction',Discourse Processes,45:4,298 — 322

To link to this Article: DOI: 10.1080/01638530802145395

URL: http://dx.doi.org/10.1080/01638530802145395

Discourse Processes, 45:298–322, 2008

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/01638530802145395

Routledge
Taylor & Francis Group

Agent Technologies Designed to Facilitate Interactive Knowledge Construction

Arthur C. Graesser, Moongee Jeon, and David Dufty
Department of Psychology
University of Memphis

During the last decade, interdisciplinary researchers have developed technologies with animated pedagogical agents that interact with the student in language and other communication channels (such as facial expressions and gestures). These pedagogical agents model good learning strategies and coach the students in actively constructing knowledge during learning. This article describes computer technologies that have been developed during the last decade with tutors 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 tutors target high school and college students who learn about topics in science and technology. The primary example is AutoTutor, a system on the Internet that helps students compose answers to deep-reasoning questions and solutions to problems by holding a conversation. AutoTutor's dialogue moves include *feedback* (positive, neutral, and negative), *pumps* for more information ("Tell me more."), *hints, prompts* to fill in missing words, *summaries, corrections* of student misconceptions, and *answers* to student questions. Other learning technologies with agents include the Human Use Regulatory Affairs Advisor (HURAA); Source, Evidence, Explanation, and Knowledge (SEEK) Web Tutor; Interactive Strategy Trainer for Active Reading and Thinking (iSTART); Instruction with Deep-level Reasoning questions In Vicarious Environments (iDRIVE); and Acquiring Research Investigative and Evaluative Skills (ARIES). These systems have been tested on their effectiveness in facilitating knowledge construction. They also have uncovered insights on the prospects of designing agents to effectively communicate in language and discourse.

Correspondence concerning this article should be addressed to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152–3230. E-mail: a-graesser@memphis.edu

Animated conversational agents play a central role in some of the recent advanced learning environments (Atkinson, 2002; Baylor & Kim, 2005; Graesser, Chipman, Haynes, & Olney, 2005; Johnson, Rickel, & Lester, 2000; McNamara, Levinstein, & Boonthum, 2004; Moreno & Mayer, 2004; Reeves & Nass, 1996). These agents interact with students and help them learn by either modelling good pedagogy or by holding a conversation. The 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. Intelligent agents with speech recognition essentially hold a face-to-face, mixed-initiative dialogue with the student, just as humans do (Cole et al., 2003; Graesser, Jackson, & McDaniel, 2007; Gratch et al., 2002; Johnson & Beal, 2005). Single agents model individuals with different knowledge, personalities, physical features, and styles. Ensembles of agents model social interaction. These systems are major milestones that could only be achieved by advances in discourse processing, computational linguistics, learning sciences, and other fields.

From the standpoint of learning, there are at least two fundamental reasons why these agents would be effective in facilitating knowledge construction. First, it is well documented that one-to-one tutoring is one of the most effective methods of helping students learn. Meta-analyses show learning gains from non-expert human tutors of 0.42 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-to-one tutoring (Graesser, Person, & Magliano, 1995), but most researchers attribute the facilitation to the tutors' adapting to the learners' cognitive states (Anderson, Corbett, Koedinger, & Pelletier, 1995; VanLehn et al., 2007) and emotions (Graesser et al., 2007; Lepper & Henderlong, 2000). Second, agents can demonstrate (i.e., model) good learning strategies. Students rarely have the opportunity to observe other students exhibiting good learning strategies in the classroom and other typical settings in school systems. 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. Agents not only enact these strategies, individually or in groups, but can also think aloud while they do so (McNamara et al., 2004).

This article describes some of the conversational agents that have been developed at the University of Memphis in the interdisciplinary Institute for Intelligent Systems. We take a close look at AutoTutor, an intelligent tutoring system that helps college students learn technical topics by holding a conversation in language. As we describe these projects, we offer some conclusions about the status of learning and communication in learning technologies with these pedagogical agents.

AutoTutor

AutoTutor was the first pedagogical agent developed and tested at the University of Memphis (Graesser, Chipman, et al., 2005; Graesser, Weimer-Hastings, Weimer-Hastings, Kreuz, & the Tutoring Research Group [TRG], 1999). AutoTutor is an intelligent tutoring system that helps students learn through tutorial dialogue in language. AutoTutor's dialogues are organized around difficult questions and problems that require reasoning and explanations in the answers. For example, the following are two example challenging questions from two of the subject matters that get tutored: Newtonian physics and computer literacy:

> PHYSICS QUESTION: 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, and why?
>
> COMPUTER LITERACY QUESTION: When you turn on the computer, how is the operating system first activated and loaded into RAM?

These questions require the learner to construct approximately three to seven sentences in an ideal answer and to exhibit reasoning in natural language. These are hardly the fill-in-the-blank questions or multiple-choice questions that many associate with learning technologies on computers. It takes a conversation to answer each one of these challenging questions. The dialogue for one of these challenging questions typically requires 50 to 200 conversational turns between AutoTutor and the student.

Table 1 illustrates AutoTutor with a conversation that was extracted from an actual tutoring session. This session was with a relatively verbose, knowledgeable student about physics, so the conversation is comparatively short. Most students are not highly verbal and knowledgeable about physics, so it takes many more conversational turns before the student can construct a good answer. When students are asked these challenging questions, their initial answers are typically only one or two sentences in length. However, one or two sentences provide insufficient information to adequately answer the question, so tutorial dialogue is needed to flesh-out a complete answer. AutoTutor engages the student in a mixed-initiative dialogue that draws out more of what the student knows and that assists the student in the construction of an improved answer. Therefore, how does AutoTutor converse with the student? We need to examine the structure and planning of the dialogue in more detail to unveil how AutoTutor systematically steers the interaction.

TABLE 1
Example Dialogue With AutoTutor

TUTOR: Now let's move on. A lightweight car and a massive truck have a head-on collision. On which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion? Explain why.

STUDENT: The lightweight vehicle experiences the greater impact and the greater change in motion because the mass of the lighter vehicle is less, therefore having less inertia to resist the change in motion. The momentum of the truck is greater than that of the car.

TUTOR: Right! What does Newton's third law of motion say about the forces that bodies exert on each other?

STUDENT: The forces must be equal in magnitude and opposite in reaction.

TUTOR: Yes, correct! Can you add to that?

STUDENT: The forces must be equal and opposite reactions acting on opposite bodies.

TUTOR: Ya! We've almost covered it, but there's one part I think is still missing. I am sure you can get this. During an interaction, each body exerts an equal but opposite force on the other ______________?

STUDENT: Body.

TUTOR: Yes, right! On the other body. So when two balls, one plastic and the other iron, hit each other, does the heavier or neither ball exert the greater impact force on the other?

STUDENT: Neither.

TUTOR: Yes, correct! Neither, the forces are equal in magnitude. Let's take a moment for general physics questions. Ask me one and I'll do my best to answer it.

STUDENT: No.

TUTOR: Alright. To recap what we have just discussed, please give me your complete answer to the original question, as presented at the top of the screen.

STUDENT: Although the lighter car experiences a greater change in motion, both vehicles exert an equal impact force, but opposite in direction. The larger object has a larger mass, and therefore a larger inertial resistance than that of the car.

TUTOR: That is right! Putting everything together, um, according to Newton's third law of motion, if one body exerts a force on the other body then the other body must exert an equal and opposite force on the first body. Therefore, the car and the truck both experience the same impact force. By Newton's second law, the car will experience the greater change in its motion because its mass is much smaller.

This question is finished so let's go on to the next question.

The Nature of Dialogue With AutoTutor

Many experts of language and discourse processes would be profoundly sceptical that a computer could hold its own in managing a conversation with a human. After all, it is widely acknowledged that computers have limitations in the depth of their comprehension; in building a model of the human's knowledge, beliefs, and intentions; in responding to unexpected human contributions; and in dynamically formulating novel contributions that adapt to the learner (Walker et al., 2003). We, of course, recognize these limitations.

However, it is also apparent that a well-designed computer tutor can manage a reasonably smooth and pedagogically effective conversation. There are reasons for being optimistic. One reason is that the common ground (shared knowledge; Clark & Brennan, 1991; Schober & Clark, 1989) between the student and tutor is low or modest, so the student does not assume it completely understands the computer and frequently cannot identify communication errors and misalignments (Pickering & Garrod, 2004). Expectations on grounding are also lowered when the computer tutor admits it is not perfect and expresses that it does not always understand the student (which also applies to human tutors, of course). A second reason for optimism is that there have been major advances in computational linguistics (Jurafsky & Martin, 2000), statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007), and discourse processes (Graesser, Gernsbacher, & Goldman, 2003). These advances have had a noticeable impact on the performance of many automated components of language and discourse processing. A third reason is that naturalistic human tutoring is a conversational register with properties that afford automation, as is clarified later. In contrast, automated conversations would be a disaster in other conversational registers that require high common ground and precision. We would not advocate an AutoSpouse or AutoPhysician, for example.

The structure of the dialogue in both AutoTutor and human tutoring (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Graesser et al., 1995; Shah, Evens, Michael, & Rovick, 2002) can be segregated into three levels or aspects: (a) expectation- and misconception-tailored dialogue, (b) a five-step dialogue frame, and (c) composition of a conversational turn. These three levels can be automated and produce respectable tutorial dialogue.

Expectation- and misconception-tailored dialogue. This is the primary pedagogical method of scaffolding good student answers. Both AutoTutor (Graesser et al., 2005) and human tutors (Graesser et al., 1995) typically have a list of *expectations* (anticipated good answers) and a list of anticipated *misconceptions* associated with each main question. For example, expectations E1 and E2 and misconceptions M1 and M2 are relevant to the example physics problems:

- E1. The magnitudes of the forces exerted by A and B on each other are equal.
- E2. If A exerts a force on B, then B exerts a force on A in the opposite direction.
- M1. A lighter/smaller object exerts no force on a heavier/larger object.
- M2. Heavier objects accelerate faster for the same force than lighter objects.

AutoTutor guides the student in articulating the expectations through a number of dialogue moves: pumps (what else?), hints, and prompts for the student to fill in missing words. Hints and prompts are carefully selected by AutoTutor to produce content in the answers that fill in missing content words, phrases, and propositions. For example, a hint to get the student to articulate expectation E1 might be, "What about the forces exerted by the vehicles on each other?" This hint would ideally elicit the answer, "The magnitudes of the forces are equal." A prompt to get the student to say "equal" would be, "What are the magnitudes of the forces of the two vehicles on each other?" As the learner expresses information over many turns, the list of expectations is eventually covered, and the main question is scored as answered. Complete coverage of the answer requires AutoTutor to have a pool of hints and prompts available to extract all of the content words, phrases, and propositions in each expectation. AutoTutor adaptively selects those hints and prompts that fill missing constituents and thereby achieves pattern completion.

AutoTutor is dynamically adaptive to the learner in other ways than coaching them to articulate expectations. There is the conversational goal of correcting misconceptions that arise in the student's talk. When the student articulates a misconception, AutoTutor acknowledges the error and corrects it. There is the conversational goal of giving feedback to the student on their contributions. AutoTutor gives short feedback on the quality of student contributions: positive (very good, bravo), negative (not quite, almost), or neutral (uh huh, okay). AutoTutor accommodates a mixed-initiative dialogue by attempting to answer the student's questions. The answers to the questions are retrieved from glossaries or from paragraphs in textbooks via intelligent information retrieval. AutoTutor asks a counter-clarification question (e.g., I don't understand your questions, so could you ask it in another way?) when it does not understand the student's question.

Five-step dialogue frame. This dialogue frame is prevalent in human tutoring (Graesser & Person, 1994; VanLehn et al., in press) and is implemented in AutoTutor. The five steps of the dialogue frame are as follows:

1. Tutor asks main question.
2. Student gives initial answer.
3. Tutor gives short feedback on the quality of the student's answer in Step 2.
4. Tutor and student collaboratively interact via expectation- and misconception-tailored dialogue.
5. Tutor verifies that the student understands (e.g., Do you understand?).

Students often answer that they understand in Step 5, when most do not. A good tutor would press the student further by asking more questions to verify the students' understanding, but even good tutors rarely do this, unfortunately. Most tutors end up giving a summary answer to the main question and then select another main question. A good tutor would ask the student to provide the summary (as in the example dialogue in Table 1) rather than it being provided by the tutor, but even good tutors rarely do that.

Managing one conversational turn. Each turn of AutoTutor in the conversational dialogue has three information slots (i.e., units, constituents). The first slot of most turns is short feedback on the quality of the student'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 coverage of the ideal answer with either prompts for specific words, hints, assertions with correct information, corrections of misconceptions, or answers to student questions. The third slot is a cue to the student for the floor to shift from AutoTutor as the speaker to the student. 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 within a turn.

The three levels of AutoTutor go a long way in simulating a human tutor. AutoTutor can keep the dialogue on track because it is always comparing what the student says to anticipated input (i.e., the expectations and misconceptions in the curriculum script). Pattern matching operations and pattern completion mechanisms drive the comparison. These matching and completion operations are based on latent semantic analysis (Landauer et al., 2007) and symbolic interpretation algorithms (Rus & Graesser, 2006) that are beyond the scope of this article to address. AutoTutor cannot interpret student contributions that have no matches to content in the curriculum script. This, of course, limits true mixed-initiative dialogue; that is, AutoTutor cannot explore the topic changes and tangents of students as the students introduce them. However, available studies of naturalistic tutoring (Chi, Siler, & Jeong, 2004; Chi et al., 2001; Graesser et al., 1995) reveal that (a) human tutors rarely tolerate true mixed-initiative dialogue with students changing topics that steer the conversation off course; and (b) most students rarely change topics, rarely ask questions, and rarely take the initiative to grab the conversational floor. Instead, it is the tutor that takes the lead and drives the dialogue. AutoTutor and human tutors are very similar in these respects.

The conversations managed by AutoTutor are hardly perfect, but are smooth enough for students to get through the sessions with minimal difficulties. 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, & the TRG, 2002). A series of studies were conducted that randomly sampled AutoTutor's turns. One half of the turns were generated by AutoTutor, and one half were substituted by a human expert tutor on the basis of the dialogue history. Bystander participants were presented these tutoring moves in a written transcript 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 successfully passed the "bystander Turing test" for individual tutoring turns. Not surprisingly, however, 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 sufficient depth of language comprehension and global coherence. Nevertheless, AutoTutor is close enough to human tutorial dialogue to keep the conversation going and also to promote learning.

A more critical analysis of the AutoTutor dialogue points to four major limitations that require improvement. One problem lies in errors in interpreting the content of student turns. In an ideal world, the pattern matching operations between student contributions and expectations would be perfect, but errors invariably occur. AutoTutor's evaluation of whether an expectation (or misconception) is expressed by a student is significantly correlated with the evaluation of experts ($r = .50$; Olde, Franceschetti, Karnavat, Graesser, & the TRG, 2002) and almost as high as the correlation between two experts ($r = .63$), but the correlation is far from perfect. Sometimes, when such errors occur, the students get frustrated and conclude that the tutor is not listening. This interpretation problem can be mitigated to some extent by improving the depth of the interpretation modules, including some facilities for inferences and entailment (Rus & Graesser, 2006).

A second problem consists of misclassification of the speech acts in student turns. The student turns are segmented into speech acts and each speech act is assigned to one of approximately 20 speech act categories. These categories include assertions, questions in 16 different categories, short responses (yeah, right), meta-cognitive expressions (I don't understand, I see), and meta-communicative expressions (What did you say?). The accuracy of classifying the student speech acts into categories varies from .87 to .96 (Olney et al., 2003), which is almost, but not quite, perfect. The dialogue coherence breaks down when some misclassification errors occur, which ends up confusing students. More efforts are needed to improve the speech act classification accuracy and to manage the dialogue to minimize exposure of unwanted consequences (Shah et al., 2002).

A third problem is that AutoTutor does not build on those student contributions that fail to match any expectation or misconception. AutoTutor is not building on what the student is expressing, so the student may conclude that AutoTutor is unresponsive. This, of course, is a major limitation in the mixed-initiative dialogue capabilities of AutoTutor. AutoTutor may never be able to interpret unexpected input from scratch at a deep level, but it could conceivably be fortified with generic dialogue moves (e.g., Tell me more about X; What is the relation between X and Y?) that encourage the student to elaborate on what they are trying to express. AutoTutor could periodically weave-in the main expectations of the curriculum script as the student expresses such tangential elaborations. It should be noted, nevertheless, that human tutors also fail to meaningfully respond to student contributions that are not on their content radar (see Chi et al., 2004).

A fourth problem occurs when the AutoTutor does not generate relevant and informative answers to the student questions. AutoTutor can handle only approximately one half of the student questions, so one half of AutoTutor's replies are either incorrect, constitute requests for clarification (I don't understand your question, so could you rephrases it?), or pass the burden on to the student (That's a good question, so how would you answer it?). The incidence of student questions quickly diminishes when the student concludes the system fails to provide informative answers (Graesser, McNamara, & VanLehn, 2005; Van der Meij, 1987). Improvements in the question-answering facilities are needed to minimize this fourth problem.

Versions of AutoTutor

We have created many versions of AutoTutor that were designed to incorporate particular pedagogical goals and cover different topics. So far, the topics have covered computer literacy, physics, biology, tactical planning, and critical thinking. Our first version of AutoTutor covered introductory computer literacy including the topics of hardware, the operating system, and the Internet. Each of these topics had 12 challenging questions that required deep reasoning, such as *why, how, what if, what if not, how is X similar to Y*? In most versions of AutoTutor, the students type in their contributions via keyboard, whereas recent versions allow spoken input. We use the commercially available Dragon Naturally Speaking™ (Version 6) speech recognition system for speech-to-text translation.

In most versions of AutoTutor, the interface has the three major windows shown in Figure 1. 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. When the students type in their contributions, there is a window at the bottom that echoes what the student types in. In versions with speech recognition, there are two buttons on the keyboard that the learner presses to start speaking and stop speaking. The interface sometimes includes a dialogue window that presents the history of the turn-by-turn tutorial dialogue for the challenging questions; the students can scroll back as far as they want in this dialogue history, although very few students pursue this dialogue recovery.

Several versions of AutoTutor have been developed since 1997, when the system was created. Most versions of AutoTutor have animated conversational

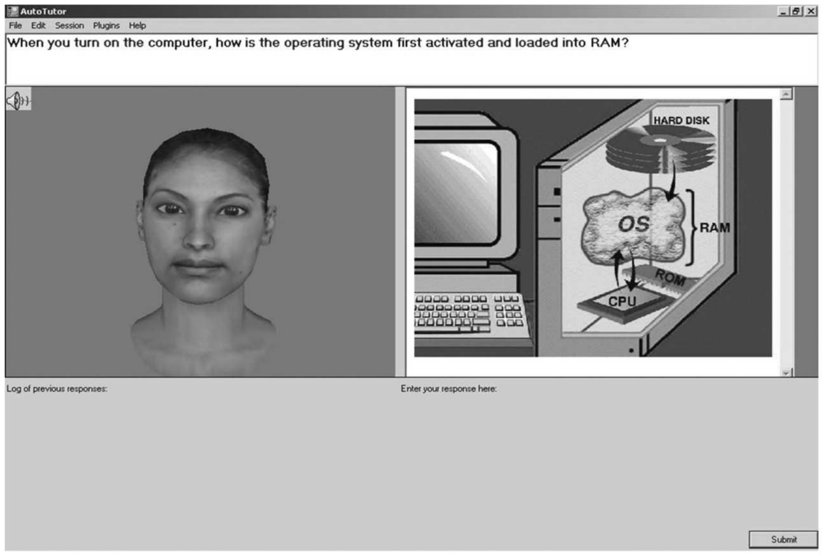

FIGURE 1 Interface of AutoTutor.

agents with synthesized speech, a small number of facial expressions, and some rudimentary hand and head 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). Stated differently, it is the content that matters rather than the presentation medium.

One version of AutoTutor, called *AutoTutor-3D,* guides learners on using interactive simulations of physics microworlds (Graesser et al., 2005; Jackson, Olney, Graesser, & Kim, 2006). For each of the physics problems, we developed an interactive simulation world with people, objects, and the spatial settings associated with the problem. The student manipulates parameters of the situation (e.g., mass of objects, speed of objects, distance between objects) and then asks the system to simulate what will happen. Students are also prompted to describe what they see. Their actions and descriptions are evaluated with respect to covering the expectations or matching misconceptions. AutoTutor manages the dialogue with hints and suggestions that scaffold the learning process with dialogue.

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 dialogue patterns during tutoring, the content covered, facial expressions, body posture, and speech intonation (D'Mello, Craig, & Graesser, 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., 2008). The accuracy of the computer classifying emotions on the basis of dialogue history is not perfect, but hovers around 65% to 70%, depending on the emotion, when 50% is chance in binary decisions of whether a particular emotion does or does not occur (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008). These percentages are expected to increase with the addition of the other channels of communication; namely, facial expressions (Ekman, 2003), body posture (Kapoor & Picard, 2005), and speech intonation (Litman & Forbes-Riley, 2004). It should be noted that trained human judges are not much more reliable in classifying these emotions.

The next step in this research is to explore 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. The emotions exhibited by AutoTutor is also an important consideration, just as it is for human tutoring (Lepper & Henderlong, 2000). Should AutoTutor be empathetic to a frustrated student or be earnest, forceful, or upbeat? Answers to such questions await future research.

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 0.8 standard deviation units in the areas of computer literacy (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, et al., 2004) and Newtonian physics (VanLehn et al., in press). These evaluations place previous versions of AutoTutor somewhere between an untrained human tutor (Cohen et al., 1982) and an intelligent tutoring system with ideal tutoring strategies (Corbett, 2001). The assessments of learning gains from AutoTutor have varied between 0.0 and 2.1 sigma ($M = 0.8$), depending on the learning performance measure, the comparison condition, the subject matter, and the version of AutoTutor. Approximately one dozen measures of learning have been collected in these assessments on the topics of computer literacy and physics, including (a) multiple-choice questions on shallow knowledge that tap definitions, facts, and properties of concepts; (b) multiple-choice questions on deep knowledge that taps causal reasoning, justifications of claims, and functional underpinnings of procedures; (c) essay quality when students attempt to answer challenging problems; (d) a cloze task that has participants fill in missing words of texts that articulate explanatory reasoning on the subject matter; and (e) performance on problems that require problem solving.

Assessments of learning in these various conditions have uncovered a number of findings that are illuminating, if not provocative (Graesser, Lu, et al., 2004; Jackson et al., 2006; VanLehn et al., 2007):

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.
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* (Glenberg, Wilkinson, & Epstein, 1982) 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 (Graesser, Lu, et al., 2004; Jackson et al., 2006; VanLehn et al., 2007). The conversational AutoTutor has (a) a 0.80 effect size (sigma) compared with pretests, reading a textbook, or doing nothing; (b) a 0.22 sigma compared with reading text book segments directly relevant to the AutoTutor problems; (c) a 0.07 sigma compared with reading a script that succinctly answers the questions posed by AutoTutor; (d) a 0.13 sigma compared with AutoTutor presenting speech acts in print instead of the talking head; (e) a 0.08 sigma compared with expert human tutors in computer-mediated conversation; and (f) a −0.20 sigma compared with a version of AutoTutor that is enhanced with interactive 3D simulations (i.e., the interactive simulations are better).

OTHER LEARNING ENVIRONMENTS WITH AGENTS

The AutoTutor project stimulated a large number of other projects with agents in the interdisciplinary Institute for Intelligent Systems at the University of Memphis. Most of the conversational agents were talking heads or full body agents, whereas others included only the voice because there was a worry that the face or body would distract the learner from the material. There is some evidence in research on multimedia that visual content with a voice-only narrator has advantages in promoting learning gains (Mayer, 2005; Whittaker, 2003). Most of the conversational agents guide the interaction with the student continuously, but some have agents communicate messages only under specific conditions or when the student asks for the help. The conditions under which information should be provided to the user of a system is far from being resolved in research on learning environments, surveys, and human–computer interaction in general (Conrad, Schober, & Coiner, 2007; Schober & Conrad, 2002).

Human Use Regulatory Affairs Advisor (HURAA)

The next system with an agent we developed was HURAA, 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 learned the U.S. policies on the ethical use of human participants in research. HURAA was designed to train high-ranking military personnel on research ethics in a small amount of time (less than 1 hr) and to provide a repository of up-to-date information on research ethics that could be retrieved by trainee questions.

The animated conversational agent of HURAA appeared in the upper left of the Web page and served as a navigational guide to the trainee. The trainees were military personnel, so the persona of the agent was an amalgamation between Colin Powell and a doctor with a white lab coat. The agent made suggestions on what to do next and answered the trainee's questions. Below the agent were labels for the major learning modules that spanned a broad array of learning technologies: didactic lessons to be read, engaging videos, hypertext, case scenarios to be evaluated on ethics, question-asking facilities, and a large repository of documents. The exact nature of these learning technologies is not directly relevant to this article, which is primarily concerned with the conversational agents. It was the agent that guided the learner to complete tasks that matched their level of knowledge and cognitive profile.

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 0.56 and 1.19 sigma ($M = 0.78$; Hu & Graesser, 2004). Second, HURAA's answers to trainee 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 several measures compared with the control condition: case-based reasoning; the speed of accessing information when trainees were given difficult questions that required information search; and perceptions of the system with respect to interest, enjoyment, amount learned, and ease of learning. Another somewhat disappointing result was that there were no significant differences in any of the measures we collected when we conducted an experiment that compared the agent's messages being presented in different media (i.e., the full animated conversational agent, text-only, voice-only, and text + voice; Graesser, Ventura, et al., 2003). As with the studies of AutoTutor, it is the content of what the agent says in the conversation that matters rather than the medium of message delivery.

We were somewhat surprised at the feedback from some of the military personnel on the use of agents. The majority of the personnel were not keen on the agents when we showed prototypes of HURAA in 2001. The critics considered the agents useless, frivolous, or distracting. They recommended that the agent could be easily turned off. The defense community also disagreed on what the persona should be for the agent. Our attempts to have them consider the agent facility was not fortified by the experiment reported by Graesser et al. (2003) that the animated agent had no incremental impact on learning gains. The community was still suffering from the negative press on Microsoft's Clippy, the animated paper clip that annoyed so many users and that presented a setback to the advocates of agents. The problems with Clippy were that it barged in and interrupted users, that users could not easily get rid of it, and that Clippy was not engineered by researchers who had adequate expertise in discourse processes. Quite clearly, the dialogue management of a conversational agent is absolutely crucial to its success. It is also apparent that agents are not well received by some generations and cultures of potential users (Shneiderman & Plaisant, 2005).

SEEK (Source, Evidence, Explanation, and Knowledge) Web Tutor

Critical thinking about science requires learners to actively evaluate the truth and relevance of information, the quality of information sources, and the implications of evidence and claims (Halpern, 2002). 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 student who takes a *critical stance* considers the possibility that the truth, relevance, or quality of the information is potentially suspect. A critical stance toward scientific information is especially important in the Internet age, an era when there are millions of Web pages but no control over the quality of the scientific information.

The next agent we developed was a Web tutor to scaffold the acquisition of a critical stance to science learning. The Web tutor is called SEEK (Graesser et al., 2007; Wiley, 2001). The SEEK Web Tutor was designed to improve college students' critical stance while they search Web pages on the topic of plate tectonics. Some of the Web sites were reliable information sources on the topic, written by professionals in the National Aeronautics and Space Administration, the Public Broadcasting Station, and Scientific American. Others were erroneous accounts of earthquakes and volcanoes that appealed to the stars, the moon, and oil drilling. The student's goal in the experiments was to search the Web for the purpose of writing an essay on what caused the eruption of the Mt. St. Helen's volcano.

The SEEK Web Tutor fostered critical stance with three main facilities. The first was a "hint" button on the Google search engine page that contained suggestions on how to effectively guide the student's search. This page was a mock Google page with titles and URLs for reliable and unreliable Web sites, which could be accessed and explored. Whenever the student clicked on the hint button, there were spoken messages that gave reminders of the goal of the task (i.e., writing an essay on the causes of the Mt. St. Helen's volcano eruption in the state of Washington) and suggestions on what to do next (i.e., reading Web sites with reliable information). The agent in one version of the SEEK Web

Tutor had a talking head, but the studies we conducted were on a version that had voice only. The talking head was dropped because we worried that it would create a distraction or a split-attention effect (Kalyuga, Chandler, & Sweller, 1999) with the Web material to be learned

The second facility to foster critical stance was "pop-up ratings and justifications" that asked students to evaluate the expected reliability of the information in a site. The pop-up ratings and justifications appeared after the students first viewed a particular Web site for 20 s. The third facility consisted of a "pop-up journal" that had five questions about the reliability of the site that the learner just visited. 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 had a hint button that could be pressed to evoke spoken hints to guide the learners on answering each question. The pop-up journal was launched whenever the learner exited one of the Web sites. It forced the learner to think about each of the five core aspects of critical stance and also to verbally articulate the reasons for their ratings.

We conducted experiments that evaluated the impact of the SEEK Web Tutor in acquiring a critical stance (Graesser et al., 2007). College students explored the Web sites for approximately 1 hr with the goal of writing an essay on the causes of the eruption of Mt. St. Helen's. Participants were randomly assigned to either the *SEEK Web Tutor* condition or to a *Navigation* condition that had no training on critical stance. The 1-hr training with the SEEK Web Tutor was expected to enhance a critical stance, as assessed by over one dozen measures, including an essay on the causes of a volcano.

We were surprised to learn that 1 hr of intense training on critical stance had very little impact on college students, even when we assessed the impact of the SEEK Web Tutor on dozens of measures of study processes and learning. For example, the SEEK Web Tutor did not improve learners' ability to detect reliable information sources, the amount of study time they allocated to reliable versus unreliable sites, their judgments of the truth or falsity of 30 statements about plate tectonics, and the articulation of core ideas about plate tectonics in the essays. After assessing dozens of measures, there was only one measure that showed a benefit of the SEEK Web Tutor: Students had more expressions in the essay with language about causal explanations (such as "cause" and "explanation") compared to controls. Therefore, the SEEK Web Tutor did influence the causal language in their essays, which is a reassuring manipulation check, but had virtually no influence on the learning processes and results.

There will need to be much more training and scaffolding from the SEEK Web Tutor before robust effects emerge on the application of critical stance to Web learning. Very little can be accomplished in 1 hr of training of new strategies. Perhaps dozens of hours of SEEK Web Tutor on multiple topics and problems are needed before benefits are realized for deep science learning and the application of a critical stance. Perhaps critical stance cannot be trained even after hundreds of hours of training. It is informative to note that there have been no empirical studies that assess the impact of agents on learning and skill training after students experience dozens of hours of interactions with the agents. This is one important direction for future research.

Dyads and Groups of Agents: iSTART, iDRIVE, and ARIES

The aforementioned agents interacted with students one-to-one. Learning environments can also have pairs of agents (dyads) and larger ensembles of agents that exhibit ideal learning strategies and social interactions. It is extraordinarily difficult to train teachers and tutors to apply specific pedagogical techniques, especially when the techniques clash with the pragmatic constraints and habits of everyday conversation. However, pedagogical agents can be designed to have such precise forms of interaction. Researchers at the University of Memphis have designed several systems in which students learn by observing and interacting with ensembles of agents. The highlights of some of these systems are provided in the following.

iSTART (Interactive Strategy Trainer for Active Reading and Thinking). This is an automated strategy trainer that helps students become better readers by constructing self-explanations of the text (McNamara et al., 2004). The construction of self-explanations during reading is known to facilitate deep comprehension (Chi, de Leeuw, Chiu, & LaVancher, 1994; Pressley & Afflerbach, 1995), especially when there is some context-sensitive feedback on the explanations that get produced (Palincsar & Brown, 1984). The 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.

Groups of animated conversational agents scaffold these strategies in three phases of training. In an *Introduction Module,* a trio of animated agents (1 instructor and 2 students) collaboratively describe self-explanation strategies with each other. In a *Demonstration Module,* 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. In a final *Practice* phase, Merlin coaches and provides feedback to the trainee one-to-one while the trainee practices self-explanation 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 iSTART system attempts to interpret the trainee's contributions with latent semantic analysis and other modules with computational linguistics (McNamara, Boonthum, Levinstein, & Millis, 2007; Millis et al., 2004). Merlin gives feedback and 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 Kindergarten through Grade 12 and in college (McNamara, O'Reilly, Best, & Ozuru, 2006). The three-phase iSTART training (approximately 3 hr) has been compared with a control condition that didactically trains students on self-explanation, but without any vicariously 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 comprehension 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 hr of training, we do begin to see some impact on the mastery and application of comprehension strategies. Moreover, the facilitation by iSTART depends on world knowledge and general reading ability. 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 (1978) theory of zone of proximal development, as we discovered in our research with AutoTutor (VanLehn et al., 2007). 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.

iDRIVE (Instruction with Deep-level Reasoning questions In Vicarious Environments). iDRIVE has dyads of animated agents train students to learn science content by modeling deep 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, & the TRG, 2000; Driscoll et al., 2003; King, 1994; Rosenshine, Meister, & Chapman, 1996). However, the asking of deep questions and inquiry does not come naturally (Graesser, McNamara, & VanLehn, 2005), so the process needs to be modeled by agents or humans (Azevedo & Cromley, 2004; Goldman, Duschl, Ellenbogen, Williams, & Tzou, 2003; White & Frederiksen, 1998). 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 0.56 to

1.77 compared to a condition in which students listen to the monologue on the same content without questions.

ARIES (Acquiring Research Investigative and Evaluative Skills). This system is currently being developed in a research collaboration among the University of Memphis, Northern Illinois University, and Claremont Graduate School (Millis, Magliano, Britt, Wiemer-Hastings, Halpern, & Graesser, 2006). ARIES teaches scientific critical thinking via two animated pedagogical agents. The potential of agents taking on different social or pedagogical roles has been investigated by previous researchers (Baylor & Kim, 2005; Wiemer-Hastings & Graesser, 2000). One agent in ARIES, called the guide-agent, is an expert on scientific inquiry and serves as a knowledgeable tutor. The other agent is a fellow student, but could potentially take on other roles (e.g., a neighbor, another scientist, an evaluator of research) that are tailored to the learner. During the training sessions with ARIES, a case study is presented on the screen that describes an experiment that may or may not have a number of flaws with respect to scientific methodology. A three-way conversation transpires among the human student, the expert agent, and the student agent. The human students interact with both agents by holding mixed-initiated "trialogs" in natural language. The agents give the students texts to read, pose diagnostic questions and situated problems, give hints and feedback, encourage question asking, answer questions posed by the student, and monitor the student's progress. Our first test of ARIES has shown greater learning gains on tests of scientific inquiry than a control condition of reading a textbook for approximately the same amount of time.

CONCLUSION AND DISCUSSION

Animated conversational agents are destined to have a major impact on the human-computer interfaces of the future. This article has concentrated on the role of agents in learning environments. However, it is easy to imagine the various ways they can be used in eCommerce, surveys, medical applications, entertainment, and nearly any enterprise that benefits from advanced technologies with conversational facilities. Agents could sell cars on the Web, administer questionnaires on drug use, remind the elderly to take their medicine, and entertain children or adults for hours.

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. For example, one member of the IIS (Institute for Intelligent Systems; Natalie Person) designed a version of AutoTutor that has a rude personality. Instead of giving earnest and short feedback, the RudeTutor gives positive feedback that is sarcastic (e.g., Aren't you the little genius) and negative feedback that is derogatory (e.g., I thought you were bright, but I sure pegged you wrong). This simple substitution of short feedback dialogue moves ended up converting a rather boring earnest AutoTutor (we call PrudeTutor) to an enjoyable, captivating cyber partner. Of course, some students would rather interact with the PrudeTutor than the RudeTutor. The agent 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. Different agent ensembles 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 article, because very little is accomplished in a 1-hr training session. The agents can be embedded in game environments that can captivate many learners for hours and thereby automatize skills, strategies, and knowledge applications.

An imperfect agent may also help learning. This notion has indeed been pursued in the *Teachable Agent* research of Biswas, Leelawong, Schwartz, Vye, and the Teachable Agents Group at Vanderbilt (2005). Human students attempt to help a fellow student agent who has misconceptions and incomplete knowledge. The process of the human student trying to help the cyber student actually ends up facilitating learning in the human. In fact, the Teachable Agent project was motivated by a widely cited finding that peer tutors end up learning more than the tutee. There are many ways the teachable agent can be imperfect, and the imperfection can vary in subtlety and substance. The imperfect Teachable Agent need not even understand or adapt to the human student; it can exhibit its imperfections no matter how hard the human tries to change them.

Our studies of conversational agents have led us to a number of noteworthy or counterintuitive conclusions about discourse and the interactive construction of knowledge. Among these are the following three conclusions:

1. *Students can learn from their conversations with these agents, although the agents do not perfectly understand them at a deep and precise level*: An approximate understanding of the student may be sufficient when combined with dialogue moves that steer the conversation in a way that gets the student to construct answers. Learning was resilient to a number of conversational flaws of the agents that we documented.
2. *Learning is better explained by the content of the conversational moves than the communication media or modalities*: In assessments of learning, differences between printed and spoken dialogue were very subtle compared to what the agent said.
3. *A reasonably smooth tutorial dialogue can be simulated by a small number of dialogue structures and planning modules*: The major components

in AutoTutor were expectation- and misconception-tailored dialogue, the five-step tutoring frame, management of a conversational turn, and adaptive dialogue facilities.

A skeptic may once again raise the objection that the conversational agents will ultimately be a disaster when they do not completely understand the human (Shneiderman & Plaisant, 2005), yet raise their expectations that they do (Norman, 1994). We suspect this may be true when there are high expectations on precision and common ground. However, this is not the case when tutoring on verbal content for students who have little or no subject matter knowledge and when the tutor hedges on how well it understands the student. An AutoTutor might be just the right fit for this niche. We recall when we tested the first version of AutoTutor with speech recognition, the Dragon speech recognition system was correctly translating less than 60% of the content words, and AutoTutor's semantic evaluator was far from perfect. AutoTutor produced reasonable responses to most of what the student said, although its understanding was considerably less than perfect. We suspected that many of the college students who used this version of AutoTutor had the illusion that AutoTutor was comprehending them. But indeed, that just may be the way it is when students try to communicate with human tutors. The fundamental test of the pedagogical value of AutoTutor does not lie in its ability to comprehend perfectly, but rather in its comparisons to humans and in its ability to facilitate learning.

ACKNOWLEDGMENTS

The research on AutoTutor was supported by the National Science Foundation (NSF; SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918), the Institute of Education Sciences (IES; R305H050169), and the (DoD) (Department of Defense) Multidisciplinary University Research Initiative administered by (ONR) (Office of Naval Research) 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 the NSF, IES, DoD, or ONR.

REFERENCES

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

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

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

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

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & the Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*, 363–392.

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.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasely (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Cohen, P. A., Kulik, J. A., & 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., et al. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human computer interaction. *Proceedings of the Institute of Electrical and Electronics Engineers, 91*, 1391–1405.

Corbett, A. T. (2001). Cognitive computer tutors: Solving the two-sigma problem. *User Modeling: Proceedings of the 8th international conference of the 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.

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.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A.C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1–2), 45–80.

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.

Ekman, P. (2003). *Emotions revealed*. New York: Times Books.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10*, 597–602.

Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S., & Tzou, C. T. (2003). Science inquiry in a digital age: Possibilities for making thinking visible. In H. van Oostendorp (Ed.), *Cognition in a digital world* (pp. 253–284). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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., et al. (2008). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research, 19*, 293–312.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. (Eds.). (2003). *Handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Hu, X., Person, P., Jackson, T., & 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. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology, 47*, 19–22.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., et al. (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, & J. Kay (Eds.), *Proceedings of artificial intelligence in education* (pp. 47–54). Amsterdam: IOS Press.

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*, 495–522.

Graesser, A. C., Ventura, M., Jackson, G. T., Mueller, J., Hu, X., & Person, N. (2003). The impact of conversational navigational guides on the learning, use, and perceptions of users of a Web site. In *Proceedings of the AAAI Spring symposium 2003 on Agent-Mediated Knowledge Management* (pp. 9–14). Palo Alto, CA: AAAI Press.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group. (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. (2007). A SEEK Web Tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*(2–3), 89–105.

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 ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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 meeting of the Cognitive Science Society* (pp. 1557–1562). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, W. L., & Beal, C. (2005). Iterative evaluation of a large-scale intelligent game for language learning. In C. Looi, G. McCalla, B. Bredeweg, & 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., & 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.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371.

Kapoor, A., & Picard, R. (2005). Multimodal affect recognition in learning environments. In S. Boll & L. Chaisorn (Eds.), Proceedings of the 13th annual ACM international conferences on multimedia (pp. 677–682). New York: ACM Press.

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.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook on latent semantic analysis.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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.

Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer–human tutoring dialogues. In O. Rambow & S. Balari (Eds.), *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics* (pp. 352–359). East Stroudsburg, PA: Association for Computational Linguistics.

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: Lawrence Erlbaum Associates, Inc.

Mayer, R. E. (2005). *Multimedia learning*. Cambridge, MA: Cambridge University Press.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 227–241). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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.

Millis, K. K., Kim, H. J., Todaro, S., Magliano, J., Wiemer-Hastings, K., & McNamara, D. S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, and Computers, 36*, 213–221.

Millis, K. K., Magliano, J., Britt, A., Wiemer-Hastings, K., Halpern, D., & Graesser, A. C. (2006). *Acquiring Research Investigative and Evaluative Skills (ARIES) for scientific inquiry.* Unpublished manuscript, Northern Illinois University.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*, 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 Association for Computing Machinery, 37*(7), 68–71.

Olde, B. A., Franceschetti, D. R., Karnavat, A., Graesser, A. C., & the Tutoring Research Group. (2002). The right stuff: Do you need to sanitize your corpus when using latent semantic analysis? In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 708–713). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Olney, A., Louwerse, M. M., Mathews, E. C., Marineau, J., Mitchell, H. H., & Graesser, A. C. (2003). Utterance classification in AutoTutor. In J. Burstein & C. Leacock (Eds.), *Building educational applications using natural language processing: Proceedings of the Human Language Technology—North American chapter of the Association for Computational Linguistics Conference 2003 workshop* (pp. 1–8). Philadelphia: Association for Computational Linguistics.

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.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Brain and Behavioral Sciences, 27*, 169–190.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Rus, V., & Graesser, A. C. (2006). Deeper natural language processing for evaluating student answers in intelligent tutoring systems. In Y. Gil & R. Mooney (Eds.), *Proceedings of the 21st national conference on Artificial Intelligence* (pp. 1495–1500). Menlo Park, CA: AAAI Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addresses and overhearers. *Cognitive Psychology, 21*, 211–232.

Schober, M. F., & Conrad, F. (2007). Survey interviews and new communication technologies. In F. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 1–30). New York: Wiley.

Schober, M. F., & Conrad, F. G. (2002). A collaborative view of standardized survey interviews. In D. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. Van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 67–94). New York: Wiley.

Shah, F., Evens, M., Michael, J., & Rovick, A. (2002). Classifying student initiatives and tutor responses in human keyboard-to-keyboard tutoring sessions. *Discourse Processes, 33*, 23–52.

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

Van der Meij, H. (1987). Assumptions of information-seeking questions. *Questioning Exchange, 1*, 111–118.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*, 3–62.

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

Walker, M., Whittaker, S., Stent, A., Maloor, P., Moore, J., Johnson, M., et al. (2003). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science, 28*, 811–840.

Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp. 243–286). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Wiemer-Hastings, P., & Graesser, A. C. (2000). Supporting composition feedback with LSA in Select-a-Kibitzer. *Interactive Learning Environments, 8*, 149–169.

Wiley, J. (2001). Supporting understanding through task and browser design. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1136–1143). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.