EDUCATIONAL PSYCHOLOGIST, *40*(4), 225–234

Scaffolding Deep Comprehension Strategies Through Point&Query, AutoTutor, and iSTART

Arthur C. Graesser and Danielle S. McNamara
Department of Psychology
University of Memphis

Kurt VanLehn
Department of Computer Science
University of Pittsburgh

It is well-documented that most students do not have adequate proficiencies in inquiry and metacognition, particularly at deeper levels of comprehension that require explanatory reasoning. The proficiencies are not routinely provided by teachers and normal tutors so it is worthwhile to turn to computer-based learning environments. This article describes some of our recent computer systems that were designed to facilitate explanation-centered learning through strategies of inquiry and metacognition while students learn science and technology content. *Point&Query* augments hypertext, hypermedia, and other learning environments with question–answer facilities that are under the learner control. *AutoTutor* and *iSTART* use animated conversational agents to scaffold strategies of inquiry, metacognition, and explanation construction. AutoTutor coaches students in generating answers to questions that require explanations (e.g., why, what-if, how) by holding a mixed-initiative dialogue in natural language. iSTART models and coaches students in constructing self-explanations and in applying other metacomprehension strategies while reading text. These systems have shown promising results in tests of learning gains and learning strategies.

Imagine an active, curious, self-regulated learner who asks good questions, persistently hunts for answers, critically evaluates the quality of the fetched answers, constructs deep explanations of the subjective matter, applies the explanations to difficult problems, and consciously reflects on these cognitive activities. That is precisely the sort of learner and learning process that we have been attempting to cultivate in our recent computer-based learning environments (CBLE). It is rare to find a student who spontaneously and skillfully enacts self-regulated learning, inquiry learning, metacognitive strategies, and explanation-centered learning. These processes are also rarely exhibited in normal classrooms and in typical one-on-one sessions with human tutors (Baker, 1996; Graesser, Person, & Magliano, 1995). However, it is possible to build CBLEs that are systematically designed to scaffold different phases of the inquiry process. It is therefore worthwhile to consider CBLEs as a potential practical solution to the poverty of self-regulation, inquiry, metacognition, and explanatory reasoning in today's educational settings.

Researchers have dissected these sophisticated forms of learning into theoretical subcomponents. *Self-regulated learning* occurs when learners create their own goals for learning and then follow up in achieving these goals by monitoring, regulating, and controlling their thoughts and behavior (Azevedo & Cromley, 2004; Winne, 2001; Zimmerman, 2001). *Inquiry learning* is a subclass of self-regulated learning that emphasizes inquiry. The learners ask questions, hunt for answers, evaluate the quality of the answers, and revise their questions in cyclical trajectory until their curiosity is satisfied. Ideally the questions are sincere information-seeking questions that reflect personal curiosity rather than questions handed to them by teachers and peers. Inquiry is sometimes viewed as a miniscientific method that consists of hypothesis generation, experimentation, validation of the hypothesis, and hypothesis modification (Bransford, Brown, & Cocking, 2000; Collins, 1988; White & Frederiksen, 1998). *Explanation-centered* learning occurs when learners attempt to build explanations of the material they comprehend and to

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

apply the explanatory concepts to their reasoning and problem solving (Chi, de Leeuw, Chiu, & LaVancher, 1994; VanLehn, Jones, & Chi, 1992). These explanations may take the form of causal chains or networks in complex systems, logical justifications of claims, and goal-driven plans that motivate human action (Graesser, Baggett, & Williams, 1996). *Metacognition* is defined here as conscious and deliberate thoughts about a person's behavior, emotions, and other thoughts (Flavell, 1979; Hacker, 1998). One research issue addresses the extent to which the self-regulated, inquiry, and explanation-centered learning processes are consciously detected, monitored and regulated, as opposed to the learner implementing relatively unconscious, automatic procedures. *Metacomprehension* is a subclass of metacognition that taps comprehension components, as opposed to memory, reasoning, problem solving, and emotions.

This article describes some CBLEs that were designed to improve explanation-centered learning by planting or refining strategies of inquiry and metacognition. The first section focuses on the process of learners' question asking because inquiry and self-regulated learning will be severely limited if students have trouble asking good questions. A *Point&Query* system was developed to stimulate question asking and to expose learners to deep causal questions. Next we turn to *AutoTutor*, a computer tutor that coaches the learner in building explanations during question answering by holding a conversation in natural language. The final section describes *iSTART* (Interactive Strategy Training for Active Reading and Thinking), which was designed to improve self-explanations and other metacomprehension strategies while reading text. These three systems illustrate how one might scaffold most phases of inquiry and metacognition.

QUESTION ASKING, INQUIRY, AND Point&Query SOFTWARE

There are several barriers in setting up a learning environment that promotes inquiry learning. One salient barrier is that most students are not particularly curious, find it difficult to ask questions, and have trouble generating hypotheses. The average college student asks only 0.17 questions per hour in the classroom, with the rate increasing to 27 questions per hour in one-on-one human tutoring (Graesser & Person, 1994). So it takes 6 hr for a typical college student to ask one question in a classroom. A second barrier is that the quality of the students' questions is disappointing. Students in tutoring sessions on research methods in college asked only about six deep-reasoning questions per hour in a tutoring session that encourages them to ask questions (Graesser & Person, 1994). Deep reasoning questions are about explanations and invite lengthier answers (such as why, why-not, how, and what-if). The quality of student questions may be quite poor in part because the questions of teachers are typically poor in quality. That is, teachers tend to pose shallow rather than deep questions and short-answer rather than long-answer questions (T. J. Dillon, 1988). There are many other barriers in the inquiry process, such as inadequate strategies for finding answers, evaluating answers, monitoring progress, revising questions, and so on (Azevedo & Cromley, 2004). However, the poverty of student questions is a serious barrier that prevents other components of inquiry from developing.

One approach to facilitating student question asking is to model or to explicitly train students how to ask questions. For example, Palincsar and Brown (1984) reported that reading comprehension improves in the reciprocal teaching method, where the tutor models good question-asking skills, invites the learner to demonstrate these skills, and gives feedback on the learner's activities (see also Lysynchuk, Pressley, & Vye, 1990). Reciprocal teaching includes one to three dozen training and practice sessions with the young children. However, King (1992, 1994) found that a 1-hr session that explicitly trains young students to ask good questions improves comprehension and learning from texts and lectures. Rosenshine, Meister, and Chapman (1996) conducted a meta-analysis on 26 empirical studies that compared conditions that instructed or modeled question-asking skills to learning conditions without question-asking training. The median effect size on outcome measures was 0.36 for the standardized tests, 0.87 for experimenter-generated multiple-choice tests, and 0.85 for summaries of texts.

More recently, we have used animated conversational agents (i.e., talking heads) to model the asking of deep-reasoning questions while college students learned about topics in computer literacy (Craig, Gholson, Ventura, Graesser, & TRG, 2000). A series of eight computer topics were discussed by two male animated agents, a computer-controlled virtual tutor and a virtual tutee, located on opposite sides of a monitor. During acquisition, learners either overheard the virtual tutee carry on a dialog (with many questions) with the virtual tutor, or they overheard a more monologue-like discourse by a tutor (no tutee questions). Across the eight topics, the virtual tutee asked a total of 66 (mostly) deep-level reasoning questions and the virtual tutor immediately answered each. A recall task was administered on the computer literacy content covered, followed by a transfer task on question asking. In the memory task, there was a significant 21% increase in the content recalled in the dialog condition than the monologue condition. In the transfer task, the learners were presented with a series of eight new computer literacy topics and were given the opportunity to ask questions on each. Compared with the monologue condition, the students in the dialogue condition asked 39% more questions and recalled 40% more of the content, both effects being significant. This first approach to improving question asking follows the theoretical tradition of Vygotsky (1978): Other agents (human or computer) provide the models, feedback, and scaffolding for learning inquiry skills.

A rather different approach to facilitating inquiry follows the theoretical tradition of Piaget (1952) and subsequently by

Berlyne (1960) and Festinger (1957). In essence, inquiry is spawned when learners experience cognitive disequilibrium. According to a cognitive model of question asking that was recently developed (called PREG; Graesser & Olde, 2003; Otero & Graesser, 2001), learners face cognitive disequilibrium when they encounter obstacles to goals, anomalies, contradictions, incompatibilities with prior knowledge, salient contrasts, obvious gaps in knowledge, and uncertainty in the face of decisions. These different forms of cognitive disequilibrium trigger questions. The PREG model also has a set of rules that predict the particular questions that readers should ask on the basis of the characteristics of the text, the type of disequilibrium, the reader's background knowledge, and metacognitive standards of comprehension (Otero & Graesser, 2001). It is beyond the scope of this article to present the details of this model, but one important claim is that question-asking mechanisms are inextricably bound to the conceptual content of the subject matter. Training students with generic question-asking strategies will be limited without a fine-grained specification of relevant knowledge representations.

We have empirically tested the claim that cognitive disequilibrium influences the quantity and quality of learner questions. In a study by Graesser and McMahen (1993), college students were instructed to ask questions while they read stories or solved algebra word problems. There were different versions of each story or algebra word problem that systematically manipulated the texts with transformations that involved inserting contradictions, inserting irrelevant information, or deleting critical information. As predicted by the PREG model, these transformations significantly increased the number of student questions and their questions were relevant to the transformations.

In a different line of research, college students first read an illustrated text on everyday devices from the popular book by David Macaulay, *The Way Things Work* (Graesser & Olde, 2003). After reading about a device for 5 min (e.g., the cylinder lock), the participant was given a breakdown scenario (e.g., the key turns but the bolt does not move). The participant generated questions for 3 min while reflecting on the causes and possible repairs of the breakdown. The participants were also assessed on a large battery of cognitive abilities and personality measures. Good questions were defined as those that referred to a plausible fault that would explain the breakdown. Graesser and Olde (2003) found that high-quality questions were asked by college students with high mechanical comprehension scores and electronics knowledge. The quality of student questions was one of the two best measures (out of 30 measures) of how well a student comprehended the illustrated texts according to an objective multiple choice test. Deep comprehenders asked better questions, but not necessarily more questions. We conducted a follow-up study on these illustrated texts and breakdown scenarios by considering a nonverbal form of inquiry, namely eye tracking (Graesser, Lu, Olde, Cooper-Pye, & Whitten, in press). The central hypothesis was that deep comprehenders would spend more time gazing on areas depicting faults in the text or diagram; the faults were parts or events that potentially explained the breakdown scenarios. Graesser et al. (in press) reported that participants did indeed have more eye fixations on faults as a function of their aptitude in electronics and mechanical systems. Thus, in addition to verbal questions, inquiry can be manifested by the student's actions (e.g., manipulating components of the device) and by eye movements (such as focusing the eyes on the cause of the breakdown).

We propose that an adequate account of inquiry needs to embrace the traditions of both Piaget and Vygotsky. The learning environment needs to have affordances that put the learner in the state of cognitive disequilibrium, but there also needs to be scaffolding of question-asking skills by social agents (human or computer).

In the early 1990s, software was developed to push the limits of learner question asking and to expose the learner to a broad profile of question categories (Graesser, Langston, & Baggett, 1993). The system was called Point&Query (P&Q). Students learned entirely by asking questions and interpreting answers to questions. The original P&Q software was developed for the subject matter of woodwind instruments and was suitable for high school and college students. The system was a hypertext–hypermedia system, with the augmentation of a question asking-and-answering facility. To ask a question, the learner would point to a hot spot on the display (e.g., the double reed of an oboe) by clicking a mouse. Then a list of questions about the double reed of an oboe would be presented (e.g., What does a double reed look like? What does an oboe sound like? How does a double reed affect sound quality?). The learner would click on the desired question and an answer to the question was presented. Therefore, the learner could ask a question very easily—by two quick clicks of a mouse. On the average, a learner ends up asking 120 questions per hour, which is approximately 700 times the rate of questions in the classroom. The learner also is exposed to good questions because high-quality questions are presented on the menu of question options. It is conceivable that the P&Q software could have a revolutionary impact on learning and thinking skills. Imagine students spending 4 hr a day with the P&Q software, as they learn about topics in science, mathematics, humanities, and the arts. Learning modules on the Web could be transformed to a P&Q environment, as has been already achieved for the subject matter of research ethics (Graesser, Hu, Person, Jackson, & Toth, 2004). Students would learn and automatize excellent question-asking skills, perhaps to the extent of rekindling the curiosity of a 4-year-old child.

Evaluations of the P&Q software on question asking revealed, however, that it is not sufficient to simply expose the students to good menus of questions associated with hot spots in a large landscape of hypertext–hypermedia content. When participants are left to surf the hyperspace on their own, they tend to drift toward posing shallow questions. That is, the percentage of the learner's P&Q choices that were shallow questions was higher than chance among the questions available in the hyperspace (Graesser et al., 1993). Deeper questions explain the causal mechanisms. Such questions include *why* questions (Why did event E occur?), *how* questions (How does process P occur?), *what-if* questions (What are the consequences of event E occurring?), and *what-if-not* questions (What if state S did not exist?). The learner needs to have a goal, task, or challenge that places them in cognitive disequilibrium and requires causal explanations before they tread in deeper waters. This was apparent in a study where Graesser et al. (1993) randomly assigned the college students to one of three conditions with different instructional goals: deep goals, shallow goals, versus no-bias control. In the *deep goal* condition, the learners were instructed that, after they studied the woodwind hyperspace, they would be expected to design a new woodwind instrument that had a deep pure tone. The design of a new instrument requires deep knowledge, such as causal reasoning about the impact of size of instruments on pitch and volume, and the impact of mouthpieces (i.e., air reed, single read, versus double reed) on the purity of tones. In the *shallow goal* condition, their goal was to design an ensemble with five instruments for a New Year's Eve party with 40-year-old yuppies. The aesthetic appeal of the instruments was more important than causal reasoning in this condition, so the learners could rely on shallow knowledge and questions. In the *no-bias* goal condition, the students were given vague instructions that they would be administered a test after exploring the woodwind hyperspace. The results of the study were clear-cut regarding the learners' selection of questions in the woodwind hyperspace. Their selection of shallow questions was above chance and nearly equivalent in the shallow goal condition and no-bias control; the selection of deep causal questions was below chance. In contrast, the selection of causal questions was above chance only in the deep goal condition, where causal reasoning was required.

In summary, it apparently is not sufficient to simply expose learners to a learning environment with a balanced distribution of questions (both shallow and deep) for deep explanation-centered learning to be achieved. We are convinced that there needs to be a goal or task that puts the learner in cognitive disequilibrium, such as challenging them with a difficult problem to solve or with a claim that clashes with a valued belief system. We are also convinced that learning will not be achieved without some scaffolding of additional phases in self-regulated inquiry (Azevedo & Cromley, 2004; A. Dillon & Gabbard, 1998; Goldman, 1996; Hadwin & Winne, 2001), such as setting subgoals, evaluating the quality of accessed information, and judging the relevance of accessed information to the primary goals. One direction for future research is to orthogonally vary cognitive disequilibrium and scaffolding to test the predicted interaction on learning gains. The learning environments with the Point&Query facility could have different versions that systematically manipulate whether or not the animated conversational agents (a) present suitable challenges to the learner, (b) model illuminating inquiry processes, and (c) give feedback to learners who drift too far away from the critical paths in the hyperspace.

COACHING ANSWERS TO DEEP QUESTIONS THROUGH AutoTutor

If inquiry is truly self-regulated, then after the learner generates a question the learner needs to go through the self-initiated process of finding or constructing an answer to the question. However, composing an answer can be quite challenging when the ideal answer is lengthy or requires deep explanatory reasoning. For example, a typical student produces only one or two sentences when asked a difficult conceptual physics question, such as the one below.

> 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?

An ideal answer is a paragraph of information in length (roughly 10 sentences), but the initial answer to such a question by a college student is typically only one or two sentences. This is where tutorial dialogue is particularly helpful. A good tutor engages the learner in a dialogue that assists the learner in the evolution of an improved answer that draws out more of the learner's knowledge, that fills in missing information, and that corrects the learner's misconceptions. The dialogue between tutor and student may be 100 *turns* (i.e., the learner expresses something, then the tutor, then the learner, and so on) before a good answer to this single physics question emerges.

AutoTutor is a computer tutor that attempts to simulate the dialogue moves of a human tutor (Graesser et al., 2004; Graesser, Person, & Harter, 2001; Graesser, VanLehn, Rose, Jordan, & Harter, 2001). AutoTutor holds a conversation in natural language that coaches the student in constructing a good explanation in an answer, that corrects misconceptions, and that answers student questions. AutoTutor delivers its dialogue moves with an animated conversational agent that has a text-to-speech engine, facial expressions, gestures, and pointing. Animated agents have been become increasingly popular in learning environments on the Web, Internet, and desktop applications (Atkinson, 2002; Johnson, Rickel, & Lester, 2000). The learner contributions are currently typed into AutoTutor, although we do have a prototype version with speech recognition.

The primary method of scaffolding explanations through dialogue is what we call *expectation and misconception tailored dialogue* (EMT dialogue). Both AutoTutor (and human tutors) typically have a list of anticipated good answers (called *expectations*) and a list of *misconceptions* associated with each main question. For example, there are approximately 10 sentence-like expectations (e.g., *force equals mass times acceleration*) that AutoTutor would like to cover in a good answer to the example physics problem involving a collision and neck injuries. One goal of the tutor is to coach the student in covering the list of 10 expectations. This is accomplished by AutoTutor generating pumps (*what else?*), hints, prompts for specific information, assertions, and other dialogue moves until the student or tutor covers each expectation on the list. As the learner expresses information over many turns, the list of expectations is eventually covered and the main question is scored as answered. A second goal is to correct misconceptions that are manifested in the student's talk by simply correcting the errors as soon as they are manifested. Most human tutors quickly correct student errors so that students do not flounder down unproductive avenues. A third goal is to adaptively respond to the student by giving short feedback on the quality of student contributions (positive, negative, or neutral) and by answering the student's questions. A fourth goal is to manage the dialogue in a fashion that appears coherent and accommodates unusual speech acts by learners.

One way to convey what AutoTutor can do is through an example dialogue. The Appendix presents an excerpt of a conversation with a college student on the example conceptual physics question. There is an annotated analysis of the example dialogue that specifies the categorized dialogue moves of AutoTutor, the classified speech acts of the student, and assorted comments to help the reader interpret what is going on. The content expressed by either AutoTutor or the student in the Appendix is signified in italics. Discourse categories of AutoTutor dialogue moves have been added in capitals, whereas other information is added in normal font.

The example in the Appendix illustrates some of the important characteristics of a dialogue with AutoTutor. AutoTutor needs to adapt to what the student says so it needs to appropriately classify the content of the student turns. When the student asks a question, AutoTutor needs to answer it. When the student makes one or more assertions in a turn, then AutoTutor needs to give feedback on the quality of the information and build productively on what the student says. AutoTutor therefore analyzes each student turn by first segmenting the student turns into speech act units and then assigning these units to categories, such as assertion, short answer, metacognition, metacommunication, verification question, and comparison question. There are approximately 20 categories of student speech acts; 16 of these are different categories of student questions. AutoTutor attempts to accommodate virtually any student question, assertion, comment, or extraneous speech act. This requires interpreting the student speech acts by implementing modules developed in the field of computational linguistics (Jurafsky & Martin, 2000). After AutoTutor interprets what the student says, it needs to formulate what to say next in AutoTutor subsequent turn. Each turn of AutoTutor requires the generation of one or more dialogue moves that adaptively respond to what the student just expressed and that advance the conversation in a constructive fashion. The dialogue moves within a turn are connected by dialogue markers, as illustrated in the Appendix. Some dialogue moves are very responsive to the student-preceding turn, such as the short feedback (positive, neutral, versus negative), the answers to student questions, and corrections of student misconceptions. Other dialogue moves push the dialogue forward in an attempt to cover the expectations in a good answer to the main question. These forward-directed dialogue moves include pumps (e.g., *Tell me more, What else?*), hints, prompts for specific words or phrases, and assertions.

AutoTutor has been evaluated on learning gains in several experiments on the topics of computer literacy (Graesser et al., 2004) and conceptual physics (VanLehn et al., 2004). The results of these studies have been quite positive. It is well-established that one-to-one human tutoring is a powerful method of promoting learning (Cohen, Kulik, & Kulik, 1982), even though most human tutors have moderate subject matter knowledge and no training in pedagogy or tutoring. These unaccomplished human tutors enhanced learning with an effect size of 0.4 standard deviation units (called sigma), whereas intelligent tutoring systems with sophisticated pedagogical tactics, but no natural language dialogue, produce effect sizes of approximately 1.0 sigma (Corbett, 2001). Previous versions of AutoTutor have produced gains of 0.2 to 1.5 sigma (a mean of 0.8), depending on the learning performance measure, the comparison condition (either pretest scores or a control condition in which the learner reads the textbook for an equivalent amount of time as the tutoring session), the subject matter, and the version of AutoTutor (Graesser, Lu, et al., 2004). Approximately a 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. These evaluations place AutoTutor somewhere between an unaccomplished human tutor and an intelligent tutoring system.

One study compared AutoTutor with a condition in which college students in a physics course read the course textbook on the same topics for a study time comparable to AutoTutor (approximately 2 hr) and with a control condition in which no physics material was assigned (Graesser et al., 2003). Two separate versions of a multiple choice test were administered prior to and subsequent to the training. The adjusted posttest scores (that partialed out pretest scores) were .727, .610, and .608 in the AutoTutor, read-textbook, and no-read-control, respectively. AutoTutor produced significantly better learning than the two comparison conditions, as predicted.

An equally remarkable result was that reading the textbook was equivalent to studying nothing. The absence of learning from the textbook reading can perhaps be explained by the lack of active engagement in the learning process. Students in the textbook condition may not have been engaged in the learning process, whereas AutoTutor forced them to be actively engaged through interactive dialogue. However, there are two other potential reasons for the lack of learning from textbook reading, both of which address metacomprehension. First, the readers' metacomprehension standards may have been set at a shallow level while they were reading the textbook; consequently, they did not attempt to acquire causal explanations of the physics. Many readers are prone to settle for shallow comprehension of material unless they are challenged with problems/questions that place them in cognitive disequilibrium and that encourage deeper comprehension (Hacker, 1998; Otero & Graesser, 2001; Otero & Kintsch, 1992). Second, the readers may have had difficulties calibrating their comprehension while reading, which is typically the case for readers of all ages (Glenberg & Epstein, 1985; Maki, 1998). According to the meta-analysis reported by Maki (1998), there is only a .27 correlation between objective tests on reading material and readers' judgments on how well they comprehended the material. Thus, the students in the textbook reading condition of Graesser et al. (2003) may not have been trying to comprehend very deeply or may have been incapable by virtue of poor comprehension calibration skills. It takes a challenging experience like learning from AutoTutor to construct explanations of the material at deeper levels. Quite clearly, AutoTutor is not the only class of scaffolds that encourages explanation-centered learning and deeper standards of comprehension. We are currently investigating alternative materials, tasks, and dialogue patterns that produce equivalent or better learning gains than AutoTutor (VanLehn et al., 2005).

AutoTutor was designed to foster explanation-centered learning, with apparent success, but there is a relevant auxiliary question of how well AutoTutor promotes improved question-asking skills in the learner. To answer this question, we analyzed the questions asked by college students in a physics course who interacted with AutoTutor compared to those who interacted with one of five expert human tutors through computer-mediated communication. Each question in these two corpora was classified into one of the 16 question categories used by AutoTutor and previous tutoring corpora (Graesser & Person, 1994). The number of learner questions per 100 turns was twice as high for learners who received AutoTutor compared to those who received tutoring from the expert human tutor, 13.2 versus 6.4 questions per 100 student turns, respectively. The proportion of questions that were classified as deep was significantly higher for AutoTutor than the human tutors, 24.6% versus 9.0%, respectively. Therefore, the tutorial dialogue of AutoTutor stimulated more learner questions and also deeper questions.

METACOMPREHENSION, SELF-EXPLANATIONS, AND iSTART

Whereas AutoTutor supports explanation-centered learning through tutorial dialogue, iSTART helps young adolescent to college-aged students learn metacomprehension strategies that support deeper comprehension while they read. iSTART is a Web-based reading strategy trainer that uses animated agents to discuss, model, and provide feedback about reading strategies that improve comprehension of difficult science texts (McNamara, Levenstein, & Boonthum, 2004). iSTART was designed based on a successful classroom intervention called Self-explanation Reading Training (SERT; McNamara, 2004; O'Reilly, Best, & McNamara, 2004). SERT was motivated by two sets of empirical findings. First, students who self-explain text are more successful at solving problems, more likely to generate inferences, construct more coherent mental models, and develop a deeper understanding of the concepts covered in the text (Chi et al., 1994; Chi & VanLehn, 1991; Magliano, Trabasso, & Graesser, 1999; VanLehn et al., 1992). However, these advantages only emerge for students who are skilled self-explainers. Therefore, SERT combines self-explanation with another line of research showing the importance of reading strategies to promote successful comprehension (Bereiter & Bird; 1985; Palincsar & Brown, 1984). The combination of self-explanation with reading strategy training (i.e., metacognition and metacomprehension) helps students understand difficult, unfamiliar concepts, such as those typically encountered in science textbook material.

The SERT and iSTART interventions teach readers to self-explain using five reading strategies: comprehension monitoring, paraphrasing, making bridging inferences, predictions, and elaborations. *Comprehension monitoring* enables the reader to recognize a failure of understanding and it is this recognition that triggers the use of additional active reading strategies. *Paraphrasing* helps students remember the surface structure of the text by transforming it into more familiar ideas. However, students are encouraged to go beyond this basic sentence-focused processing by invoking knowledge-building strategies that link the content of the sentences to other information, either from the text or from the students prior knowledge. The process of making *bridging inferences* improves comprehension by linking the current sentence to the material previously covered in the text. Such inferences allow the reader to form a more cohesive global representation of the text content. Students may also use *prediction* to anticipate the content in subsequent text, either by guessing what is coming next or by reminding themselves to watch out for some particular item that will aid comprehension. Finally, readers may associate the current sentence with their own related prior knowledge using a strategy called *elaboration*. Readers are encouraged to draw on common sense, mundane reasoning, and domain-general knowledge when they do not have sufficient knowledge about the topic of the text. Elaboration essentially ensures that the information in the text is linked to information that the reader already knows. These connections to prior knowledge result in a more coherent and stable representation of the text content (Kintsch, 1998).

The animated agents of iSTART provide three phases of training. First, the trainee is provided with instruction on self-explanation and reading strategies in an *Introduction Module*. There is a trio of animated agents (an instructor and two students) that cooperate with each other, provide information, pose questions, and provide explanations of the reading strategies. The interactions between the characters vicariously simulate the active processing necessary to learn the strategies. The instructor character presents definitions and examples for each strategy and questions the student characters knowledge of the strategies. The student characters banter among themselves as they ask the instructor for examples or clarifications. After the presentation of each strategy, the trainees complete brief multiple-choice quizzes to assess their learning.

In the second phase, called the *Demonstration Module*, two Microsoft Agent characters (Merlin and Genie) demonstrate the use of self-explanation and the trainee identifies the strategies being used. A science text is presented on the computer screen one sentence at a time. Genie (representing a student) reads the sentence aloud and produces a self-explanation, which appears in a self-explanation box. Merlin (the teacher) continues by asking the trainee to indicate which strategies Genie employed in producing the self-explanation. The trainee answers by clicking on a strategy in a strategy box with five options (monitoring, paraphrasing, bridging, predicting, and elaborating). Merlin follows up by asking the student to identify and locate the various reading strategies contained in Genie self-explanation. For example, if the student answered that Genie had provided an elaboration in his self-explanation, then Merlin might ask the student to click on the part of the self-explanation that contained the elaboration. Merlin gives Genie verbal feedback on the quality of his self-explanation, which mimics the interchanges that the student will encounter in the practice module. For example, sometimes Merlin states that the self-explanation is too short, prompting Genie to add to his self-explanation.

In the third phase, called *Practice*, Merlin coaches and provides feedback to the trainee while the trainee practices self-explanation using the repertoire of reading strategies. The goal is to help the trainee acquire the skills necessary to integrate prior text and prior knowledge with the sentence content. For each sentence, Merlin reads the sentence and asks the trainee to self-explain it by typing a self-explanation. Merlin gives feedback, sometimes asking the trainee to modify unsatisfactory self-explanations. Once the self-explanation is satisfactory, Merlin asks the trainee to identify what strategy was used and where in the sentence they used it, after which Merlin provides general feedback. The agent interactions with the trainee are moderated by the quality of the explanation. For example, more positive feedback is given for longer, more relevant explanations, whereas increased interactions and support are provided for shorter, less relevant explanations (McNamara, Boonthum, Levinstein, & Millis, in press; McNamara et al., 2004; Millis et al., 2004).

Recent studies have evaluated the impact of iSTART on both reading strategies and comprehension (McNamara et al., 2004; O'Reilly, Sinclair, & McNamara, 2004). The three-phase iSTART training has been compared with a control condition that didactically trains students on self-explanation, but without any vicariously modeling and feedback via the agents. After training in the iSTART and control conditions, the participants are asked to self-explain a transfer text (e.g., on heart disease) and are subsequently given a comprehension test. The results have revealed that metacognitive strategies and comprehension are facilitated by iSTART. Moreover, the impact of the strategy training the resulting cognitive representations depends on the student's general reading ability. For example, O'Reilly et al. (2004) demonstrated that iSTART helps both high and low strategy-knowledge students, but in different ways. They found that 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' theory of zone of proximal development (Vygotsky, 1978). That is, 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 research on reading strategies support the need for user-adaptive metacognitive training in intelligent tutoring systems. iSTART was developed so that reading strategy training could be brought to scale, and thus provided to a wide range of students. To that end, current research and development on iSTART is centered on its use in classroom settings. The current goals of the iSTART project are to expand the types and levels of reading strategy training and to create a teacher interface to support its use in the classroom.

CLOSING COMMENTS

At the beginning of this article we sketched a view of an ideal student: An active, curious, self-regulated learner who asks good questions, persistently hunts for answers, critically evaluates the quality of the fetched answers, constructs deep explanations of the subjective matter, applies the explanatory content to difficult problems, and consciously reflects on these cognitive activities. Available research strongly supports the conclusions that very few students are ideal learners, that these cognitive proclivities are rarely cultivated in classrooms, and that they fail to emerge by simply exposing students to rich learning environments. Much more is needed to equip an ideal learner. Learners need to be challenged with questions, problems, tasks, and goals that place them in cognitive disequilibrium and thereby encourage standards of comprehension that are explanation centered. Learners need scaffolding for deep inquiry (i.e., good questions and answers) and metacognitive strategies. These mechanisms are sufficiently complex that they will most likely be furnished in computer-based learning environments; the alternative is to provide extensive teacher–tutor training on sophisticated and subtle pedagogical and cognitive mechanisms. Point&Query, AutoTutor, iSTART, and other recent learning environments are a good start, but the landscape of more effective systems is wide open. We are convinced that the social and pedagogical fabric of classrooms, groups, and human tutoring sessions is inherently limited compared to what computers can realistically provide.

ACKNOWLEDGMENTS

The research on AutoTutor was supported by the National Science Foundation (NSF; SBR 9720314, REC 0106965, REC 0126265, ITR 0325428) and the DoD Multidisciplinary University Research Initiative administered by the Office of Naval Research (ONR) under Grant N00014–00–1–0600. This research on iSTART was supported by NSF Interagency Education Research Initiative Grant REC–0089271. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Department of Defense, ONR, or NSF.

We thank the Tutoring Research Group at the University of Memphis (visit http://www.autotutor.org), the Center for Interdisciplinary Research on Constructive Learning Environments Research Group at the University of Pittsburgh (http://www.pitt.edu/~circle), and the Cognitive Science and Educational Practice group at the University of Memphis (http://csep.psyc.memphis.edu).

REFERENCES

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.

Baker, L. (1996). Social influences on metacognitive development in reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 331–352). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2,* 131–156.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity.* New York: McGraw-Hill.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: 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., & VanLehn, K. (1991). The content of physics self-explanations. *Journal of the Learning Science, 1,* 69–105.

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.

Collins, A. (1988). Different goals of inquiry teaching. *Questioning Exchange, 2,* 39–45.

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 TRG. (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.

Dillon, A., & Gabbard, R. (1998). Hypermedia as educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research, 68,* 322–349.

Dillon, T. J. (1988). *Questioning and teaching: A manual of practice.* New York: Teachers College Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychology, 34,* 906–911.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 702–718.

Goldman, S. R. (1996). Reading, writing, and learning in hypermedia environments. In H. Van Oostendorp & S. de Mul (Eds.), *Cognitive aspects of electronic text processing* (pp. 7–42). Norwood, NJ: Ablex.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10,* S17–S32.

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, October-November,* 29–39.

Graesser, A. C., Jackson, G. T., Mathews, E. C., Mitchell, H. H., Olney, A., Ventura, M., et al. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialogue. In R. Alterman, & D. Hirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 1–6). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Langston, M. C., & Baggett, W. B. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation: Vol. 29. Categorization by humans and machines* (pp. 411–436). Orlando, FL: Academic.

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,* 180–192.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (in press). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology, 85,* 136–151.

Graesser, A. C., & Olde, B. (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., Harter, D., & TRG. (2001). Teaching tactics and dialogue in AutoTutor. *International Journal of Artificial Intelligence in Education, 12,* 257–279.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology, 9,* 359–387.

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

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hadwin, A., & Winne, P. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation, 7,* 313–334.

Johnson, W. L., & Rickel, J. W., & Lester, J. C. (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. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303–323.

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.

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

Lysynchuk, L. M., Pressley, M, & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *Elementary School Journal, 90,* 469–484.

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.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91,* 615–629.

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

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (in press). Using LSA and word-based measures to assess self-explanations in iSTART. In T. Landauer, D. S., McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A road to meaning.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S., Levenstein, I. B., & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. *Behavioral Research Methods, Instruments, and Computers, 36,* 222–233.

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, & Computers, 36,* 213–221.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Reilly, T. P., Sinclair, G. P., & McNamara, D. S. (2004). iSTART: A web-based reading strategy intervention that improves students' science comprehension. In Kinshuk, D. G. Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age: CELDA 2004* (pp. 173–180). Lisbon, Portugal: IADIS.

Otero, J., & Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction, 19,* 143–175.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what the read. *Psychological Science, 3,* 229–235.

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

Piaget, J. (1952). *The origins of intelligence*. Madison, CT; International Universities Press.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27,* 91–109.

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.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2005). When is reading just as effective as one-on-one interactive human tutoring? In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences, 2,* 1–60.

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

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16,* 3–117.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

APPENDIX
Example Excerpt in an AutoTutor–Student Dialogue With Classified Dialogue Moves of AutoTutor, Classified Speech Acts of the Student, and Comments

AutoTutor-1: *When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries? Explain why.* (MAIN QUESTION)

Student-2: *I don't know.* (This expression by the student is classified as a Metacognitive speech act because it conveys what the student knows about his/her own knowledge.)

AutoTutor-3: *Could you at least try to give me an answer? That way, we'll have something as a starting place to work from.* (PUMP. This dialogue move encourages the student to say more, but does not direct the student to express particular information.)

Let's try this together. (DISCOURSE MARKER. This is a discourse marker that connects different dialogue moves within a turn.)

When the collision provides a large force to accelerate the head, what could happen to the neck? (HINT. This dialogue move leads the student to articulate a particular sentence, clauses, or proposition.)

Student-4: *What is a force?* (The student can ask questions during an AutoTutor session.)

AutoTutor-5: *I'll take a crack at answering your question.* (DISCOURSE MARKER) *Force is any influence that tends to accelerate an object; a push or pull; measured in newtons. Force is a vector quantity.* (ANSWER. AutoTutor answers the student direct question.)

Was that what you were looking for? (QUESTION SOLICITATION. AutoTutor allows the student the opportunity to rephrase his question.)

[At this point, the dialogue continues until AutoTutor believes it has detected a misconception, as shown in the following excerpt.]

AutoTutor-20: *See if you know the answer to this.* (DISCOURSE MARKER. AutoTutor prepares the student for a direct question.)

When the head and body are moving at the same acceleration, they are moving at the same ________? (PROMPT. This dialogue move encourages the student to fill in a particular word or phrase of information, in this case the word "rate." Prompts are needed for students who are low verbal and to elicit a particular noun, verb, or adjective.)

Student-21: *distance*

AutoTutor-22: *Moving at the same rate.* (PROMPT COMPLETION. AutoTutor splices in the correct information if the student provides an incorrect answer to its prompt.)

[At this point, the dialogue continues until all relevant principles are covered in the conversation, followed by a summary by the tutor.]