

Running head: TUTORING

Instruction Based on Tutoring

Arthur C. Graesser, Andrew Olney, and Whitney Cade

University of Memphis

Graesser, A. C., Olney, A., Cade, W. (2009). Instruction based on tutoring. In R.E. Mayer and P.A. Alexander (Eds.), *Handbook of Research on Learning and Instruction*. New York Routledge Press.

Send correspondence to:

Art Graesser
Department of Psychology & Institute for Intelligent Systems
202 Psychology Building
University of Memphis
Memphis, TN 38152-3230
901-678-4857
901-678-2579 (fax)
a-graesser@memphis.edu

Instruction Based on Tutoring

Tutoring is the typical solution that students, parents, teachers, principals and school systems turn to when the students are not achieving expected grades and educational standards. There are serious worries in the community when a school is not meeting the standards of a high stakes test, and teachers are anxious about the prospects of losing their jobs due to the criteria and policies of *No Child Left Behind*. Schools and families worry when a student runs the risk of losing a scholarship or when an athlete may be cut from a team. Tutors step in to help under these conditions. Wealthier families might end up paying \$200 per hour for an accomplished tutor to rescue save a son or daughter. However, these expectations may be rather high, considering that most tutors are same-age peers of the students, slightly older cross-age tutors, citizens in the community, and paraprofessionals who have had little or no training on tutoring pedagogy. Nevertheless, their tutoring can be effective in helping students learn, as we will document in this chapter.

Although most tutors in school systems have little or no tutoring training, there are many examples of excellent tutoring programs that are grounded in the science of learning. One notable example is the *Reciprocal Teaching* method that helps students learn how to read text at deeper levels (Palincsar & Brown, 1984, 1988). The tutoring method engages the tutor and students in a dialogue that jointly constructs the meaning of the text. The dialogue is supported with the use of four strategies: generating questions, summarizing text segments, clarifying unfamiliar words and underlying global ideas, and predicting what will happen next in the text. These strategies are applied in a context-sensitive manner rather than mechanically applied in scripted lessons. Moreover, the tutors systematically change their

style of tutoring as the lessons proceed. When students are initially introduced to Reciprocal Teaching, the tutor models the application of these strategies by actively bringing meaning to the written word (called content strategies) and also monitoring one's own thinking and learning from text (called meta-cognitive strategies). Over the course of time, the students assume increased responsibility for leading the dialogues. That is, after the modeling phase, the tutor has the students try out the strategies while the tutor gives feedback and scaffolds strategy improvements. Eventually the students take more and more control as the tutor fades from the process and occasionally intervenes much like a coach. This *modeling-scaffolding-fading* instructional process has a long history in the psychology of learning (Collins, Brown, & Newman, 1989; Rogoff & Gardner, 1984; Vygotsky, 1978).

The Reciprocal Teaching method has been tested in dozens of studies and has been shown to improve students' reading skills. Rosenshine and Meister (1994) conducted a meta-analysis of 16 studies of Reciprocal Teaching that were conducted with students from age seven to adulthood. The method was compared with traditional basal reading instruction, explicit instruction in reading comprehension, and reading and answering questions. When experimenter-developed comprehension tests were used, the median effect size was 0.88 sigma¹. When standardized measures were used to assess comprehension, the median effect size favoring Reciprocal Teaching was 0.32. The Reciprocal Teaching method has also been applied in classroom contexts with trained teachers applying the method in front of a classroom of students or in small groups. The method was recently

¹ A sigma is a measure in standard deviation units that compares a mean in the experimental treatment to the mean in a comparison condition.

accepted as a method to try in *What Works Clearinghouse*, a mechanism funded by the US Department of Education to test promising methods of instruction in a large number of schools throughout the country.

Despite encouraging examples like Reciprocal Teaching, there are practical challenges in relying on humans to supply high-quality one-on-one human tutoring. It is very costly to train tutors on tutoring strategies. There is a high burn-out rate when both skilled and unskilled tutors face the realities of how difficult it is to help students learn. Indeed, the burnout rate also is extremely high for teachers according to the many surveys that have been conducted on teacher retention. Fortunately, the tutoring enterprise has expanded beyond human tutoring and into the realm of computer tutoring. Computers are available 24/7, do not get fatigued, do not burn out, and can reliably apply pedagogical strategies. There are now intelligent tutoring systems (ITS) and other advanced learning environments that implement sophisticated instructional procedures. ITS are able to induce the characteristics of individual learners at a fine-grained level, to assign problems or tasks that are sensitive to the students' profile, and to generate specific tutoring actions that attempt to optimize learning according to scientific principles. Unlike human tutors, ITS provide precise control over the instructional activities, which of course is a methodological virtue. ITS have the capacity to scale up in delivering learning assistance to many more students than can be provided by human tutors.

The Cognitive Tutors developed by the Pittsburgh Science of Learning Center is one noteworthy ITS family (Anderson, Corbett, Koedinger, & Pelletier, 1995; Koedinger, Anderson, Hadley, & Mark, 1997; Ritter, Koedinger, Anderson, & Corbett, 2007). The Cognitive Tutors help students learn algebra, geometry, and programming languages by

applying learning principles inspired by the ACT-R cognitive model (Anderson, 1990).

There is a textbook and curriculum to provide the content and the context of learning these mathematically intensive subject matters, but the salient contribution of the Cognitive Tutors is to help students solve problems. The tutor scaffolds the students to take steps in solving the problem by prompting them to actively take the steps, by comparing the students' actions to ideal models of correct solutions, by giving students feedback on their actions, and by providing hints and other forms of help. The Cognitive Tutor mechanism incorporates a combination of declarative knowledge (facts) and procedural knowledge. Students are expected to learn through enough practice in varied contexts on problems that are tied to the curriculum.

The Cognitive Tutors are now used in over 2000 school systems throughout the country and are among the methods being assessed in the *What Works Clearinghouse*. These tutors have been heavily evaluated over the course of 35 years. The effect sizes on experimenter-developed tests are approximately 1.0 sigma compared to normal classroom teaching (Corbett, 2001). According to Ritter et al. (2007), standardized tests show overall effect sizes of 0.3 sigma, but particularly shine for the subcomponents of problem solving and multiple representations, which show effect sizes of 0.7 to 1.2. The What Works Clearinghouse investigations show an effect size of 0.4 sigma. The Cognitive Tutors are an excellent example of how scientific principles of learning can be implemented in a technology that not only helps learning but also scales up to widespread use in thousands of school systems.

This chapter reviews the research on both human and computer tutors. The first section covers human tutoring, including pedagogical theories and empirical evidence for its

effectiveness. The second puts the lens on computer tutors that directly manipulate tutoring strategies and the consequences on learning. The final section identifies some future directions for the field to pursue.

Human Tutoring and Its Impact on Learning

It could be argued that tutoring was the very first form of instruction. Children were trained one-on-one by parents, other relatives, and members of the village who had particular specialized skills. The apprenticeship model reigned for several millennia before we encountered the industrial revolution and classroom education (Collins & Halverson, 2009). Throughout that part of history, the modelling-scaffolding-fading process was probably the most sophisticated form of early tutoring. The alternative of lecturing was probably more prevalent: The master simply lectured to the apprentice, the apprentice nodded (knowingly or unknowingly), and the master undoubtedly grew frustrated when very few of the ideas and skills were sinking in. Lecturing is ubiquitous in the repertoire of today's unskilled tutors (Graesser, Person, & Magliano, 1995), but there are some other strategies that come naturally, as will be elaborated in this chapter.

How much does human tutoring help learning?

Evaluations of one-on-one tutoring have shown that the method is quite effective, even when the tutors are unskilled tutors. Unskilled tutors are defined in this chapter as tutors who are not experts on subject matter knowledge, are not trained systematically on tutoring skills, and are virtually never evaluated on their impact on student learning. Unskilled tutors are paraprofessionals, parents, community citizens, cross-age tutors, or same-age peers. Meta-analyses show learning gains from typical human tutors, the majority being unskilled, of

approximately 0.4 sigma when compared to classroom controls and other suitable controls (Cohen, Kulik & Kulik, 1982). There are many possible explanations of these learning gains from tutors who are unskilled. Perhaps the tutor can detect whether or not the student is generally mastering the subject matter on the basis of the student's verbal responses, from their nonverbal reactions, or from the student's attempts to perform a task. The tutor would then re-plan and make adjustments to help the student move forward. Perhaps the one-on-one attention motivates the student or encourages sufficient mastery to prevent embarrassing performance deficits. Perhaps the nature of conversation encourages a "meeting of the minds" with sufficient common ground for learning to be built on a solid discourse foundation. The question is still unsettled why one-on-one tutoring is so effective when the tutor is unskilled.

Available evidence suggests that the expertise of the tutor does matter, but the evidence is not strong. Collaborative peer tutoring shows an effect size advantage of 0.2 to 0.9 sigma (Johnson & Johnson, 1992; Mathes & Fuchs, 1994; Slavin, 1990; Topping, 1996), which appears to be slightly lower than older unskilled human tutors. Peer tutoring is a low-cost effective solution because expert tutors are expensive and hard to find. Unfortunately, there have not been many systematic studies on learning gains from expert tutors because they are expensive, they are difficult to recruit in research projects, and tutors tend to stay in the tutoring profession for a short amount of time (Person, Lehman, & Ozbun, 2007). However, available studies report effect sizes of 0.8 to 2.0 (Bloom, 1984; Chi, Roy, & Hausmann, 2008; VanLehn et al., 2007), which is presumably higher than other forms of tutoring. The question is still unsettled on the impact of tutoring expertise on learning gains.

The impact of tutoring expertise on student learning is complicated by the fact that much of the answer lies in what the student does, not what the tutor does. Constructionist theories of learning have routinely emphasized the importance of getting the student to construct the knowledge, as opposed to an instruction delivery system that transfers the knowledge to the student (Bransford, Brown & Cocking, 2000; Mayer, 2009). Students learn by expressing, doing, explaining, and being responsible for their knowledge construction, as opposed to being passive recipients of exposure to information. There is considerable evidence for the constructivist thesis in general, (Bransford, Brown, & Cocking, 2000), but this chapter will consider the evidence for constructivism in tutoring per se. One form of evidence is that the tutors in these same-age and cross-age collaborations tend to learn more than the tutees (Cohen et al., 1982; Mathes & Fuchs, 1994; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003). Playing the role of tutor rather than tutee undoubtedly increases study, effort, initiative, and organization, all of which contribute to learning. In peer tutoring, students often are randomly assigned to tutor versus peer, so any advantages of the tutor role can not be explained by prior abilities. Another form of evidence lies in who contributes most to the tutoring session. Is it the student or tutor? Correlational evidence reveals that students learn more when they contribute a higher percentage of the words and ideas to the tutoring sessions (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Litman et al., 2006). A good tutor apparently says very little when the student is on a roll and learning. Yet another form of evidence is that it does not help much for the tutor to articulate explanations, solutions, and other critical content in the form of information delivery, without making any attempt to connect with what the learner knows (Chi et al., 2001; VanLehn, Siler, Murray, Yamauchi, &

Baggett, 2003). Explanations and other forms of high quality information are of course important when students are maximally receptive, for example after they try to solve a problem and fail (Schwartz & Bransford, 1998). However, information delivery very often has a limited impact on the student when the content involves complex conceptualizations.

The obvious question that learning scientists have been asking over the years is why tutoring is effective in promoting learning? There are many approaches to answering this question. One approach is to conduct meta-analyses that relate learning gains with characteristics of the subject matter, tutee, tutor, and general structure of the tutoring session. There is evidence, for example, that (1) learning gains tend to be higher for well-structured, precise domains (mathematics, physics) than for ill-structured domains (reading), (2) that learning gains from tutors are more pronounced for tutees who start out with comparatively lower amounts of knowledge and skill, (3) that the quality of tutor training is much more important than the quantity of training, and that (4) that a tutoring session shows more benefits when there are particular pedagogical activities (Cohen et al., 1982; Fuchs et al., 1994; King, Staffieri, & Adelgais, 1998; Mathes & Fuchs, 1994; Rohrbeck et al., 2003).

A second approach is to perform a very detailed analysis of the tutoring session structure, tasks, curriculum content, discourse, actions, and cognitive activities manifested in the sessions and to speculate how these might account for the advantages of tutoring (Chi et al., Roy, & Hausmann, 2008; Chi et al., 2001; Graesser & Person, 1994; Graesser, Person, & Magliano, 1995; Hacker & Graesser, 2007; Lepper, Drake, & O'Donnell-Johnson, 1997; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Merrill, & Landes, 1995; Person & Graesser, 1999; 2003; Person, Kreuz, Zwaan, & Graesser, 1995; Shah, Evens, Michael, &

Rovick, 2002; VanLehn et al., 2003). This chapter will address these process factors in more detail.

A third approach is to manipulate the tutoring activities through trained human tutors or computer tutors and to observe the impact of the manipulations on learning gains (Chi et al., 2001, 2008; Graesser, Lu et al., 2004; Litman et al., 2003; VanLehn et al., 2003, 2007). Manipulation studies allow us to infer what characteristics of the tutoring directly cause increases in learning gains, barring potential confounding variables.

What are the frequent tutoring strategies and processes?

As discussed, the typical tutors in school systems are unskilled. These tutors are nonetheless effective in helping students learn so it is worthwhile to explore what tutoring strategies and processes they frequently implement. Graesser and Person analyzed the discourse patterns of 13 normal unskilled tutors in great detail (Graesser, & Person, 1994; Graesser et al., 1995; Person & Graesser, 1999). They videotaped over 100 hours of naturalistic tutoring in a corpus of unskilled tutors who tutored middle school students in mathematics or college students in research methods. The research team transcribed the tutorial dialogues, classified the speech act utterances into discourse categories, and analyzed the rate of particular discourse patterns. We refer to this as the Graesser-Person *unskilled tutor corpus*. Regarding expert tutors, Person et al. (2007) conducted a literature review of studies with accomplished tutors. Unfortunately, the sample sizes of expert tutors have been extremely small ($N < 3$) in empirical investigations of expert tutoring and often the same expert tutors are used in different research studies; at times the tutors are co-authors on publications. Claims about expert tutoring may therefore be biased by the idiosyncratic

characteristics of the small sample of tutors and the tutors' authorship role. Person et al. (2007) recently conducted a study on a sample of 12 tutors who were nominated by teachers in the Memphis community who were truly outstanding. The discourse patterns of these outstanding tutors in Person's *expert tutor corpus* were dissected in great detail.

Unfortunately, neither the unskilled tutor corpus nor the expert tutor corpus had outcome scores. There is a large void in the literature on detailed analyses of human tutorial dialogue that are related to outcome measures and that have a large sample of tutors. Part of the problem lies in logistical problems in obtaining such data. The subject matters of the tutoring sessions are difficult to predict in advance so it is difficult to proactively identify suitable pre-test and posttest measures from normative testbanks. Nevertheless, these tutoring corpora can be analyzed to identify the tutoring processes.

As one might expect, unskilled human tutors are not prone to implement sophisticated tutoring strategies that have been proposed in the fields of education, the learning sciences, and developers of ITSs (Graesser et al., 1995; Graesser, D'Mello, & Person, 2009; Person et al., 1995). Tutors rarely implement pedagogical techniques such as *bona fide* Socratic tutoring strategies, modeling-scaffolding-fading, Reciprocal Teaching, frontier learning, building on prerequisites, or diagnosis/remediation of deep misconceptions. In Socratic tutoring, the tutor asks learners illuminating questions that lead the learners to discover and correct their own misconceptions in an active, self-regulated fashion (Collins et al., 1975). Thus, Socratic tutoring is not merely bombarding the student with a large number of questions, as some practitioners and researchers erroneously believe. In modeling-scaffolding-fading, the tutor first models a desired skill,

then gets the learners to perform the skill while the tutor provides feedback and explanation, and finally fades from the process until the learners perform the skill all by themselves (Rogoff & Gardner, 1984). As discussed earlier, in Reciprocal Teaching, the tutor and learner take turns reading and thinking aloud with the goal of focusing on question generation, summarization, clarification, and prediction (Palincsar & Brown, 1984). Tutors who use frontier learning select problems and give guidance in a fashion that slightly extends the boundaries of what the learner already knows or has mastered (Sleeman & Brown, 1982). Tutors who build on prerequisites cover the prerequisite concepts or skills in a session before moving to more complex problems and tasks that require mastery of the prerequisites (Gagne, 1985).

One would expect tutors to be able to help the students correct their idiosyncratic deficits in knowledge and skills. Tutors are no doubt sensitive to some of these deficits but available data suggest there are limitations. Two examples speak to such limitations. If a tutor is truly adaptive to the student's learning profile, then the tutor should initiate some discussion or activity at the beginning of the session that diagnoses what the student is struggling with. This adaptation is manifested when the tutor: (a) inspects previous test materials and scores of the student, (b) selects problems in the tutoring session that are associated with the student's deficits, and (c) asks the tutee what they are having problems with. A tutor would lack the principle of adaptation if the tutor immediately presents problems to work on in a scripted fashion for all students. Whereas *a* and *c* occur with some frequency, tutors are not prone to do *b* (Chi et al., 2008; Graesser et al., 1995).

The second example addresses meta-cognitive knowledge of the student. Tutors frequently ask students comprehension-gauging questions, such as *Do you understand?*, *Are you following?*, and *Does that make sense?* If the student's comprehension calibration skills are accurate, then the student should answer YES when the tutee understands and NO when there is little or no understanding. One counterintuitive finding in the tutoring literature is that there sometimes is a positive correlation between a student's knowledge of the material (based on pre-test scores or post-test scores) and their likelihood of saying NO rather than YES to the tutors' comprehension-gauging questions (Chi, Bassock, Lewis, Reimann, & Glaser, 1989; Graesser et al., 1995). So it is the knowledgeable students who tend to say *No, I don't understand*. This result suggests that deeper learners have higher standards of comprehension (Baker, 1985; Otero & Graesser, 2001) and that many students have poor comprehension calibration skills. The finding that students have disappointing comprehension calibration is well documented in the metacognitive literature, where meta-analyses have shown only a 0.27 correlation between comprehension scores on expository texts and the students' judgments on how well they understand the texts (Dunlosky & Lipko, 2007; Glenberg, Wilkinson, & Epstein, 1982; Maki, 1998). It is perhaps not surprising that the student's comprehension calibration is poor because they are low in domain knowledge. From the perspective of the tutor, many tutors mistakenly believe the students' answers to the comprehension gauging questions and that reflects insensitivity to the students' knowledge states. A good tutor would periodically ask follow-up questions when students say YES, they understand.

The above examples suggest that human tutors are insensitive to the students' knowledge states, but such a generalization would be too sweeping. Tutors are often

adaptive to the students' knowledge and skills at a micro-level, as opposed to the macro-levels in the above two examples. The distinction is what VanLehn (2006) calls the inner loop versus the outer loop. The inner loop consists of covering individual steps or expectations within a problem whereas the outer loop involves the selection of problems, the judgment of mastery of a problem, and other more global aspects of the tutorial interaction. Available analyses of human tutoring suggest that human tutors are more sensitive to the students' knowledge at the inner loop than the outer loop.

Graesser and Person's analyses of tutorial dialogue uncovered a number of frequent dialogue structures (Graesser & Person, 1994; Graesser et al., 1995; Graesser, Hu, & McNamara, 2005). These structures are also prominent in the work of other researchers who have conducted fine-grained analyses of tutoring (Chi et al., 2001, 2004, 2008; Graesser, Hu, & McNamara, 2005; Litman et al., 2006; Shah et al., 2002). The following three dialogue structures are prominent: (1) the 5-step Tutoring Frame, (2) Expectation and Misconception Tailored (EMT) dialogue, and (3) Conversational Turn Management. All of these structures are in the inner loop: 3 is embedded in 2, which in turn is embedded in 1. It should be noted that it is the tutor who takes the initiative in implementing these structures, not the student. It is rare to have the student take charge of the tutorial session in a self-regulated manner.

5-Step Tutoring Frame. Once a problem or difficult main question is selected to work on, the 5-Step Tutoring Frame is launched, as specified below.

- (1) TUTOR asks a difficult question or presents a problem.
- (2) STUDENT gives an initial answer.

(3) TUTOR gives short feedback on the quality of the answer.

(4) TUTOR and STUDENT have a multi-turn dialogue to improve the answer.

(5) TUTOR assesses whether the student understands the correct answer.

This 5-step tutoring frame involves collaborative discussion, joint action, and encouragement for the student to construct knowledge rather than merely receiving knowledge.

The first 3 steps often occur in a classroom context, but the questions are easier short-answer questions. The Initiate-Respond-Evaluate (IRE) sequence in a classroom consists of the teacher initiating a question, the student giving a short-answer response, and the teacher giving a positive or negative evaluation of the response (Sinclair & Coulthart, 1975). This is illustrated in the exchange below on the subject matter of Newtonian physics.

(1) TEACHER: According to Newton's second law, force equals mass times what?

(2) STUDENT: acceleration

(3) TEACHER: Right, mass times acceleration.

Or

(2) STUDENT: velocity

(3) TEACHER: Wrong, it's not velocity, it is acceleration.

Thus, tutoring goes beyond the IRE sequence in the classroom by having more difficult questions and more collaborative interactions during step 4 of the 5-Step Tutoring Frame.

Expectation and Misconception Tailored (EMT) dialogue. Human tutors typically have a list of expectations (anticipated good answers, steps in a procedure) 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 problem below.

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?

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

The tutor 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 selected by the tutor to get the student to articulate 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.

Human tutors are dynamically adaptive to the learner in ways other than scaffolding them to articulate expectations. There also is the goal of correcting misconceptions that arise in the student’s responses. When the student articulates a misconception, the tutor acknowledges the error and corrects it. There is another conversational goal of giving

feedback to the student on their contributions. For example, the tutor gives short *feedback* on the quality of student contributions. The tutor accommodates a mixed-initiative dialogue by attempting to answer the student's questions when the student is sufficiently inquisitive to ask questions. However, it is well documented that students rarely ask questions, even in tutoring environments (Graesser & Person, 1994; Graesser, McNamara, & VanLehn, 2005), because they have limited self-regulated learning strategies (Azevedo & Cromley, 2004). Tutors are considered more adaptive to the student to the extent that they correct student misconceptions, give correct feedback, and answer student questions.

Conversational Turn Management. Human tutors structure their conversational turns systematically. Nearly every turn of the tutor has three information slots (i.e., units, constituents). The first slot of most turns is feedback on the quality of the learner's last turn. This feedback is either positive (*very good, yeah*), neutral (*uh huh, I see*), or negative (*not quite, not really*). The second slot advances the interaction with either prompts for specific information, hints, assertions with correct information, corrections of misconceptions, or answers to student questions. The third slot is a cue for the floor to shift from the tutor as the speaker to the learner. For example, the human ends each turn with a question or a gesture to cue the learner to do the talking.

The three conversational structures together present challenging problems or questions to the student, adaptively scaffold good answers through collaborative interactions, provide feedback when students express erroneous information, and answers student questions that infrequently are asked. What is absent is sophisticated pedagogical strategies. This is perhaps unsurprising because these strategies are complex and took

centuries to discover by scholars. However, it is a very important finding to document because it is conceivable that deep learning could improve tremendously by training human tutors and programming computer tutors to implement the sophisticated strategies.

The pedagogical strategies of expert tutors are very similar to those of unskilled tutors and computer tutors in most ways (Cade, Copeland, Person, & D'Mello, 2008; Person, Lehman, & Ozbun, 2007). However, Cade et al. (2008) did identify a few notable trends in pedagogy in the expert tutor corpus. The expert tutors did occasionally implement modeling-scaffolding-fading, although the relative frequencies of the dialogue moves for this pedagogical strategy were not impressively high. The tutors did a modest amount of modeling, a large amount of scaffolding, and very little fading. These tutors periodically had just-in-time direct instruction or mini-lectures when the student was struggling with a particular conceptualization. These content-sensitive mini-lectures allegedly were sensitive to what the student was having trouble with rather than being routinely delivered to all students. The expert tutors also appeared to differ from unskilled tutors on some meta-cognitive dimensions, as addressed below. However, it is important to qualify these claims about expert tutors because there was never a systematic comparison of tutors with different expertise in any given study. Instead, the relative frequencies of tutor strategies and discourse moves were computed in the expert tutor corpus and compared with the relative frequencies of the same theoretical categories in published studies with unskilled tutors. One pressing research need is to systematically compare tutors with varying expertise.

Meta-cognition and meta-communication

Graesser, D'Mello, and Person (2009) have documented some of the illusions that typical human tutors have about cognition and communication. These illusions may get in the way of optimizing learning. Expert tutors also may be less likely to fall prey to these illusions. The five illusions below were identified.

- (1) *Illusion of grounding.* The unwarranted assumption that the speaker and listener have shared knowledge about a word, referent, or idea being discussed in the tutoring session. Failure to establish common ground threatens successful communication and the joint construction of knowledge (Clark, 1996). A good tutor is sufficiently skeptical of the student's level of understanding that the tutor trouble-shoots potential communication breakdowns between the tutor and student.
- (2) *Illusion of feedback accuracy.* The unwarranted assumption that the feedback that the other person gives to a speaker's contribution is accurate. For example, tutors incorrectly believe the students' answers to their comprehension gauging questions (e.g., Do you understand?).
- (3) *Illusion of discourse alignment.* The unwarranted assumption that the listener does understand or is expected to understand the discourse function, intention, and meaning of the speaker's dialogue contributions. For example, tutors sometimes give hints, but the students do not realize they are hints.
- (4) *Illusion of student mastery.* The unwarranted assumption that the student has mastered much more than the student has really mastered. For example, the fact

that a student expresses a word or phrase does not mean that the student understands an underlying complex idea.

- (5) *Illusion of knowledge transfer*. The speaker's unwarranted assumption that the listener understands whatever the speaker says and thereby knowledge is accurately transferred. For example, the tutor assumes that the student understands whatever the tutor says, when in fact the student absorbs very little.

Both the tutor and student may each have these illusions and thereby compromise the effectiveness of tutoring.

These illusions undermine the tutor's building an accurate and detailed model of the cognitive states of the student, or what is called the *student model* by ITS researchers. Indeed, there are reasons for being pessimistic about the quality of the student model that tutors construct. A more realistic picture is that the tutor has only an approximate appraisal of the cognitive states of students and that they formulate responses that do not require fine tuning of the student model (Chi et al., 2004; Graesser et al., 1995). There are three sources of evidence for this claim. First, the short feedback to students on the quality of the students' contributions is often incorrect. In particular, the short feedback has a higher likelihood of being positive than negative after student contributions that are vague or error-ridden (Graesser et al., 1995). Tutors have the tendency to be polite or to resist discouraging the student by giving a large amount of negative feedback (Person et al., 1995). Second, tutors do not have a high likelihood of detecting misconceptions and error-ridden contributions of students (Chi, Siler, & Jeong, 2004; VanLehn et al., 2007). Third, as mentioned earlier, tutors do not select new cases or problems to work on that are

sensitive to the abilities and knowledge deficits of students (Chi et al., 2008). One would expect the selection of problems to be tailored to the student's profile according to the zone of proximal development, i.e., not too easy or not too hard, but just right. However, Chi et al. (2008) reported that there was no relation between problem selection and student's profile. Data such as these lead one to conclude that tutors have a modest ability to conduct student modeling.

A good tutor is sufficiently skeptical of the student's level of understanding. The tutor trouble-shoots potential communication breakdowns between the tutor and tutee. This is illustrated in the simple exchange below.

TUTOR: We know from Newton's law that net force equals mass times acceleration.

This law

STUDENT: Yeah, that is Newton's second law.

TUTOR: Do you get this?

STUDENT: Yeah. I know that one.

TUTOR: Okay, let's make sure. Force equals mass times what?

STUDENT: times velocity.

TUTOR: No, it's mass times acceleration.

A good tutor assumes that the student understands very little of what the tutor says and that knowledge transfer approaches zero. Person et al. (2007) has reported that expert tutors are more likely to verify that the student understands what the tutor expresses by asking follow up questions or giving follow-up trouble-shooting problems.

Emotions during tutoring

It is important to consider motivation and emotion in tutoring in addition to the cognitive subject matter. Indeed, connections between complex learning and emotions have received increasing attention in the fields of psychology and education (Deci & Ryan, 2002; Dweck, 2002; Gee, 2003; Lepper & Henderlong, 2000; Linnenbrink & Pintrich, 2002; Meyer & Turner, 2006). Studies who have tracked the emotions during tutoring have identified the predominate emotions, namely confusion, frustration, boredom, anxiety, and flow/engagement, with delight and surprise occurring less frequently (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello, Graesser et al., 2007, 2008). These data are informative, but the important question is how these emotions can be coordinated productively with learning.

The central assumption is that it is important for tutors to adopt pedagogical and motivational strategies that are effectively coordinated with the students' emotions. Lepper, Drake, and O'Donnell (1998) proposed an INSPIRE model to promote this integration. This model encourages the tutor to nurture the student by being empathetic and attentive to the student's needs, to assign tasks that are not too easy or difficult, to give indirect feedback on erroneous student contributions rather than harsh feedback, to encourage the student to work hard and face challenges, to empower the student with useful skills, and to pursue topics they are curious about. One of the interesting tutor strategies is to assign an easy problem to the student, but to claim that the problem is difficult and to encourage the student to give it a try anyway. When the student readily solves the problem, the student

builds self-confidence and self-efficacy in conquering difficult material (Winne, 2001; Zimmerman, 2001).

Several theories linking emotions and learning have been proposed. Meyer and Turner (2006) identified three theories that are particularly relevant to understanding the links between emotions and learning: academic risk taking, flow, and goals (Meyer & Turner, 2006). The academic risk theory contrasts (a) the adventuresome learners who want to be challenged with difficult tasks, take risks of failure, and manage negative emotions when they occur and (b) the cautious learners who tackle easier tasks, take fewer risks, and minimize failure and the resulting negative emotions (Clifford, 1991). According to flow theory, the learner is in a state of flow (Csikszentmihaly, 1990) when the learner is so deeply engaged in learning the material that time and fatigue disappear. When students are in the flow state, they are at an optimal zone of facing challenges and conquering the challenges by applying their knowledge and skills. Goal theory emphasizes the role of goals in predicting and regulating emotions (Dweck, 2002; Stein & Hernandez, 2007). Outcomes that achieve challenging goals result in positive emotions whereas outcomes that jeopardize goal accomplishment result in negative emotions.

A complementary perspective is to focus on learning impasses and obstacles rather than on flow and goals. Obstacles to goals are particularly diagnostic of both learning and emotions. For example, the affective state of confusion correlates with learning gains perhaps because it is a direct reflection of deep thinking (Craig et al., 2004; D'Mello et al., 2008; Graesser, Jackson, & McDaniel, 2007). Confusion is diagnostic of *cognitive disequilibrium*, a state that occurs when learners face obstacles to goals, contradictions,

incongruities, anomalies, uncertainty, and salient contrasts (Graesser, Lu, Olde, Pye-Cooper, & Whitten, 2005; Otero & Graesser, 2001). Cognitive equilibrium is ideally restored after thought, reflection, problem solving and other effortful deliberations. It is important to differentiate being productively confused, which leads to learning and ultimately positive emotions, from being hopelessly confused, which has no pedagogical value.

Research is conspicuously absent on how the tutees perceive the causes and consequences of these emotions and what they think they should do to regulate each affect state. The negative emotions are particularly in need of research. When a student is frustrated from being stuck, the student might attribute the frustration either to themselves (“I’m not at all good at physics”), the tutor (“My tutor doesn’t understand this either”), or the materials (“This must be a lousy textbook”). Solutions to handle the frustration would presumably depend on these attributions of cause of the frustration. When a student is confused, some students may view this as a positive event to stimulate thinking and show their metal in conquering the challenge; other students will attribute the confusion to their poor ability, an inadequate tutor, or poorly prepared academic materials. When a student is bored, they are likely to blame the tutor or material rather than themselves. Tutors of the future will need to manage the tutorial interaction in a fashion that is sensitive to the students’ emotions in addition to their cognitive states.

Tutoring Strategies that Influence Deep Learning

So far this chapter has provided evidence for the effectiveness of human tutoring and has identified various strategies and processes of naturalistic human tutoring. The obvious next question is which of the strategies help learning? Surprisingly, there is not an abundance of

research on this question because it is difficult to control what human tutors do in controlled experiments. Computer tutors are needed to impose the control that is needed to systematically manipulate tutoring strategies and observing the impact on learning. This research is covered in the next section.

Chi, Roy, and Hausmann (2008) compared 5 conditions in order to test a hypothesis they were advancing called the *active/constructive/interactive/observing* hypothesis. As the expression indicates, the hypothesis asserts that learning is facilitated from active student learning, knowledge construction, and collaborative interaction, as we have discussed in this chapter. The other aspect of the expression refers to observing a tutoring session vicariously. Their ideal experimental condition involves 4 people: two student participants watching and discussing a tutorial interaction between a tutor and another student. The participants learn a great deal from this interactive vicarious observation condition because it has all four components (action, construction, interaction, observation). To test this, students trying to learn physics were randomly assigned to this mean treatment (condition 1) and compared to one-on-one tutoring (condition 2), vicarious observation (all alone) of the tutoring session (condition 3), collaboratively interacting with another student without observing the interaction (condition 4), and studying from a text alone (condition 5). Conditions 1 and 2 were approximately the same in learning gains and significantly higher than conditions 3-5. So it appears that multiple components among the four are needed for learning to be optimal.

As discussed earlier, there is evidence that learning from tutorial interactions improves when the learner constructs explanations and when the student does more of the taking than the tutor (Litman et al., 2006; Siler & VanLehn, 2009). However, Chi et al. (2001) examined

the type of tutor moves in detail for students learning physics. For deep learners, it was the tutor moves that encouraged reflection that helped; for shallow learners, the tutor's responses to scaffolding and explanations were important. Unfortunately, the sample sizes in these studies are modest and very much in need of replication. The door is clearly open for discovering the particular dialogue moves of tutors predict learning.

Research on reading tutors have also investigated what aspects of tutoring help reading at deeper levels of comprehension (McNamara, 2007). Three notable examples are *Reciprocal Teaching* (Palincsar & Brown, 1984), *Self Explanation Reading Training, SERT* (McNamara, 2004), and *Questioning the Author* (Beck, McKeown, Hamilton, & Kucan, 1997). The available evidence, including meta-analyses and reviews of research (Roscoe & Chi, 2007; Rosenshine, Meister, & Chapman, 1996; Rosenshine & Meister, 1996), is that the scaffolding of explanations and of deep questions and answers are particularly important components. Explanations involve causal chains and networks, plans of agents, and logical justifications of claims. Deep questions have been defined systematically (Graesser & Person, 1994; Graesser, Ozuru, & Sullins, 2010), but include questions stem such as why, how, what if, what if not, and so what? In contrast, the strategy to predict future content in the text has little or no impact on improving reading at deeper levels, whereas summarization and clarification are somewhere in between.

Part of the challenge of conducting experimental research on human tutoring is that it is difficult to train tutors to adopt particular strategies. They rely on their normal conversational and pedagogical styles. It is nearly impossible to run repeated measures designs where a tutor adopts a normal style on some days and an experimental style on other days. The treatments

end up contaminating each other and it is difficult to force the human tutors to adopt changes in their language and discourse, particularly those levels that are unconscious and involuntary. However, computers can supply such experimental control. We turn to computer tutors in the next section.

Computer Tutors

The Intelligent Tutoring Systems (ITS) enterprise was launched in the late 1970's and christened with the edited volume aptly entitled *Intelligent Tutoring Systems* (Sleeman & Brown, 1982). The goal was to develop computerized learning environments that had powerful intelligent algorithms that would optimally adapt to the learner and to formulate computer moves that optimized learning for the individual learners.

ITS were viewed as a generation beyond computer-based training (CBT). A prototypical CBT system involves mastery learning, such that the learner (a) studies material presented in a lesson, (b) gets tested with a multiple choice test or another objective test, (c) gets feedback on the test performance, (d) re-studies the material if the performance in c is below threshold, and (e) progresses to a new topic if performance exceeds threshold. The order of topics presented and tested can follow different pedagogical models that range in complexity from ordering on prerequisites (Gagne, 1985) to a knowledge space model and Bayesian computation that attempts to fill learning deficits and correct misconceptions (Doignon & Falmagne, 1999) and to other models that allow dynamic sequencing and navigation (O'Neil & Perez, 2003). These more complex models are often viewed as bona fide ITS. The materials presented in a lesson can vary in CBT, including organized text with figures, tables, and diagrams (essentially books on the web),

multimedia, problems to solve, example problems with solutions worked out, and other classes of learning objects. CBT has been investigated extensively for decades and has an associated mature technology. Meta-analyses show effect sizes of 0.39 sigma compared to classrooms (Dodds & Fletcher, 2004), whereas Mayer's (2009) analyses of multimedia have substantially higher sigma. The amount of time that learners spend studying the material in CBT has a 0.35 correlation with learning performance (Taraban, Ryneerson, & Stalcup, 2001) and can be optimized by contingencies that distribute practice (Pashler et al., 2007).

These CBT systems are an important class of learning environments that can serve as tutors. However, the next generation of ITS went a giant step further that enhanced the adaptability, grain-size, and power of computerized learning environments. The processes of tracking knowledge (called user modeling) and adaptively responding to the learner incorporate computational models in artificial intelligence and cognitive science, such as production systems, case-based reasoning, Bayes networks, theorem proving, and constraint satisfaction algorithms. It is beyond the scope of this chapter to sharply divide systems that are CBTs versus ITS, but one dimension that is useful is the space of possible interactions that can be achieved with the two classes of systems. For ITS, every tutorial interaction is unique and the space of possible interactions is infinite. For CBT, interaction histories can be identical for multiple students and the interaction space is finite, if not small. Nevertheless, the distinction between CBT and ITS is not of central concern to this chapter other than to say that the distinction is blurry and that both classes of learning environments appear intelligent to the learners.

Successful systems have been developed for mathematically well-formed topics, including algebra, geometry, programming languages (the Cognitive Tutors, Anderson et al., 1995; Koedinger et al., 1997; Ritter et al., 2007), physics (Andes, Atlas, and Why/Atlas, VanLehn et al., 2002; VanLehn et al., 2007), electronics (Lesgold, Lajoie, Bunzo, & Eggen, 1992), and information technology (Mitrovic, Martin, & Suraweera, , 2007). These systems show impressive learning gains (1.00 sigma, approximately, Corbett, 2001; Dodds & Fletcher, 2004), particularly for deeper levels of comprehension. School systems are adopting ITSs at an increasing pace, particularly those developed at LearnLab and Carnegie Learning in the Pittsburgh area. ITS are expensive to build but are now in the phase of scaling up for widespread use. As mentioned earlier, the Cognitive Tutors are currently in over 2000 schools in the United States.

Recent ITS handle knowledge domains that have a stronger verbal foundation as opposed to mathematics and precise analytical reasoning. The Intelligent Essay Assessor (Landauer, Laham, & Foltz, 2000; Landauer, 2007) and e-Rater (Burstein, 2003) grade essays on science, history, and other topics as reliably as experts of English composition. AutoTutor (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Jeon, & Dufty, 2008; Graesser, Lu et al., 2004;) helps college students learn about computer literacy, physics, and critical thinking skills by holding conversations in natural language. AutoTutor shows learning gains of approximately 0.80 sigma compared with reading a textbook for an equivalent amount of time (Graesser, Lu et al., 2004; VanLehn, 2007). These systems automatically analyze language and discourse by incorporating recent advances in computational linguistics (Jurafsky & Martin, 2008) and information retrieval, notably

latent semantic analysis (Landauer, McNamara, Dennis, & Kintsch, 2007; Millis et al., 2004).

At this point we turn to some recent ITSs that have been tested on thousands of students and have proven effective in helping students learn. These systems also have scientific principles of learning that guide their design. Most of them fit within VanLehn's analyses of outer loop and inner loop, with step-by-step scaffolding of solutions to problems or answers to questions. The systems include the Cognitive Tutors, constraint-based tutors, case-based tutors, and tutors with animated conversational agents.

Cognitive Tutors

One of the salient success stories of transferring science to useful technology is captured in the Cognitive Tutors, the tutoring systems built by researchers at Carnegie Mellon University and produced by Carnegie Learning, Inc. Cognitive Tutor is built on careful research grounded in cognitive theory and extensive real-world trials, and has realized the ultimate goal of improvements over the status quo of classroom teaching. Its widespread implementation has drawn interest to both its inner psychological mechanisms and to its efficacy.

The Cognitive Tutor instruction is based on a cognitive model developed by Anderson (1990), called adaptive control of thought (ACT, or ACT-R in its updated form). The Cognitive Tutor has a curriculum of problems, with each problem having anticipated expectations and misconceptions, as we discussed in the previous section. The software can adaptively identify a student's problem solving strategy through their actions and comparisons with the expectations and misconceptions. This comparison process is called

model tracing. More specifically, the system constantly compares the student's actions to correct and incorrect potential actions that are represented in the cognitive model. Through these pattern matching comparison operations, the system can detect the many of the misconceptions that underlie the student's activities. The system is able to trace the student's progress using these comparisons and to give feedback when it is appropriate. When the system has decided that enough of the skill requirements have been met for mastery, the tutor and student move on to a new section.

Cognitive Tutor makes use of two kinds of knowledge, called declarative and procedural knowledge, to represent the skills the student must acquire in learning to solve a problem. Declarative knowledge is primarily concerned with static factual information that can readily be expressed in language. In contrast, procedural knowledge handles how to do things. Though procedural knowledge is often limited to a specific context, it is also more deeply ingrained in a student's knowledge structure and often acted upon more quickly. Conversely, declarative knowledge can be slower and more deliberate (particularly if it is not well-learned), but it applies to a broader range of situations than does procedural knowledge (Ritter et al., 2007).

Any common math problem consists of a combination of declarative and procedural knowledge. For example, consider a student trying to solve the problem below.

$$\begin{array}{r} 336 \\ +848 \\ \hline 4 \end{array}$$

The student would have to have the declarative knowledge component “ $6 + 8 = 14$ ” stored in memory in addition to making sure that the production rules are in place to retrieve this

fact. Production rules are contextualized procedural knowledge that shape the core of the Cognitive Tutor. Production rules help determine the manner in which student behavior is interpreted and also the knowledge students should gain as part of the learning process. In our example problem adopted from Anderson and Gluck (2001), one of the relevant production rules would be:

IF the goal is to add n_1 and n_2 in a column
& $n_1 + n_2 = n_3$
THEN set as a subgoal to write n_3 in that column.

According to ACT-R, an important part of cognition is simply a large set of production rules that accumulate in long-term memory, that get activated by the contents of working memory, and that are dynamically composed in sequence when a particular problem is solved. An important part of learning is accessing and mastering these production rules in addition to the declarative knowledge. Behaviors performed by the student reflect the production rules and declarative knowledge, so the system can reconstruct what knowledge the student has already mastered versus has yet to learn (Anderson & Gluck, 2001). Using this student model, ACT-R can target those knowledge components that are missing in a student's education and can select problems that specifically address those components. If successful, this would of course be a giant step forward beyond normal human tutoring.

Cognitive Tutor has a large number of skills for the student to learn. In four of their curricula (Bridge to Algebra, Algebra 1, Geometry or Algebra 2), there are 2,400 skills (or collection of knowledge components) for the student to learn (Ritter et al., 2009). This is a very large space of detailed content, far beyond school standards, the knowledge of human teachers, and conventional CBT. The goal of Cognitive Tutor is to scaffold the correct

methods for solving problems with the student until they become automatized after multiple problems in multiple contexts. This is accomplished by breaking down knowledge into smaller components, which can then be resurrected in particular problems in order to strengthen the student's knowledge into something well specified and procedural. Since any given task is made up of a combination of procedural and declarative knowledge components, the ultimate goal is to proceduralize the retrieval of this declarative knowledge in order to speed up and strengthen its availability, thereby making the right facts and procedures highly accessible during problem solving (Ritter et al., 2007). Additionally, a series of well-learned knowledge components can be put together to solve larger, more complex tasks (Lee & Anderson, 2001).

Cognitive Tutor also offers helps and hints, which aid the student in problem solving. Students can see their progress in Cognitive Tutor by looking at their *skillmeter*, which logs how many skills the student has acquired and depicts them in progress bars. The system gives just-in-time feedback when the student's knowledge is buggy (Ritter et al., 2007). Cognitive Tutor includes a mechanism where the student can solicit hints to overcome an impasse. More specifically, there are three levels of hints. The first level may simply remind the student of the goal, whereas the second level may offer more specific help, and the third level comes close to directly offering the student the answer for a particular step in the problem solving. An example of an intermediate hint would be "As you can see in the diagram, Angles 1 and 2 are adjacent angles. How does this information help to find the measure of Angle 2?" when the student is learning about angles (Roll et al., 2006). These hints can be highly effective when used properly. However, some students

attempt to “game the system,” or abuse the hint function to quickly get through a lesson (Aleven, McLaren, Roll, & Koedinger, 2006; Baker, Corbett, Koedinger, & Wagner, 2004). Gaming the system has been associated with lower learning outcomes for students and may be a consequence of learned helplessness.

As discussed earlier in this chapter, Cognitive Tutor is a widely used ITS and produces impressive learning gains in both experimental and classroom settings. Corbett (2001) tested various components of a cognitive tutor that teaches computer programming, looking for those aspects of the system that produce the most learning. Modeling tracing, when compared with no model tracing, had an effect size of 0.75 sigma. When students were exposed to a model tracing system that encourages mastery of skills, Corbett found a 0.89 sigma over simple model tracing. Significant effect sizes have also been found in their investigations into Cognitive Tutor in the school system. Their first classroom studies in Pittsburgh showed that Cognitive Tutor students excelled with an average of 0.6 sigma when compared to students in a traditional Algebra class. According to Ritter et al. (2007), standardized tests show overall effect sizes of 0.3 sigma and the *What Works Clearinghouse* investigations show an effect size of 0.4 sigma.

To be accurate, however, the results of the Cognitive Tutor are not uniformly rosy. For example, the representations and problems in which Cognitive Tutor students showed extremely high gains over traditional algebra students were experimenter-designed (Koedinger et al., 1997). A large-scale study in Miami with over 6,000 students showed that Cognitive Tutor students scored 0.22 sigma over their traditional Algebra student counterparts, but only scored 0.02 sigma better than the traditional students on the

statewide standardized test (Shneyderman, 2001). It is widely acknowledged that there are challenges in scaling up any intervention and that the results will invariably be mixed.

Cognitive Tutor will continue to be modified in the future to optimize learning gains in a wide range of contexts and populations.

Constraint-based Tutors

Constraint-based modeling (CBM) is an approach first proposed by Ohlsson (Ohlsson, 1994, 1992) and later extended by (Ohlsson & Mitrovic, 2007). The core idea of CBM is to model the declarative structure of a good solution rather than the procedural steps leading to a good solution. Thus, CBM contrasts heavily with the model-tracing approach to student modeling which models each step of an expert solution, perhaps ignoring alternate solutions. CBM instead has much conceptual similarity with declarative styles of programming, like Prolog (Bratko, 1986), which define relationships between entities rather than operations on entities. This abstraction thus focuses on what properties a good solution must have rather than how it is obtained.

Ohlsson (1994) gives a concrete example of constraint-based modeling in the domain of subtraction. Subtraction has two core concepts, each giving rise to a constraint. The first core concept is place value, meaning that the position of the digit affects its quantity, e.g. 9 in the tens place represents the quantity 90. The second core concept of subtraction is regrouping, in which the digits expressing a quantity may change without changing the value of the quantity, e.g. $90 = 9*10 + 0*1 = 80*10 + 10*1$, so long as the decrement in one digit is offset by an increment in the other. The two constraints that follow from these core concepts are:

Constraint 1: Increments and corresponding decrements must occur together

(otherwise the value of the numeral has changed)

Constraint 2: An increment of ten should not occur unless the digit in the position to the left is decremented by one.

The key observation is that a correct solution can never violate either of these constraints, no matter what order of operations is followed. Thus, the style of constraints is declarative rather than procedural.

In CBM, the declarative structure of a good solution is composed of a set of state constraints (Ohlsson, 1994). Each constraint is composed of a relevance condition (R) and a satisfaction condition (S). The relevance condition specifies when the constraint is relevant. Only in these conditions is the state constraint meaningful. The satisfaction condition specifies whether the state constraint has been violated. A relevant, satisfied state constraint corresponds to an aspect of the solution that is correct. A relevant, unsatisfied state constraint indicates a problem in the solution.

There allegedly are two major advantages to CBM over traditional student models like model-tracing and buggy libraries (Ohlsson, 1994). First, CBM is able to account for a wider array of student behavior, i.e. greater deviations from the “correct” solution path, than either of these methods. This advantage stems from the basic property of CBM that solution paths aren't modeled. The second major advantage of CBM is that it is substantially less effort- intensive to create student models using CBM than traditional methods. Both of these claimed advantages have been explored in the literature, leading to a more elaborate and nuanced understanding of the tradeoffs between CBM and other

modeling approaches.

Several CBM tutoring systems have been built by Mitrovic and colleagues with encouraging results (Mitrovic, Martin, & Suraweera, 2007; Mitrovic & Ohlsson, 1999; Mitrovic, Martin, & Mayo, 2002; Suraweera & Mitrovic, 2004). Particularly noteworthy are those that support learning of database and SQL (Structured Query Language) design. These systems have been incorporated into Addison-Wesley's *Database Place*, accessible by anyone who has bought a database textbook from Addison-Wesley (Mitrovic, Martin, & Suraweera, 2007). KERMIT (Suraweera & Mitrovic, 2004) is an entity-relationship tutor that focuses on database design. In a two hour pretest-posttest study with randomized assignment, students using KERMIT had significantly higher learning gains than students from the same class who used KERMIT with disabled constraint-based feedback, with an effect size of 0.63.

It is informative to compare CBM with the model-tracing architecture of the Cognitive Tutors. This has been one of the fundamental debates in the ITS literature in recent years. Table 1 contrasts properties of the two models according to Mitrovic, Koedinger, and Martin (2003). For example, it appears that CBM is less effortful to build, but less capable of giving specific advice to the learner, whereas model-tracing is the opposite: more effortful to build but more capable of giving specific advice (Mitrovic et al., 2003). Kodaganallur, Weitz, and Rosenthal (2006) conducted a more thorough investigation of two complete systems in the domain of hypothesis testing. When compared with model-tracing, CBM accounted for a narrower array of student behavior, required buggy libraries, was unable to give procedural remediation, was incapable of

giving fine grained feedback, and was likely to give incorrect feedback to proper solutions. However, a number of methodological problems with this analysis were noted by Mitrovic and Ohlsson (2007). It suffices to say that the debate continues on the relative strengths and liabilities of these two architectures, but it is beyond the scope of this chapter to elaborate and resolve the controversy.

Table 1. Comparative analysis of CBM and MT (Mitrovic, Koedinger, & Martin, 2003)

Property	Model Tracing	Constraint-Based Modeling
Knowledge representation	Production rules (procedural)	Constraints (declarative)
Cognitive fidelity	Tends to be higher	Tends to be lower
What is evaluated	Action	Problem state
Problem solving strategy	Implemented ones	Flexible to any strategy
Solutions	Tend to be computed, but can be stored	One correct solution stored, but can be computed
Feedback	Tends to be immediate, but can be delayed	Tends to be delayed, but can be immediate
Problem-solving hints	Yes	Only on missing elements, but not strategy
Problem solved	'Done' productions	No violated constraints
Diagnosis if no match	Solution is incorrect	Solution is correct
Bugs represented	Yes	No
Implementation effort	Tends to be harder, but can be made easier with loss of other advantages	Tends to be easier, but can be made harder to gain other advantages

Case-based Reasoning

ITS with Case-based reasoning (CBR) are inspired by cognitive theories in psychology, education, and computer science that emphasize the importance of specific cases, exemplars, or scenarios in constraining and guiding our reasoning (Leake, 1996;

Ross, 1987; Schank, 1999). There are two basic premises in CBR that are differentially emphasized in the literature. The first basic premise is that memory is organized by cases consisting of a problem description, its solution, and associated outcomes (Watson & Marir, 1994). Accordingly, problem solving makes use of previously encountered cases rather than by proceeding from first principles. Although there is wide variation in the field as to the adaptation and use of cases, implemented systems generally follow four steps (Aamodt & Plaza, 1994):

RETRIEVE the most similar case(s); (indexing problem)

REUSE the case(s) to attempt to solve the problem; (adaptation problem)

REVISE the proposed solution if necessary, and

RETAIN the new solution as a part of a new case.

The second basic premise of CBR is that memory is dynamically organized around cases (Schank, 1999), meaning that the outcome of the four steps above can not only cause cases to be re-indexed using an existing scheme but can also drive the indexing scheme itself to change. Therefore learning in the CBR paradigm goes beyond adding new cases after successful REUSE and beyond adding new cases after failed REUSE and successful REVISION. Even though success and failure can drive the creation of new cases, they can also drive the way all existing cases are organized and thus their future use (Leake, 1996).

Like ACT-R, CBR has been used as an underlying theory for the development of learning environments (Kolodner et al., 2003; Kolodner, Cox, & Gonzalez-Calero, 2005; Schank, Fano, Bell, & Jona, 1994). As a theory of learning, CBR implies that human learners engage in case-based analogical reasoning as they solve problems and learn by

solving the problems. One might expect CBR learning environments to proceed in a similar fashion to model-tracing and CBM tutors: implement a tutor capable of solving the problem, allow the student to solve the problem, and then provide feedback based on the differences between the student's solution and the tutors. However, perhaps because of the complications involved with implementing a model-based reasoner on top of a CBR system, an ITS analogous to model-tracing and constraint-based modeling has yet to be implemented. Instead, CBR learning environments give learners the resources to implement CBR on their own. That is, designers of these CBR systems present learners with activities designed to promote CBR activities: identifying problems, retrieving cases, adapting solutions, predicting outcomes, evaluating outcomes, and updating a case library.

There are two CBR paradigms that proceed in this fashion of facilitating design by human learners. The first is exemplified by two environments called *Goal-Based Scenarios* (Schank, Fano, Bell, & Jona, 1994) and *Learning by Design* (Kolodner, Cox, & Gonzalez-Calero, 2005), both of which are highly integrated with classroom activities. Goal-Based Scenarios use simulated worlds as a context for learning (Schank et al., 1994). For example *Broadcast News*, puts students in the scenario of having to create their own news program. Cases are news sources from the previous day, and these are RETRIEVED via tasks students perform to establish the social issues in each story. Experts are available to answer questions and provide feedback, thereby helping the students complete the REUSE and REVISE phases. In contrast, *Learning by Design* frames the learning task as a design task (Kolodner et al., 2003). An example of this is designing a parachute. The *Learning by Design* addresses this task in 6 phases, with learning occurring in groups: clarifying the

question, making a hypothesis, designing an investigation (using existing cases as input, REUSE), conducting the investigation (REUSE), analyzing results (REVISE), and group presentation. Kolodner et al. (2005) review educational outcomes in this paradigm and draw two conclusions. The first is that the reviewed CBR classes have significantly larger simple learning gains (posttest – pretest) compared to control classrooms, although effect sizes have not been reported. The second conclusion is that CBR classes have greater skill in collaborating and in scientific reasoning than matched peers. However, more rigorous tests of these claims await future research.

The second CBR paradigm involves support from a computer environment for one-on-one learning. Perhaps the best known work of this type is the law-based learning environment by Alevan and colleagues (Alevan, 2003; Ashley & Brüninghaus, 2009). The most current system, CATO, is a CBR system for legal argumentation, i.e. the arguments attorneys make using past legal cases. CATO uses its case library and domain background knowledge to organize multi-case arguments, reason about significant differences between cases, and determine which cases are most relevant to the current situation. Students using CATO practice two types of tasks, theory-testing and legal argumentation, both of which rely heavily on CATO's case library. Theory-testing requires the student to predict a ruling on a hypothetical case by first forming a hypothesis, retrieving relevant cases from CATO (RETRIEVE), and then evaluating their hypothesis in light of the retrieved cases (REUSE, REVISE). Legal argumentation requires the student to write legal arguments for both the defendant and plaintiff on a hypothetical case. Students first study the hypothetical case and then retrieve relevant cases from CATO (RETRIEVE). Next the students study example

arguments that CATO generates dynamically based on the selected cases, in a kind of multi-case REUSE/REVISE. Students iteratively use this dynamic generation capability to explore the outcomes of combining different sets of cases, successively refining their arguments until they are complete (RETRIEVE/REUSE/REVISE). For argumentation, learning with CATO was not significantly different than learning from a human instructor when matched for time and content, in both cases in a law school setting (Alevan, 2003).

Conversational Agents

Animated conversational agents play a central role in some of the recent advanced learning environments (Atkinson, 2002; Baylor & Kim, 2005; Gholson et al., 2009; Graesser, Chipman, Haynes, & Olney, 2005; Johnson, Rickel, & Lester, 2000; McNamara, Levinstein, & Boonthum, 2004; Moreno & Mayer, 2007; 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.

AutoTutor is an intelligent tutoring system that helps students learn through tutorial dialogue in language (Graesser, Dufty, & Jeon, 2008; Graesser, Hu, & McNamara, 2005; Graesser et al., 2005). AutoTutor's dialogues are organized around difficult questions and problems that require reasoning and explanations in the answers. For example, below 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 3-7 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 50-100 conversational turns between AutoTutor and the student.

The structure of the dialogue in both AutoTutor attempts to simulate that of human tutors, as was discussed earlier. More specifically, AutoTutor implements the three conversational structures: (1) a 5-step dialogue frame, (2) expectation and misconception-

tailored dialogue, and (3) conversational turn management. These three levels can be automated and produce respectable tutorial dialogue. 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, as we discussed in the previous section on human tutoring, (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 learning gains of AutoTutor have been evaluated in over 20 experiments conducted during the last 12 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 et al., 2004) and Newtonian physics (VanLehn, Graesser et al., 2007). AutoTutor's learning gains have varied between 0 and 2.1 sigma (a mean of 0.8), depending on the learning performance measure, the comparison condition, the subject

matter, and the version of AutoTutor. Approximately a dozen measures of learning have been collected in these assessments on the topics of computer literacy and physics, including: (1) multiple choice questions on shallow knowledge that tap definitions, facts and properties of concepts, (2) multiple choice questions on deep knowledge that taps causal reasoning, justifications of claims, and functional underpinnings of procedures, (3) essay quality when students attempt to answer challenging problems, (4) a cloze task that has subjects fill in missing words of texts that articulate explanatory reasoning on the subject matter, and (5) performance on problems that require problem solving. AutoTutor is most impressive for the multiple choice questions that tap deep reasoning.

The agents described above interact with students one-to-one. Learning environments can also have pairs of agents interact with the student as a triologue 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.

As an example, *iSTART (Interactive Strategy Trainer for Active Reading and Thinking)* is an automated strategy trainer that helps students become better readers by constructing self-explanations of the text (McNamara et al., 2004). The construction of self-explanations during reading is known to facilitate deep comprehension (Chi et al., 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 (an instructor and two 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. 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 K12 and college (McNamara, Best, O'Reilly, & Ozuru, 2006). The three-phase iSTART training (approximately 3 hours) has been compared with a control condition that didactically trains students on self-explanation, but without any 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 (.04 to 1.4 sigma) for strategy use and for comprehension.

Future Directions

This chapter has hopefully made a convincing case that tutoring by humans and computers is a powerful learning environment. It could be argued that tutoring is the most effective learning environment we know of in addition to being the oldest. Tutoring has been around for millennia and has been shown to help learning in several meta-analyses. However, there are still a large number of unanswered fundamental questions that need attention in future research.

Rather surprisingly, there needs to be a systematic line of research that investigates the impact of tutoring expertise on learning gains as well as learner emotions and motivation. We had hoped to find a rigorous study that randomly assigns students to tutors with varying levels of expertise and that collects suitable outcome measures. The fact that we came up empty is remarkable, but it also sets the stage for new research initiatives. To what extent is student learning and motivation facilitated as a function of increased tutor training on pedagogy and/or increased subject matter knowledge? To what extent does tutoring experience matter? How do different schools of tutoring pedagogy compare? Are there interactions between tutor pedagogy, subject matter, and student profiles? How do we best train tutors? Decades of research is needed to answer these questions?

Computer tutors allow more control over the tutoring process than human tutors can provide. This opens up the possibility of new programs of research that systematically compare different versions of an ITS and different types of ITS. All ITS have multiple

modules, such as the knowledge base, the student's ability and mastery profile, decision rules that select problems, scaffolding strategies, help systems, feedback, media on the human-computer interface, and so on. Which of these components are responsible for any learning gains of the ITS? It is possible to systematically manipulate the quality or presence of each component in *lesion* studies. The number of conditions in manipulation studies of course grows with the number of components. If there are 6 major components, with each level varying in 2 levels of quality, then there would be $2^6 = 64$ conditions in a factorial design. That would require nearly 2000 students in a between-subjects design with 30 students randomly assigned to each of the 64 conditions. It indeed might be realistic to perform such a lesion study to the extent that the ITS enterprise scales up and delivers training on the web (see Heffernan, Koedinger, & Razzaq, 2008). The alternative would be to selectively focus on one or two modules at a time.

Comparisons need to be made different computer tutors that handle the same subject matter. Algebra, for example, can be trained with the Cognitive Tutors, ALEKS, Constraint-based Models, and perhaps even Case-based learning environments. Which of these provides the most effective learning for different populations of learners? It may be that there are aptitude-treatment interactions and no clear winner. Eventually we would sort out which tutoring architecture works best for each population of learners. The subject matters with verbal reasoning, as opposed to mathematical computation, may need a different ITS architecture. The conversational agents are expected to play an important role in these topics that require verbal reasoning. Questions remain on how effective the conversational agents are compared to the more conventional graphical user interfaces. Is it

best to make the interface resemble a face-to-face conversation with a human? Or does such anthropomorphic realism present a distraction from the subject matter?

One of the provocative tests in the future will pit human versus machine as tutors. Most people place their bets on the human tutors under the assumption that they will be more sensitive to the student's profile and be more creatively adaptive in guiding the student. However, the detailed analyses of human tutoring challenge such assumptions in light of the many illusions that humans have about communication and the modest pedagogical strategies in their repertoire. Computers may do a better job in cracking the illusions of communication, in inducing student knowledge states, and in implementing complex intelligent tutoring strategies. A plausible case could easily be made for betting on the computer over the human tutor. Perhaps the ideal computer tutor emulates humans in some ways and complex non-human computations in other ways. Comparisons between human and computer tutors need to be made in a manner equilibrates the conditions on content, time on task, and other extraneous variables that are secondary to pedagogy. As data roll in from these empirical studies, we make only one prediction with any semblance of confidence: There will be unpredictable and counterintuitive discoveries.

References

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39-59.
- Aleven, V. (2003). Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence*, 150(1-2), 183-237.
- Aleven, V., McLaren, Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16, 101-128.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.
- Anderson, J. R. & Gluck, K. (2001). What role do cognitive architectures play in intelligent tutoring systems? In D. Klahr & S. M. Carver (Eds.) *Cognition & Instruction: Twenty-five years of progress*, 227-262. Hillsdale, NJ: Erlbaum.
- Ashley, K.D., & Brüninghaus, S. (2009). Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17, 125-165.
- Atkinson, R. (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. (1985). Differences in standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, 20, 298-313.
- Baker, R.S., Corbett, A.T., Koedinger, K.R., & Wagner, A.Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "Game the System". *Proceedings of ACM CHI 2004: Computer-Human Interaction*, 383-390.
- Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15, 5-115.
- Beck, I.L., McKeown, M.G., Hamilton, R.L., & Kucan, L. (1997). *Questioning the Author: An approach for enhancing student engagement with text*. Delaware: International Reading Association.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn* (expanded ed.). Washington, D.C.: National Academy Press.
- Bratko, I. (1986). *Prolog Programming for Artificial Intelligence*. Wokingham, England: Addison Wesley Publishing Company.
- Burstein, J. (2003). The E-rater scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*, 133-122. Mahwah, NJ: Erlbaum.
- Cade, W., Copeland, J. Person, N., and D'Mello, S. K. (2008). Dialogue modes in expert tutoring. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the*

- Ninth International Conference on Intelligent Tutoring Systems* (pp. 470-479). Berlin, Heidelberg: Springer-Verlag
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.
- 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., Roy, M., & Hausmann, R. G. M. (2008) Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, *32*(2), 301-341.
- Chi, M.T.H., Siler, S.A. & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, *22*(3), 363-387.
- Chi, M.T.H., Siler, S., Yamauchi, T., Jeong, H. & Hausmann, R. (2001). Learning from human tutoring. *Cognitive Science*, *25*, 471- 534.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- 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., Hacıoglu, K., Ma, J., & Movellan, J., (2003). Perceptive animated interfaces: First steps toward a new paradigm for human computer interaction. *Proceedings of the IEEE*, *91*, 1391-1405.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. New York: Teacher College Press.
- Collins, A., Warnock, E. H., Aeillo, N., Miller, M. L. (1975). Reasoning from incomplete knowledge. In D. G. Bobrow A. Collins (Eds.), *Representation and understanding* (pp. 453-494). New York: Academic.
- Corbett, A.T. (2001). Cognitive computer tutors: Solving the two-sigma problem. *User Modeling: Proceedings of the Eighth International Conference, UM 2001*, 137-147.
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling & User-Adapted Interaction*, *4*, 253-278.
- Craig, S.D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, *29*, 241-250.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper-Row.
- D'Mello, S. K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., and Graesser, A. C. (2008). Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User-Adapted Interaction*, *18*(1-2), 45-80.
- D'Mello, S.K., Picard, R., & Graesser, A.C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*, 53-61.

- Deci, E. L., & Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
- Dodds, P. V. W., & Fletcher, J. D. (2004) Opportunities for new “smart” learning environments enabled by next generation web capabilities. *Journal of Education Multimedia and Hypermedia*, 13(4), 391-404.
- Doignon, J.P. & Falmagne, J. C. (1999). *Knowledge Spaces*. Berlin, Germany: Springer.
- Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228-232.
- Dweck, C. S. (2002). Messages that motivate: How praise molds students’ beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
- Fuchs, L., Fuchs, D., Bentz, J., Phillips, N., & Hamlett, C. (1994). The nature of students’ interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart, & Winston.
- Gee, J.P. (2003). *What video games have to teach us about language and literacy*. New York: Macmillan.
- Gholson, B., Witherspoon, A., Morgan, B., Brittingham, J. K., Coles, R., Graesser, A. C., Sullins, J., & Craig, S. D. (2009). Exploring the deep-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer literacy and Newtonian physics. *Instructional Science*, 37, 487-493.
- Glenberg, A. M., Wilkinson, A. C., and Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10, 597-602.
- Graesser, A. C., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618.
- Graesser, A. C., D’Mello, S. K., & Person, N., (2009). Meta-knowledge in tutoring. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.
- Graesser, A.C., Hu, X., & McNamara, D.S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A.F. Healy (Ed.), *Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. Washington, D.C.: American Psychological Association.
- 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., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, 45, 298–322.

- Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180-193.
- Graesser, A.C., Lu, S., Olde, B.A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, 33, 1235-1247.
- 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., & 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, 1-28.
- Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the TRG (1999). Auto Tutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, 1, 35-51.
- 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.
- Hacker, D.J., & Graesser, A.C. (2007). The role of dialogue in reciprocal teaching and naturalistic tutoring. In R. Horowitz (Ed.), *Talk about text: How speech and writing interact in school learning*. Mahwah, NJ: Erlbaum.
- Heffernan, N.T., Koedinger, K.R., & Razzaq, L. (2008) Expanding the model-tracing architecture: A 3rd generation intelligent tutor for Algebra symbolization. *The International Journal of Artificial Intelligence in Education*. 18(2). 153-178
- Johnson, D.W., & Johnson, R.T. (1992). Implementing cooperative learning. *Contemporary Education*, 63(3), 173-180.
- Johnson, W.L., & Beal, C. (2005). Iterative evaluation of a large-scale, intelligent game for learning language. In C. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Artificial Intelligence in Education* (pp. 290-297). Amsterdam: IOS Press.
- 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. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.
- King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, 90, 134-152.
- Kodaganallur, V., Weitz, R. R., & Rosenthal, D. (2006). An assessment of constraint-based Tutors: A response to Mitrovic and Ohlsson's critique of "A comparison of model-tracing and constraint-based intelligent tutoring paradigms". *International Journal of Artificial Intelligence in Education*, 16(3), 291-321.

- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.
- Kolodner, J., Camp, P., Crismond, D., Fasse, B., Gray, J., & Holbrook, J.. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *Journal of the Learning Sciences*, 12, 495-547.
- Kolodner, J., Cox, M., & Gonzalez-Calero, P. (2005). Case-based reasoning-inspired approaches to education. *The Knowledge Engineering Review*, 20(3), 299-303.
- Landauer, T. K. (2007) LSA as a theory of meaning. In T. Landauer, D. McNamara, D. Simon, & W. Kintsch (Eds.) *Handbook of Latent Semantic Analysis*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- Landauer, T., McNamara, D. S., Dennis, S., Kintsch, W. (2007) (Eds.), *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum
- Leake, D. (1996). CBR in context: The present and future. In *Case-Based Reasoning: Experiences, Lessons, and Future Directions* (pp. 3-30). AAAI Press/MIT Press.
- Lee, F. J. & Anderson, J. R. (2001). Does learning of a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, 42(3), 267-316.
- Lehman, B. A., Matthews, M., D'Mello, S. K., and Person, N. (2008). Understanding students' affective states during learning. *Ninth International Conference on Intelligent Tutoring Systems (ITS'08)*.
- Lepper, M. R., Drake, M., & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds), *Scaffolding student learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.
- Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp.257-307). San Diego, CA: Academic Press.
- Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.
- Lesgold, A., Lajoie, S. P., Bunzo, M., & Eggan, G. (1992). SHER-LOCK: A coached practice environment for an electronics trouble-shooting job. In J. H. Larkin & R. W. Chabay (Eds.), *Computer assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 201-238). Hillsdale, NJ: Erlbaum.
- Linnenbrink, E. A. & Pintrich, P. (2002). The role of motivational beliefs in conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Litman, D.J, Rose, C.P., Forbes-Riley, K., VanLehn, K., Bhembé, D., and Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education*, 16, 145-170.

- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.
- Mathes, P. G., & Fuchs, L. S. (1994). Peer tutoring in reading for students with mild disabilities: A best evidence synthesis. *School Psychology Review*, 23, 59-80.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.
- McArthur, D., Stasz, C., & Zmuidzinis, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197 - 244.
- McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38, 1-30.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, 36, 222 - 233.
- McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34, 147-171.
- Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction*, 13(3), 315-372.
- Meyer, D. K., & Turner, J. C. (2006). Re-conceptualizing Emotion And Motivation To Learn In Classroom Contexts. *Educational Psychology Review*, 18 (4), 377-390.
- Millis, K., Kim, H. J., Todaro, S. Magliano, J. P., 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.
- Mitrovic, A., Koedinger, K., & Martin, B. (2003). A comparative analysis of cognitive tutoring and constraint-based modeling. In *User Modeling 2003*.
- Mitrovic, A., Martin, B., & Mayo, M. (2002). Using evaluation to shape ITS design: Results and experiences with SQL-Tutor. *User Modeling and User-Adapted Interaction*, 12(2), 243-279.
- Mitrovic, A., Martin, B., & Suraweera, P. (2007). Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems*, 22(4), 38-45.
- Mitrovic, A., & Ohlsson, S. (2006). A Critique of Kodaganallur, Weitz and Rosenthal, "A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms". *International Journal Artificial Intelligence in Education*, 16(3), 277-289.
- Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *International Journal on Artificial Intelligence in Education*, 10(3-4), 238-256.
- Mitrovic, A., Suraweera, P., Martin, B. and Weerasinghe, A. (2004) DB-suite: Experiences with three intelligent, web-based database tutors. *Journal of Interactive Learning Research* 15(4), 409-432.
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19, 309-326.

- O'Neil, H. & Perez, R. (Eds.). (2003). *Web-based learning: Theory, Research and Practice*. Mahwah NJ: Lawrence Erlbaum Associates.
- Ohlsson, S. (1994). Constraint-based student modeling. In J. E. Greer & G. McCalla (Eds.), *Student modelling: the key to individualized knowledge-based instruction* (pp. 167-190). Birkhäuser.
- Ohlsson, S. (1992). Constraint-based student modelling. *International Journal of Artificial Intelligence in Education*, 3(4), 429-447.
- Ohlsson, S., & Mitrovic, A. (2007). Fidelity and efficiency of knowledge representations for intelligent tutoring systems. *Technology, Instruction, Cognition and Learning (TICL)*, 5(2-3-4), 101-132.
- Otero, J., & Graesser, A.C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction*, 19, 143-175.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension- fostering and monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Palincsar, A. S., & Brown, A. L. (1988). Teaching and practicing thinking skills to promote comprehension in the context of group problem solving. *Remedial and Special Education (RASE)*, 9(1), 53-59.
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., & McDaniel, M., (2007). Organizing instruction and study to improve student learning. *IES practice guide* (NCER 2007-2004). Washington, DC: National Center for Education Research.
- Person, N. K., & Graesser, A. C. (1999). Evolution of discourse in cross-age tutoring. In A. M.O'Donnell and A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 69-86). Mahwah, NJ: Erlbaum.
- Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13, 161-188.
- Person, N., Lehman, B., & Ozbun, R. (2007). Pedagogical and motivational dialogue moves used by expert tutors. Presented at the 17th Annual Meeting of the Society for Text and Discourse. Glasgow, Scotland.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale NJ: Erlbaum.
- Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions, and new media like real people and places*. University Press, Stanford, California.
- Ritter, S., Anderson, J. R., Koedinger, K. R., Corbett, A. (2007) Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14, 249-255.
- Ritter, S., Harris, T., Nixon, T., Dickison, D., Murray, R.C. & Towle, B. (2009.) Reducing the knowledge tracing space. Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (Eds.) Educational Data Mining 2009. 151-160.
- Rogoff, B. & Gardner, W., (1984). Adult guidance of cognitive development. In: Rogoff, B. and Lave, J., Editors, 1984. *Everyday cognition: Its development in social context*, Harvard University Press, Cambridge, MA, pp. 95-116.

- Rohrbeck, C. A., Ginsburg-Block, M., Fantuzzo, J. W., & Miller, T. R. (2003). Peer assisted learning Interventions with elementary school students: A Meta-Analytic Review. *Journal of Educational Psychology*, 95 (2), 240-257.
- Roll, I., Alevan, V., McLaren, B. M., Ryu, E., Baker, R. S., & Koedinger, K. R. (2006). The Help Tutor: Does metacognitive feedback improve students' help-seeking actions, skills and learning? *8th International Conference in Intelligent Tutoring Systems*, 360-369.
- Roscoe, R.D., & Chi, M.T.H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77, 534-574.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479-530.
- 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.
- Ross, R.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.
- Rus, V., & Graesser, A.C. (2006). Deeper natural language processing for evaluating student answers in intelligent tutoring systems. In the Proceedings of the American Association of Artificial Intelligence. Menlo Park, CA: AAAI.
- Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge University Press.
- Schank, R. C., Fano, A., Bell, B., & Jona, M. (1994). The design of goal-based scenarios. *Journal of the Learning Sciences*, 3(4), 305-345.
- Schwartz, D.L., & Bransford, J.D. (1998). A time for telling. *Cognition & Instruction*, 16(4), 475-522.
- Shah, F., Evens, M.W., Michael, J., & Rovick, A. (2002). Classifying student initiatives and tutor responses in human keyboard-to keyboard tutoring sessions. *Discourse Processes*, 33, 23-52.
- Shneyderman, A. (2001). *Evaluation of the Cognitive Tutor Algebra I Program*. Miami, FL: Miami-Dade County Public Schools Office of Evaluation and Research.
- Sinclair, J. & Coulthart, M. (1975) *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.
- Slavin, R.E. (1990). *Cooperative learning: Theory, research, and practice*. New Jersey: Prentice Hall.
- Sleeman D. & J. S. Brown. (1982)(Eds.). *Intelligent Tutoring Systems*. Orlando, Florida: Academic Press, Inc.
- Stein, N. L., & Hernandez, M.W. (2007). Assessing understanding and appraisals during emotional experience: The development and use of the Narcoder. In J. A. Coan & J. J. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 298-317). New York: Oxford University Press.
- Suraweera, P., & Mitrovic, A. (2004). An intelligent tutoring system for entity relationship modelling. *International Journal of Artificial Intelligence in Education*, 14(3,4), 375-417.

- Taraban, R., Rynearson, K., & Stalcup, K. (2001). Time as a variable in learning on the World Wide Web. *Behavior Research Methods, Instruments, & Computers*, 33, 217-225.
- Topping, K. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- VanLehn, K. (2006) The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*. 16, 3, 227-265.
- 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.
- VanLehn, K., Jordan, P., Rosé, C. P., et al. (2002). The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems: 6th International Conference* (pp. 158-167). Berlin: Springer.
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W.B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209-249.
- Vygotsky, L.S. 1978. *Mind in Society*. Cambridge, MA: Harvard University Press.
- Watson, I., & Marir, F. (1994). Case-based reasoning: A review. *Knowledge Engineering Review*, 9(4), 327-354.
- 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: Erlbaum.
- 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: Erlbaum.

Author Notes

The research on was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918), the Institute of Education Sciences (R305H050169, R305B070349), and the Department of Defense Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, or DoD. The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of researchers from psychology, computer science, physics, and education (visit <http://www.autotutor.org>, <http://emotion.autotutor.org>, <http://fedex.memphis.edu/iis/>). Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230, [a-graesser@memphis.edu](mailto:agraesser@memphis.edu).