INSTRUCTION BASED ON TUTORING

Arthur C. Graesser, Vasile Rus, and Xiangen Hu

INTRODUCTION

This chapter reviews research on human tutoring and also some technologies that enhance the tutoring experience. The previous chapter on tutoring in this Handbook series (Graesser, D’Mello, & Cade 2011) identified a number of tutoring strategies that human tutors routinely enact, as well as some ideal strategies they could use if they had the capacity to implement them. Some instructional strategies are so complex to apply that it takes an intelligent computer tutoring system to track what the student knows and to adaptively help the student by generating well-selected tutoring moves. Indeed, there is a growing field of intelligent tutoring systems (ITS) that has shown promise and is continuing its momentum in scaling up to accommodate millions of learners over the developmental life-span (Graesser, Conley, & Olney, 2012; Sottilare, Graesser, Hu, & Goldberg, 2014; Woolf, 2009). Moreover, both human tutors and computer tutors have shown impressive learning gains compared to classroom teaching and other comparison conditions, as we will document in this chapter. The focus of this chapter is on the instructional strategies and mechanisms that underlie tutoring, whether the tutor is another human or a computer.

Tutoring is a form of one-on-one instruction between a tutor (human or computer) and a tutee (student). The human tutors vary in experience and expertise. Peer tutors are involved when one student is assigned the role of tutor and the other the tutee (Johnson & Johnson, 1992; Mathes & Fuchs, 1994; Slavin, 1990; Topping, 1996). A peer tutor is similar to the tutee in age, status, and background so there is high common ground at the sociocultural level even though the tutor has limited subject matter expertise and no pedagogical training. Research tutors include professors, staff, graduate students, or other relatively knowledgeable domain experts (in some cases, members of the research team) who are recruited to play the role of a tutor for the purposes of a laboratory experiment (Chi, Siler, Yamauchi, Jeong, & Hausmann, 2001; Evens & Michael, 2005; VanLehn et al., 2007). The subject matter expertise is comparatively high, whereas pedagogical expertise varies and the sociocultural common ground with the tutee is limited. The most frequent category of tutors in the real world
includes *untrained paraprofessionals, cross-aged student tutors* (older than the tutee), and adults who volunteer assistance to students in school programs (Cohen, Kulik, & Kulik, 1982). The subject matter expertise, pedagogical training, and sociocultural common ground with the tutee vary considerably in this category of tutor. The ideal tutor category presumably consists of *experienced professional tutors* (Cade, Copeland, Person, & D’Mello, 2008; Kulik & Fletcher, 2015; Morrison & Rus, 2014) in which both subject matter expertise and pedagogical expertise is high.

It is interesting to point out that these different categories of human tutors have been investigated over the last three decades, but there has been no systematic comparison of the different categories of tutors with respect to either tutoring strategies or outcome measures on learning gains. In contrast, there have been comparisons between different ITS environments (Ramachandran & Atkinson, 2008; Sabo, Atkinson, Barrus, Joseph, & Perez, 2013) and also between ITS environments and human tutors (Kulik & Fletcher, 2015; Olney et al., 2012; VanLehn, 2011; VanLehn et al., 2007).

Communication plays a central role in human tutoring sessions in addition to pedagogy (Clark, & Krych, 2004; Evens & Michael, 2005; Fox, 1993; Graesser, Keshtkar, & Li, 2014; Graesser, Person, & Magliano, 1995). A very important communication channel is the natural language dialogue, but there are also other important forms of communication, such as pointing, gestures, facial expressions, physical actions, and the learning materials (e.g., textbook, problems to solve, computer learning environment). The available research on tutoring suggests that the facial expressions and intonation of the tutor’s feedback is sometimes more informative than the tutor’s verbal contributions. These various channels of communication are coordinated as the tutor and tutee take turns performing verbal, paralinguistic, or physical actions during the tutorial exchange. Most human tutors set the agenda with a fairly tight control over the tutorial session by asking questions and introducing new topics (Chi et al., 2001; Graesser & Person, 1994; Graesser et al., 1995). The tutors do give the tutees an opportunity to identify their problem areas, which often are reflected in poor exam scores or assignments the tutees are struggling with. Most tutors also give the tutees opportunities to solve the problems or answer questions, as opposed to merely lecturing or immediately providing the solutions or answers.

As the tutor and tutee work on problems or difficult questions, the tutors fill in gaps in a good solution/answer and also correct errors. The typical communication exchange varies from approximately 10 to 100 turns during the solving of a single problem or an answer to a difficult question. Communication processes are central to tutoring so researchers have devoted considerable effort in identifying the communication strategies and moves during tutoring. Unfortunately, very few studies of human tutoring have systematically assessed the impact of specific tutoring strategies on learning outcome measures (see Chi et al., 2001). In contrast, ITS researchers have had more success in investigating such relationships because it is possible for a computer to have precise control over its dialogue moves when tutees experience particular psychological states.

Technology has played a greater role in tutoring activities in recent years. Human tutors are turning to *blended learning environments* in which the tutee spends some time with a human tutor or instructor and some time with some form of computer based training. Many types of computer environments are available, such as computer assisted instruction, repetitive skill training, hypertext and hypermedia, simulations, serious games, intelligent tutoring systems, virtual reality, and massively open on-line
courses (MOOCs). The human tutor can scaffold the tutee to use these systems effectively, can step in to cover difficult conceptual ideas, and can address motivational challenges in some tutees. Human tutoring can even be embedded in computer learning environments for access when the student experiences an impasse while using the system. For example, www.tutor.com has 3500 individuals available for chat interactions when students have trouble with computer environments in mathematics or science topics. There have been several million of these embedded tutoring chats that are currently being analysed with the long term goal of automating the exchanges (Nye, Morrison, & Samei, 2015; Rus, Maharjan, & Banjade, 2015). Similarly, the Reasoning Mind mathematics ITS (www.reasoningmind.com) has an embedded Genie that launches a human tutor to interact in chat when the student gets stuck. Peer chat interactions are routinely included in MOOCs and other computer environments as social media to enhance the learning experience (Siemens, Gasevic, & Dawson, 2015). ITS environments are being developed as either enhancements or replacements of human tutors, although most colleagues believe some form of human interaction is needed for a successful application of an ITS. More generally, the eLearning enterprise is currently exploring how much human intervention and scaffolding is needed to provide a sufficient context for students to effectively use and continue to use computer learning environments (Means, Peters, & Zheng, 2014).

The remainder of this chapter has three sections. The next section discusses research that has assessed learning gains from tutoring. This includes both human tutoring and computer tutoring. The subsequent section explores why and how tutoring is effective. This section includes speculations on how tutoring may be improved. The final section presents some recommendations for future research and development.

LEARNING GAINS FROM TUTORING

Students, parents, teachers, principals, and district leaders often turn to tutors when the students are not achieving expected grades and educational standards. Wealthier families pay hundreds of dollars per hour for an accomplished tutor to rescue a son or daughter who is having trouble in school. In recent years, tutoring has become especially popular in Asia. For example, Bray and Lykins (2012) report that nearly 90% of elementary students in Korea receive some form of private tuition, as do about 85% of high school students in Hong Kong.

Tutoring is a wise choice, given the results of studies that assess the effectiveness of tutoring on learning gains. A frequent way of assessing learning is to compute an effect size by computing Cohen’s d (Cohen, 1992), the difference in the mean learning scores of the tutoring condition (M_t) and a comparison condition (M_c) and dividing the difference by a pooled standard deviation within conditions (called a sigma, or σ); alternative indices include Hedges’ g and Glass’s Δ. According to Cohen (1992), effect sizes of 0.20, 0.50, and 0.80 are considered small, medium, and large, respectively.

Human Tutoring

An early meta-analysis of human tutoring reported learning gains of approximately 0.4 sigma when compared to classroom and other suitable controls (Cohen, Kulik & Kulik, 1982). The meta-analysis included different subject matters, tutoring techniques, and tutors who varied in experience and training. The tutors in the meta-analysis tended to have a small-to-modest amount of training in pedagogy and experience in tutoring,
such as paraprofessionals and cross-age tutors, but there were also research tutors and expert tutors in the mix. This classic meta-analysis has served as an initial approximation of tutoring effectiveness and provided very encouraging news.

There have been assessments of some specific categories of tutors. Available evidence suggests that collaborative peer tutoring shows an effect size advantage of 0.2 to 0.9 (Johnson & Johnson, 1992; Mathes & Fuchs, 1994; Topping, 1996). Peer tutoring is a low-cost practical solution because trained tutors and accomplished expert tutors are expensive and hard to find. The trained research tutors have yielded effect sizes of approximately 0.8, according to VanLehn (2011), who has compared trained human tutors with intelligent tutoring systems. Interestingly, his analysis shows that the intelligent tutoring systems in his sample showed approximately the same effect sizes as the trained human tutors. Olney et al. (2012) also reported nearly equivalent effect sizes (sigma = 1.9) for trained human tutors and his GuruTutor, an intelligent tutoring system in biology for middle school students.

There have been few studies on learning gains from expert tutors because they are expensive, they are difficult to recruit for research projects, and tutors tend to stay in the tutoring profession for a short amount of time (Person, Lehman, & Ozbun, 2007). It does appear that certified tutors and well-trained tutors yield impressive gains in learning, with effect sizes varying from 0.8 to 2.9 (Bloom, 1984; Chi, Roy, & Hausmann, 2008; Roscoe & Chi, 2007; Slavin, Karweit, & Madden, 1989; VanLehn et al., 2007). In summary, the news is good that human tutoring helps students learn better than classroom teaching and other comparison conditions, such as reading a textbook for an equivalent amount of time as being tutored (VanLehn et al., 2007). However, the question is still unsettled on the specific impact of tutoring expertise on learning gains.

It is reasonable to expect that learning gains from tutoring depend to some extent on the subject matter and the characteristics of the learner (Koedinger, Corbett, & Perfetti, 2012). Consider subject matter. Most students are engaged in processing text and numbers on a daily basis, adding up to thousands of hours per year, so it may be difficult to show improvements from a tutoring intervention that lasts 50 to 100 hours over a few months. In contrast, if the students start from scratch on a comparatively esoteric topic in science or engineering, a 50–100 hour intervention is more likely to be effective. These expectations are plausible, but available tutoring research has not adequately documented systematic differences among subject matters in human tutoring. Similarly, more research on tutoring is needed in documenting aptitude-treatment interactions. It is plausible to assume that low knowledge tutees will improve more than high knowledge tutees because the former have more room to improve. However, one could also imagine that the “malleable middle” students will improve more than the extremes of (a) the students who are very low in knowledge for whom the material is beyond their zone of proximal development and (b) the students who are very high in knowledge with skills of self-regulated learning that eclipse any intervention. One could also imagine the scenario where the rich get richer. Answers to questions about aptitude-treatment interactions and differences among subject matters await future research.

There are examples of excellent, evidence-based tutoring programs in the area of reading. We have reviewed the tutoring interventions that tap deeper levels of comprehension, as opposed to those that target basic reading skills that include word decoding and vocabulary. Several comprehension literacy programs have shown learning gains, but these are normally implemented in classroom or small group settings rather than
one-on-one human tutoring; such examples include Questioning the Author (Beck, McKeown, Hamilton, & Kucan, 1997), Self-Explanation Reading Training (McNamara, 2004), and PAGES (Lovett, Lacerenza, De Palma, & Frijters, 2012). One notable example with one-on-one human comprehension training is Reciprocal Teaching (Hacker & Graesser, 2007; Palincsar 2013; Palincsar & Brown, 1984). The tutoring protocol calls for the tutor to engage students in a dialogue in which they construct the meaning of the text together. 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. That is, the tutors are encouraged to systematically change their style of tutoring as the lessons evolve, a type of meta-strategy (Morrison & Rus, 2014).

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 metacognitive 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 study of teaching (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 conducted with students from age 7 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 0.9. When standardized measures were used to assess comprehension, the median effect size favoring Reciprocal Teaching was 0.3. 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. According to the What Works Clearinghouse web page (What Works Clearinghouse, 2010) the extent of evidence for the impact of reciprocal teaching on adolescent learners was found to be medium to large.

**Intelligent Tutoring Systems**

Intelligent tutoring systems (ITS) have shown learning gains in addition to human tutoring. Successful ITS have helped student learning for well-formed topics, such as algebra, geometry, programming languages, engineering, and some sciences. Noteworthy examples in mathematics are the Cognitive Tutors (Aleven, McLaren, Sewall, & Koeding, 2009; Anderson, Corbett, Koe ding, & Pelletier, 1995; Koedinger, Anderson, Hadley, & Mark, 1997; Ritter, Anderson, Koedinger, & Corbett, 2007), ASSISTments (N. Heffernan & C. Heffernan, 2014), and ALEKS (Craig et al., 2012; Doignon & Falmagne, 1999; Hu et al., 2012) because these have scaled up for use in thousands of schools. In the area of technology, some empirically validated systems have covered electronics (BEETLE: Dzikovska, Steinhauser, Farrow, Moore, & Campbell, 2014: SHERLOCK: Lesgold, Lajoie, Bunzo, & Eg gan, 1992), digital literacy (Digital Tutor:
Kulik & Fletcher, 2015), and information technology (KERMIT: Mitrovic, Martin, & Suraweera, 2007). In the area of physics, VanLehn and his colleagues have developed Andes, Atlas, and Why/Atlas (VanLehn, 2006, 2011; VanLehn et al., 2007). All of these systems have shown learning gains (as considered in the ensuing discussion of relevant meta-analyses).

There are also success cases in ITS environments that handle knowledge domains that have a stronger verbal foundation as opposed to mathematics and precise analytical reasoning. AutoTutor (Graesser, 2016; Graesser, Keshkar, & Li, 2014; Graesser, Li, & Forsyth, 2014; Graesser, Lu et al., 2004; Halpern et al., 2012; Kopp, Britt, Millis, & Graesser, 2012; Millis et al., 2011; Nye, Graesser, & Hu, 2014; VanLehn et al., 2007) helps college students learn about computer literacy, physics, comprehension strategies, and scientific reasoning by holding conversations in natural language. The agent is a talking head that speaks, points, gestures, and exhibits facial expressions. Pedagogical agents have indeed become increasingly popular in contemporary adaptive learning environments. Other examples of agents that have successfully improved student learning are DeepTutor (Rus, D’Mello, Graesser, & Hu, 2013), GuruTutor (Olney et al., 2012), MetaTutor (Lintean et al., 2012), Betty’s Brain (Biswas, Jeong, Kinnebruw, Sulcer, & Roscoe, 2010), Coach Mike (Lane, Noren, Auerback, Birth, & Swartout, 2011), iDRIVE (Gholson et al., 2009), iSTART (Jackson & McNamara, 2013; McNamara, O’Reilly, Best, & Ozuru, 2006), Crystal Island (Rowe, Shores, Mott, & Lester, 2011), Operation ARA (Halpern et al., 2012; Millis et al., 2011), and My Science Tutor (Ward et al., 2013). These systems have covered topics in STEM (physics, biology, computer literacy), reading comprehension, scientific reasoning, self-regulated learning, and other domains and skills. These systems automatically analyze language and discourse by incorporating recent advances in computational linguistics (Jurafsky & Martin, 2008; McCarthy & Boonthum-Denecke, 2012) and statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007).

The intelligent tutoring systems just described have all been assessed on learning gains and have shown improvements in learning. However, it is important to clarify what the comparisons are in these assessments. The weakest evidence for an ITS or a human tutoring intervention simply compares a posttest learning score with a pretest learning score in a single intervention study. Many of the assessments of ITS report such comparisons even though there are limitations with the pretest comparison. It is difficult to identify what it is that explains the learning gains. For example, it could be enhanced motivation from being given the intervention, or an increased sensitivity to how one is being tested by virtue of the pretest, or even maturation over time. One defense of the pretest comparison, however, is that learning gains are essentially zero when comparing a pretest score to a posttest score when there is no training in between (Graesser, Lu et al., 2004; VanLehn et al., 2007); this result rules out maturation as a third variable. Nevertheless, a stronger comparison is to have students randomly assigned to either the intervention or a comparison condition because such a design rules out the alternative explanations. The strongest comparison condition would have similar content and amount of study time as the intervention in an attempt to control for content and time-on-task, which was part of the assessments of AutoTutor (Graesser, Lu et al., 2004; VanLehn et al., 2007) and some of the other ITS. Unfortunately, the strong comparison conditions have not been routinely included in assessments of both ITS and human tutoring.

Meta-analyses and reviews support the claim that the computer technologies improve learning when compared to classroom teaching, reading texts, and/or other
traditional learning methods. The reported meta-analyses with such comparison conditions show positive effect sizes that vary from 0.05 (Dynarsky et al., 2007) to 1.1 (Dodds & Fletcher, 2004), but hover most often between 0.3 and 0.80 (Kulik & Fletcher, 2015; Ma et al., 2014; Fletcher, 2003; Nye et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011). Our current best meta-meta estimate from all of these meta-analyses is 0.6, somewhere between a medium and large effect, according to Cohen (1992). This performance is comparable to human tutoring, which was discussed in the previous subsection. Moreover, human tutors have not varied greatly from ITSs in direct comparisons between ITSs and trained human tutors (Hu et al., 2012; Olney et al., 2012; VanLehn, 2011; VanLehn et al., 2007).

As discussed, some subject matters would be expected to show higher effect sizes than others when comparing any ITS intervention. It is difficult to obtain high effect sizes for literacy and numeracy because these skills are ubiquitous in everyday life and habits are automatized. The Cognitive Tutor (Ritter et al., 2007) for mathematics has shown an effect size of $\sigma = 0.20$ to 0.40 in environments with minimal control over instructors. In contrast, when the student starts essentially from ground zero, such as many subject matters in science and technology, then effect sizes are expected to be more robust. As a notable example, the Digital Tutor (Fletcher & Morrison, 2012; Kulik & Fletcher, 2015) improves information technology by an effect size as high as $\sigma = 4.0$ for knowledge and $\sigma = 1.1$ for skills. Such large effect sizes would never be expected in basic literacy and numeracy.

ITS environments appear to be well suited for acquiring strategies and knowledge at deeper levels in the cognitive spectrum rather than shallow levels (Kulik & Fletcher, 2015). The deeper levels involve inferences, reasoning, problem solving, application, analysis, synthesis, evaluation, and creativity, whereas the shallow levels involve perceptual learning, classification, and memory (Anderson & Krathwohl, 2001; Bloom, 1956; Koedinger, Corbett, & Perfetti, 2012). The shallow levels can be mastered by more repetitive, conventional, computer-based training. ITS advantages from AutoTutor, for example, occur for measures that tap deeper levels requiring reasoning and inferences but not the shallow memory-based items (Graesser et al., 2004; Kulik & Fletcher, 2015; VanLehn et al., 2007).

One of the hallmarks of ITS is that they attempt to track what the student knows as well as other psychological characteristics in order to tailor the instruction to the particular student. The student model is a record of the tutee’s subject matter student knowledge, misconceptions, meta-cognitive skills, motivational states (e.g., persistence or grit), emotional states, and other psychological attributes that are relevant to tutoring (Sottilare, Graesser, Hu, & Holden, 2013; Woolf, 2009). The subject matter in the student model may simply be (a) a list of knowledge components and a measure of the extent to which the student has mastered each component based on the performance history in the ITS and (b) a list of misconceptions and errors with measures of the extent to which the student exhibits each one. These measures of the various components change dynamically over the course of the tutoring session, hopefully for the better. Interestingly, it is difficult for a human tutor to track the student model in such detail so their assumptions on what the student knows is only approximate (Chi, Siler, & Jeong, 2004; Graesser et al., 1995). The ITS presumably has an advantage over human tutors with respect to detailed student modelling. For instance, researchers have shown results with respect to the ability of ITS to automatically infer students’ knowledge from system-student interaction features (Graesser, Keshtkar, & Li, 2014; Ștefănescu, Rus, and Graesser (2014).
ITS may also show advantages in tracking other psychological attributes. The emotions that students experience during ITS learning sessions have been analysed by human judges and also by computer sensing devices that can automatically be stored in log files (Baker, D’Mello, Rodrigo, & Graesser, 2010; Calvo & D’Mello, 2010; D’Mello, 2013; D’Mello & Graesser, 2012; Graesser & D’Mello, 2012; McQuiggan & Lester, 2009). These results have shown that the most common affect states during learning are boredom, engagement/flow, confusion, frustration, delight, surprise, and anxiety. The automatic sensing devices are nearly as accurate as trained judges in detecting these affective states, based on the verbal interactions, facial expressions, body posture, and speech intonation. However, humans have difficulty identifying these affective states without training and most human tutors do not receive training on detecting human emotions. Data mining techniques have also been integrated with ITS environments to detect other psychological characteristics: gaming the system without learning, off-task behaviour, disengagement, help seeking, help abuse, wheel spinning, and grit (Aleven, McLaren, Roll, & Koedinger, 2006; Baker, Corbett, Koedinger, & Wagner, 2004; Gobert, Baker, & Wixon, 2015). These psychological states can be tracked in ITS log files, whereas it is an open question whether they can be detected by human tutors.

Given that an ITS can automatically and accurately sense and store many features in the student model, the question arises how the ITS should adaptively respond to different combinations of features. This is one of the central questions in the ITS field. However, it will take some decades of research because there are so many features and combinations of features to monitor and adaptively respond to. One might ask, for example, whether an ITS should select problems and moves that respond to detailed configurations of subject matter knowledge or whether a few categories are adequate (such as a level in a learning progression on a particular topic; Rus, D’Mello, Hu, & Graesser, 2013) or whether 2–3 overall levels of proficiency (high, medium, low) are adequate (Rus et al., 2014). The jury is still out on answers to such questions.

Similarly, there are many combinations of motivational and emotional characteristics to consider. D’Mello and Graesser (2012) have tested different versions of AutoTutor on the extent to which AutoTutor responds adaptively to student affective states. Experimental manipulations compared the original AutoTutor without any emotion tracking and emotional displays to an Affective AutoTutor version that is emotionally supportive. The supportive Affective AutoTutor had polite and encouraging positive feedback (“You’re doing extremely well.”) or gentle negative feedback after a low quality student contribution (“Not quite, but this is difficult for most students.”). When the student expressed low quality contributions, the tutor attributed the problem to the difficulty of the materials for most students rather than blaming the student being tutored.

There was also a shake-up version of Affective AutoTutor. This version tried to shake up the emotions of the student by being playfully cheeky and telling the student what emotion the student is having (“I see that you are frustrated.”). The simple substitution of this feedback dramatically changed AutoTutor’s personality. The impact of the different AutoTutor versions on learning computer literacy depended on the phase of tutoring and the student’s level of mastery. The supportive emotion-sensitive AutoTutor had either no impact (for low knowledge students) or a negative impact (for high knowledge students) on learning during early phases of the tutoring session (i.e., within the first 30 minutes of the session). During a later phase of tutoring (i.e., the next 30 minutes), the supportive AutoTutor improved learning, but only for the low knowledge students. Low-domain knowledge students also performed better
on a transfer test when they interacted with the supportive AutoTutor. The shake-up
AutoTutor had more complex patterns of results that were never fully tested because
an initial study had learning gains that were the same as the original AutoTutor even
though most adults had a positive initial impression of the shake-up AutoTutor. Per-
haps the playful shake-up tutor is motivating when boredom starts emerging for the
more confident, high-knowledge learners. In any event, more research is needed in
diverse student populations, subject matters, and learning environments before the
ITS field can adequately assess the return on investment from ITS environments that
are highly adaptive to different configurations of features in the student model.

HOW AND WHY DOES TUTORING HELP?
The previous section provided some encouraging empirical evidence that human and
computer tutoring can help students learn more than classroom experiences or read-
ing textbooks for an equivalent amount of time. This section explores how and why
tutoring is so helpful. If we can understand tutoring processes in more detail, then we
can generate and test hypotheses on how tutoring can be improved in the future.

This section has three subsections. The first subsection succinctly summarizes the
tutoring processes and frequent strategies that were identified in the previous Hand-
book chapter by Graesser, D’Mello, and Cade (2011). That chapter can be examined for
references and detailed elaborations of the claims. The second subsection interprets
some of these normal tutoring processes from a perspective that considers multiple
communication channels and emotions. There are some suggestions on how tutor-
ing can be improved even further, both by humans and computers. The third sub-
section presents an example fictitious tutoring protocol that illustrates how tutoring
would typically occur in human tutoring. We comment on the tutor and tutee dialogue
moves, followed by suggestions on how the tutor could have improved dialogue moves
to facilitate learning and motivation.

Summary of Human Tutoring Processes and Strategies
The Handbook chapter by Graesser, D’Mello, and Cade (2011) identified the typical
tutoring mechanisms and strategies that were exhibited by tutors with moderate sub-
ject matter knowledge and low-to-moderate training on pedagogical strategies. Claims
about the tutoring processes and strategies are based on dozens of publications that
included videotapes and/or transcripts of naturalistic tutoring sessions. This subsec-
tion summarizes highlights from these studies as described in the Handbook chapter.

Complex Tutoring Strategies Are Infrequent
There are several complex tutoring strategies that are highly regarded in the education
and ITS communities. When transcripts were inspected in great detail, these strategies
rarely occurred (Graesser, Person, & Magliano, 1995). As one might expect, it takes tutor-
ning training to implement these tutoring strategies, whereas the human tutors rarely had
such training. More specifically, the tutors in these studies rarely implemented bona fide
Socratic tutoring, 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.
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, in Reciprocal Teaching, the tutor and learner take turns reading and thinking aloud with the goal of lacing in 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 (Doignon & Falmagne, 1999; 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 (Gagné, 1985). Tutors who diagnose and remediate deep misconceptions are on the lookout for errors that are manifestations of more global problematic mental models (Ploetzner & VanLehn, 1997). When a deep misconception is recognized, the tutor attempts to supplant the error-ridden mental model with a correct mental model.

All of the prior strategies are very complex to implement, especially on the fly during tutoring when there are time constraints. They all require detailed student modeling and quickly formulated and adaptive dialogue moves. They require the selection of new problems or difficult questions that are carefully tuned to confront the tutee’s knowledge gaps. This is much too difficult for any human to manage, even for expert tutors. Ideally, these could be implemented in ITS, but even those systems cannot manage the computational complexity of some of these sophisticated strategies.

**Tutoring Sessions Are Organized Around Problems, Challenging Questions, and Tasks**

The outer loop of VanLehn’s (2006) analysis of tutoring consists of the selection of major tasks for the tutor and tutee to work on. The inner loop consists of the steps and dialogue interactions to manage the interaction within these major tasks. After the tutor negotiates with the tutee on the major tasks, the tutor guides the selection of major tasks and the interaction within each task. In this sense, the typical tutoring interaction is tutor-centered, not student-centered in a self-regulated manner.

**A 5-Step Tutoring Frame Guides the Major Task (Outer Loop)**

Once a problem or difficult main question is selected to work on, the 5-step tutoring frame is launched, as specified below (Graesser & Person, 1994).

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.

Step 4 in this 5-step tutoring frame involves collaborative discussion, joint action, and encouragement for the student to construct knowledge rather than merely receiving knowledge. Interestingly, steps 1, 2, and 3 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). 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 Dialogue Guides**

**Micro-Adaptation (Inner Loop)**

Human tutors typically have a list of *expectations* (anticipated good answers, steps in a procedure) and a list of anticipated *misconceptions* (and errors or bugs) associated with each main question (Graesser, Jeon, & Dufty, 2008). For example, expectation E1 and misconception M1 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.

M1: A lighter/smaller object exerts no force on a heavier/larger object.

The tutor guides the student in articulating the expectations through a number of dialogue moves: *pumps, hints,* and *prompts* for the student to fill in missing words. A pump is a generic expression to get the student to provide more information, such as “What else?” or “Tell me more.” Hints and prompts are selected by the tutor to get the student to articulate missing content words, phrases, and propositions. A hint tries to get the student to express a complex idea (e.g., proposition, clause, sentence), whereas a prompt is a question that tries to get the student to express a single word or phrase. 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 prompting 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 typically acknowledges the error and corrects it. There is another conversational goal of giving feedback to the student on their contributions (positive, negative, or neutral). The tutor also attempts to accommodate a mixed-initiative dialogue by attempting to answer the student’s questions when the student is sufficiently inquisitive to ask questions. However, most students do not ask many questions, even in tutoring environments (Graesser & Person, 1994).

**Tutor Turns Are Well Structured**

Most turns of the tutor have three informational components during the inner loop conversation of step 4 of the 5-step frame.
Tutor Turn $\rightarrow$ Short Feedback + Dialogue Advancer + Floor Shift

The first component is feedback (positive, neutral, negative) on the quality of the student’s last turn. The second component is a dialogue advancer that moves the tutoring agenda forward with either pumps, hints, prompts, assertions with correct information, corrections of misconceptions, or answers to student questions. The third component shifts the conversational floor with cues from the tutor to the student. For example, the human ends each turn with a question or a gesture to cue the student to do the talking.

*Communication between the Tutor and Student Is Limited*

There is an idealistic assumption that the tutor and tutee understand each other while they interact. In truth, they live in very different discourse worlds because their common ground of knowledge is minimal. Graesser, D’Mello and Person (2009) documented this gap in knowledge when they articulated five tutoring illusions:

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. Given the low common ground between tutor and tutee, this assumption is false.

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?”). Tutors typically give positive short feedback to erroneous or vague student contributions.

3. **Illusion of discourse alignment.** The unwarranted assumption that the listener is expected to understand the discourse function, intention, and meaning of the speaker’s dialogue contributions. 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. 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 have these illusions, which thereby compromise the effectiveness of tutoring. The illusions undermine the tutor’s ability to build an accurate and detailed student model of the student. There might be better hope for ITS technologies to more accurately track the student model. However, technology also faces limitations because a large percentage of student contributions are vague, underspecified, telegraphic, and ambiguous (Graesser & Person, 1994; Graesser et al., 2009).

Accomplished tutors are probably careful not to over-estimate a student’s level of understanding, taking care to check understanding directly. This is illustrated in the simple hypothetical 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.

Person et al. (2007) 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 in step 5 of the 5-step tutoring frame. However, there has not yet been a systematic comparison of the tutoring moves of tutors who vary in expertise.

More Research Is Needed on Expert Tutors

Very few researchers have systematically analyzed expert tutors and their work has focused on three or fewer tutors. Consequently, it is difficult to make general claims on the basis of such case studies. Cade et al. (2008) is the first study that performed detailed analyses on a reasonable sample of expert tutors. They identified higher-level dialogue states called *modes*, such as modelling, lecturing/telling, scaffolding, fading, and sense-making. Subsequent studies have investigated these modes and have also linked them with specific tutoring tactics and strategies (Morrison et al., 2015; Rus, Maharjan, & Banjade, 2015; Vail & Boyer, 2014). The central idea is that tutorial dialogues, like all human conversations, have higher-level states that tend to constrain the kinds of dialogue acts that are expected when the dialogue is in that state. Some dialogue acts are understood as attempts to switch the mode, such as a move to a closing or a move to open a new topic (Schegloff & Sacks, 1973). A tutor’s decision to switch from one mode to another is taken to reflect the tutor’s metastrategy.

Morrison et al. (2015) recently conducted a study on over 1,400 online tutoring sessions conducted by professional tutors employed by an online tutoring service www.tutor.com. Morrison et al. (2015) reported that tutoring sessions rated by experts as showing evidence of student learning and “educational soundness” included relatively more scaffolding and fading, but fewer telling and modelling episodes. An important theoretical contribution of that project is that it provided a systematic way of linking linguistic categories with pedagogical categories.

Multiple Channels of Communication and Emotions

There are multiple channels of communication in addition to language. Some tutors get lost in the abstractions of the curricula content and language to the point of becoming detached from the core communication mechanisms that involve joint attentional activity, pointing, gestures, theory of mind, intentions, and shared social norms. The detachment is likely to threaten the success of a tutoring experience and increase the illusions of tutoring that were described earlier. Attempts can be made to train human tutors to become attuned to the features associated with the multichannel perspective, but that can be tricky when they have not been trained in the pedagogical practice of some cultures. Attempts can be made to create computer systems to simulate the
multi-channel communication mechanisms in automated tutors, but there needs to be engineering and empirical assessments to see how well that can be accomplished. How well can a computer track the eye gazes, gestures, and intentions of a tutee, for example? How can a computer agent respond to the human in a convincing manner? These questions are being pursued in the ITS field.

The impact of tutoring expertise on student learning is complicated by the fact that tutoring is necessarily interactive, requiring the combined efforts of both the tutor and the learner (Chi et al., 2001; Clark & Krych, 2004; Graesser & Person, 1994). As in many communication contexts, the listener jointly helps the speaker say what the speaker wants to say (Clark, 1996). This makes it difficult to figure out who gets credit for any accomplishments that are achieved, as is the case in all collaborative problem solving and learning activities (Graesser et al., in press). It does not seem to help much for the tutor to articulate explanations, solutions, and other critical content in the form of telling without making any attempt to connect with what the learner knows (Chi et al., 2001; Graesser et al., 2009). Nevertheless, attempts have been made to discover meaningful patterns in tutors’ actions as they attempt acts of grounding in communication (Rus, Maharjan, & Banjade, 2015). They note that previous research on instructional strategies (what the instructor/tutor does) and learning strategies (what the learner does) have fundamentally different slants on what tutors do, what students do, and the dynamics of the tutor-learner interaction.

Emotions are an important dimension of the sociobiological perspective of tutoring because tutees experience a variety of emotions during learning and emotions are on the radar of infants within the first 6 months of their development. Three of the frequent learning-centered emotions are boredom, confusion, and frustration (Baker et al., 2010; D’Mello, 2013; Graesser & D’Mello, 2012) and all of these have important repercussions on learning. The relationship between emotion and learning has received increasing attention in the fields of both psychology and education, as discussed in the previous Handbook chapter. As one notable example, Lepper, Drake, and O’Donnell (1997) proposed an INSPIRE model to promote the integration of emotions and learning. 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 interesting practice 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 (Zimmerman, 2001). D’Mello and Graesser (2012) developed an AutoTutor ITS system that is sensitive to student emotions and that attempted to build self-efficacy with supportive comments. They found this approach to help learning after initial interactions with the low-knowledge learners, but not the high-knowledge learners.

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 both correlates with and has a causal impact on learning gains perhaps because it is a direct reflection of deep thinking (D’Mello, Pekrun, Lehman, & Graesser, 2014). Confusion is diagnostic of cognitive disequilibrium, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts. Cognitive equilibrium is ideally restored after thought, reflection, problem solving, and other
A. C. Graesser, V. Rus, and X. Hu

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 documented pedagogical value.

**An Example Tutoring Protocol with Comments**

This subsection presents an example fictitious tutoring protocol that reflects a typical exchange in human tutoring. The tutorial dialogue depicts what normally occurs during tutoring and serves as a spring-board for discussing more deeply what happened during the tutoring and how different tutoring moves could facilitate learning. The tutor presents the student with the main question below.

 três newspaper article reported a recent study on self-driving police cars. A sample of 1000 officers was selected to test drive a self-driving police car and then were given a survey on how much they liked the car. The survey included a 4-point liking scale: 1 = “I hate driving the car,” 2 = “I moderately dislike driving the car,” 3 = “I moderately like driving the car,” and 4 = “I love driving the car.” The mean liking ratings were 3.2, with 80% of the officers giving a rating of 3 or 4. The study convinces the police department to replace all of their cars with self-driving cars. Is this a well-designed study that supports the decision? Explain why or why not.

The tutor hopes that the student covers the following expectations.

E1: There is a problem with the design of the study.
E2: The study has no comparison condition with a car that is not self-driving.
E3: A car in a comparison might produce liking ratings that are equal or higher than the ratings of the self-driving cars.

An example of a hypothetical tutorial dialogue is presented in Table 21.1. This dialogue illustrates the 5-step dialogue frame (Graesser & Person, 1994) with a main question (turn 1), an initial student response (turns 2 and 4), tutor feedback (turns 3 and 5 and later), tutor and student collaborative dialogue (turns 5 through 15), and tutor comprehension gauging (turn 15). Following the expectation- and misconception-tailored dialogue, the tutor also provides a variety of scaffolding moves, such as pumps (turns 3 and 9), hints (turns 5 and 11), and prompts (turn 13) to encourage the student to do the talking rather than lecturing. The tutor presents some answer content also, such as explanations and corrections (turn 9) and the final summary (turn 17). The tutor turns are well structured (Tutor Turn ◊ Short Feedback + Dialogue Advancer + Floor Shift) during the collaborative interaction phase. This dialogue also illustrates why it is important to have the dialogue. The first two student turns would lead one to believe that the student has little or no ability or knowledge to answer the question. However, later in the dialogue the student exhibits relevant knowledge after the tutor corrects a student claim (turn 9) and presents a hint (turn 11). The student’s knowledge would not be detected in a typical classroom context.

The tutorial dialogue in Table 21.1 has some positive characteristics that are aligned with well-documented scientific principles of learning (Mayer, 2011; Pashler et al., 2007). For example, the tutor has scaffolding moves that encourages active student learning rather than simply lecturing. The tutor gives feedback to the student in various
forms, including short positive or negative feedback and qualitative feedback on misconceptions. The tutor asks main questions that require reasoning rather than shallow questions that invite one or two word answers. The tutor attempts to cover the correct information rather than allowing unenlightening discourse (“anything goes”). Nevertheless, the tutoring session could improve with some more judiciously selected tutoring moves, as suggested below, which incorporate the scientific principles of learning.

1. **Request student summary.** Instead of the tutor giving the summary of a good answer (turn 17), the tutor could first request that the student summarize the answer. Requesting summaries from the student promotes active generation of information, which facilitates learning, and allows the tutor to better gauge the student’s understanding. The tutor can follow the student summary with the tutor summary so the student can compare the student answer with a good answer.

2. **Don’t trust the student’s accuracy of meta-comprehension.** The tutor assumes that the student is accurate when the tutor asks “Do you understand?” (turn 15) and the student answers “I see now” (turn 16). Decades of research have shown that students are not very good at calibrating their own comprehension (Dunlosky & Lipko, 2007; Maki, 1998). The tutor needs to have dialogue moves
that troubleshoot the student’s understanding, such as a follow-up question or a request for student summary.

3. **Explore the foundations of student’s wrong answers.** The tutor immediately dismisses the student’s irrelevant answer, “Police should not have self-driving cars” (turn 4), when that was on the student’s mind. Instead of dismissing it by saying, “We are not asking for your opinion” (turn 5), the tutor could explore through dialogue what the student had in mind. That would be an opportunity for an illuminating dialogue on the differences between opinions and claims based on scientific methods, one of the foundational issues in science. Tutors can be on the lookout for opportunities to launch such dialogues on epistemological foundations in the context of the student’s mindset.

4. **Ground referring expressions and quantities.** It is not established from the dialogue whether the tutor and tutee are on the same page (i.e., common ground, Clark, 1996) on what a self-driving car is and what the 3.2 mean means. Without a common ground on these important referring expressions, a deep answer to the question will not emerge. Self-driving cars may have an unusual or ugly appearance, so that would need consideration when selecting a suitable comparison car. A student may not know whether 3.2 is above-average versus below-average and what base rate is expected. Questions from the tutor could help resolve potential misalignments on common ground, such as “What are the characteristics of a self-driving car?” or “Is 3.2 a high or low score?”

5. **Request more explanations behind student answers.** In addition to accepting a student’s short answers (such as in turn 14), the tutor could ask that the student justify the reasoning with an explanation by posing a deep-reasoning question, such as why, how, or why not? Thus, in turn 14 the tutor could ask, “Why is a comparison condition needed?” Explanation-based questions improve comprehension and learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi et al., 2001; Gholson et al., 2009; Roscoe & Chi, 2007).

6. **Plant seeds for cognitive disequilibrium.** The tutor could challenge the student by disagreeing with the student, presenting contradictions or facts that clash with common knowledge. After the student expresses “The study has a problem” (turn 6), the tutor could disagree by saying, “Many students don’t see a problem with this study. Why do you see a problem?” These events create cognitive disequilibrium and temporary confusion, but such experiences may promote deeper learning (D’Mello, Lehman, Pekrun, & Graesser, 2014; Lehman et al., 2013).

7. **Monitor student emotions.** The early exchanges of the tutorial dialogue suggested that the student was disengaged (perhaps bored) from the tutorial experience by expressing “I don’t know” (turn 2) or silence (turn 10). Tutors need to monitor and adequately respond to student affect in addition to their knowledge states (D’Mello & Graesser, 2012; Lepper & Woolverton, 2002). Available research suggests that the three most frequent affective states in tutoring and other learning environments to promote deep learning are confusion, frustration, and boredom/disengagement (Baker et al., 2010; D’Mello, 2013). Confusion and frustration may be handled by tutor dialogue moves that are relevant to the subject matter, but boredom may require tutoring moves that are related to the self-concept of the student. That requires an analysis of how the subject matter relates to the student’s life and identity.
DIRECTIONS FOR FUTURE RESEARCH

This chapter has covered progress that has been made in human and computer tutoring. We have presented evidence that tutoring is a very effective method of training and education beyond reading textbooks and listening to lectures. The evidence has accumulated from both correlational studies and true experiments that randomly assign tutees to experimental versus comparison conditions. Tutoring has been around for millennia and is firmly planted in our sociocultural mechanisms. Intelligent computer tutors can emulate many of the human tutoring moves and strategies, plus go a step beyond with ideal tutoring strategies. Indeed, contemporary intelligent tutoring systems show learning gains approximately equivalent to trained human tutors. However, some of the pedagogically complex tutoring strategies are beyond the capabilities of both humans and computers.

Successful is not limited to the cognitive realm. It is important for tutors (human or computer) to track the emotions and motivational states of tutees in addition to their cognitive and subject matter knowledge states. It is very important for tutors to be judiciously integrated in the curriculum and educational context. It is important for the tutor to understand the social and cultural dimensions of the tutoring experience.

There are some noticeable gaps in the tutoring literature that may stimulate future empirical research and scale up efforts. One could imagine dozens of future efforts, but the following five efforts appear to be particularly pressing in the immediate horizon.

1. There is a lack of experimental comparisons between different human tutoring interventions on the impact on student learning, motivation, and emotions. Without such comparisons, it is difficult to conclude which human tutoring strategies and practices are worthwhile to pursue. It is of course important to have such evaluations cover different subject matters and student populations.

2. There is considerable uncertainty on how human teachers and tutors are needed to contextualize and scaffold the use of the intelligent tutoring systems. What do teachers or tutors need to do in their curriculum or instructions to convey the value of using an ITS? How much scaffolding is needed? When does the scaffolding interfere with a student’s self-regulated learning with an ITS?

3. There needs to be more solid evidence that an adaptive ITS produces learning gains over and above a learning environment that presents the same content in some form that is not adaptive to the individual learner. What if a well-scripted rigid intervention ends up being equivalent to an ITS that is sensitively adaptive to the learners psychological states (subject matter knowledge, cognition, motivation, emotions)? Such comparisons are needed in the ITS community with respect to true experiments.

4. Teachers and tutors have little or no knowledge of the advances in human and computer tutoring research. It is not part of the standard professional development of teachers. How can tutoring wisdom be propagated to teachers, tutors, and other practitioners in the educational system? How can their views of high quality tutoring practice influence research on human tutoring and ITS?

5. There are no widely established standards or guidelines on tutoring mechanisms and strategies. Without such standards, there is no foundation for assessing tutoring effectiveness. Some tutors hinder learning, whereas others help. Without any standards, it will be impossible to evaluate tutors on quality.
One wonders what the ideal tutor will be in the future. Will it be a computer tutor that combines the wisdom of human tutors with the algorithmic precision of intelligent systems? Will it be a human tutor who directs the student to relevant intelligent tutoring systems that help the student’s life trajectory? Will it be an agent that serves as a personal assistant for the individual student who directs the student over years? These are questions for the next generation of researchers.

NOTE

1 The research was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918, ALT-0834847, DRK-12–0918409, 1108845), the Institute of Education Sciences (R305H050169, R305B070349, R305A080589, R305A080594, R305G020018, R305C120001), Army Research Lab (W911NF-12–2-0030), and the Office of Naval Research (N00014–00–1–0600, N00014–12–C-0643). We would like to thank Chip Morrison for making comments on the manuscript. 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 composed of researchers from psychology, computer science, physics, and education at University of Memphis (visit http://www.autotutor.org, http://emotion.autotutor.org, http://fedex.memphis.edu/iis/).

REFERENCES


