Computer Agents that Help Students Learn with Intelligent Strategies and Emotional Sensitivity

Arthur C. Graesser, Sidney D'Mello, and Amber Strain

University of Memphis

:

Art Graesser
Psychology Department
202 Psychology Building
400 Innovation Drive
University of Memphis
Memphis, TN, 38152-3230
901-678-4857
901-678-2579 (fax)
a-graesser@memphis.edu

KEYWORDS: Agents, tutoring, discourse, learning technologies, emotions

Abstract

Computer agents have been designed to help students learn subject matters by holding conversations with the students in natural language. For example, AutoTutor improves learning of subject matters such as computer literacy and conceptual physics by co-constructing explanations and answers to complex questions (why, how, what if, etc.). One version of AutoTutor is sensitive to the affective states of the learners in addition to their cognitive states and also responds with emotions designed to facilitate learning. These computer agents simulate the cognitive and metacognitive strategies of human tutors in addition to incorporating ideal strategies. The agents are not perfect conversation partners and comprehenders of language, but the conversations are surprisingly coherent and also help students learn.

For millennia, prior to the industrial revolution, the most common way for students to learn a skill or subject matter was to hold conversations with a mentor, master, tutor, or instructor in an apprenticeship environment (Collins & Halverson, 2008; Graesser, D'Mello, & Cade, 2012; Resnick, 2011). The student and pedagogical expert would collaboratively work on tasks and problems as the student would hopefully achieve new levels of mastery and practice the crafts. The expert would attend to the emotions of the student in addition to the student's behavior and apparent cognitive states.

We have now reached the point where conversational agents on computers can be effective substitutes for the human pedagogical experts. This article will describe some computer systems that effectively serve as virtual pedagogical experts. Skeptics often grumble that a computer could never understand a student as deeply as a human tutor, let alone respond in an intelligent manner. However, a systematic analysis of the process of human tutoring has revealed that the vast majority of tutors do not manage to deeply understand what students know and do not implement sophisticated strategies to help them learn (Graesser, Person, & Magliano, 1995; Graesser, D'Mello, & Cade, 2012). Tutors rarely implement highly regarded 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. Simply put, human tutors are remarkably unremarkable. They try to be polite, helpful, and supportive conversation partners. But they are rarely capable of diagnosing the student's deep misconceptions, repairing subtle errors, and eliciting from the student accurate complete solutions to problems. In spite of the scruffy ways of human tutors, they manage to be very effective in helping students learn, indeed more helpful than most alternative learning environments. For example, learning gains are approximately 0.4 sigma for typical unskilled tutors

in the school systems, when compared to classroom controls and other suitable controls (Cohen, Kulik, & Kulik, 1982), and vary from .2 to 2.0 for accomplished human tutors (Chi, Roy, & Hausmann, 2008; VanLehn et al., 2007). Collaborative peer tutoring even shows an effect size advantage of 0.2 to 0.9 sigma (Johnson & Johnson, 1992; Mathes & Fuchs, 1994; Topping, 1996).

What do Human Tutor's Do?

Given that human tutors are effective, what is it they do to help students learn? The tutors do lecture on mini-topics periodically, hopefully just in time to help the student. However, the more important work is organized around difficult questions and problems that require reasoning and explanations in the answers. The following is an example of a challenging question on the topic of Newtonian physics.

PHYSICS QUESTION: If a lightweight car and a massive truck have a head-on collision, which vehicle undergoes the greater change in its motion, and why?

This why question requires 3-5 sentences in an ideal answer, but students ask an average of 1.2 sentences and rarely more than 2 sentences when initially asked such a deep question. A conversation takes typically 20 to 100 turns to draw out more of what the student knows and to answer the question collaboratively.

Tutors have a number of *dialogue moves* when they construct a conversational turn and manage the collaborative dialogue in a fashion that encourages more student contributions. The major categories of dialogue moves are listed below.

(1) *Short Feedback* on the quality of the contribution in the student's previous turn, such as positive ("very good"), neutral ("okay"), versus negative ("not quite").

(2) *Pumps* encourage the student to express more information ("What else?").

(3) *Hints* guide the student to express sentence-length ideas that are important answers to the main question/problem. For example, the hint "What about the forces of the vehicles on each other?" attempts to get the student to express "The forces exerted by each vehicle on each other are equal in magnitude."

(3) *Prompts* guide the student to fill in a missing word in an important idea. To get the student express the word "magnitude," for example, AutoTutor would deliver the prompt "The forces of the two vehicles on each other are equal in *what*?"

(4) *Assertions* of AutoTutor articulate important ideas in the answer or problem, e.g., "The forces of the two vehicles on each other are equal in magnitude."

(5) *Corrections* correct erroneous ideas and misconceptions. After the student expresses the misconception "The smaller vehicle exerts less force on the larger vehicle" then AutoTutor corrects the student with the assertion in #4.

(6) *Answers* are provided when the students ask some types of questions, such as definitional questions, e.g., "What does acceleration mean?" However, students do not frequently ask questions in both human and computer tutoring sessions because the tutor is prone to drive the agenda.

(7) *Summaries* provide the complete answer to the main question/problem.

Most of the tutor's conversational turns include 2 or more of these dialogue moves. For example, after a student expresses a misconception, a tutor would have a conversational turn that generates short negative feedback, a correction, and then a hint, as illustrated below.

STUDENT: The smaller vehicle exerts less force on the larger vehicle.

TUTOR: No, the forces of the two vehicles on each other are equal in magnitude. What about the velocity of the two vehicles?

The 3-7 sentences in a full answer to the main question are eventually constructed. These sentences are parts of an explanation that capture important principles of the subject matter.

A good tutor does not merely lecture, but rather tries to get the student to express the answer because the active generation of an explanation is better than passive learning. The tutor tries to get the student to express parts of an anticipated good answer (called *expectations*) and corrects any *misconceptions* expressed by the student. This is what we call *expectation plus misconception tailored* (EMT) dialogue. In order to perform EMT dialogue effectively, the tutor presumably needs to build an accurate model of what the student knows (called *student modeling*) and also to strategically generate dialogue moves to get the student to fill in relevant information (called *strategic elicitation*). Student modeling is optimized to the extent that there is accurate *pattern matching* between the student contributions and each of the expectations and misconceptions. Strategic elicitation is optimized to the extent that the tutor's dialogue moves end up maximizing the amount of information that the student provides when achieving *pattern completion*.

The quality of student modeling and strategic elicitation in human tutoring is far from optimal. One reason is there typically is a very large gulf between what the tutor knows and the student knows. Shared knowledge and common ground (Clark, 1996) are difficult to achieve when the conceptualizations of tutor and student are so different. A second reason is the difficulty of pattern matching because natural language tends to be imprecise, fragmentary, vague, and ungrammatical. A third reason is that human tutors have a number of tutoring illusions that get in the way of optimizing student modeling and strategic elicitation. Graesser, D'Mello and Person (2009) documented the following five illusions.

(1) *Illusion of grounding.* The unwarranted assumption that the tutor and student have shared knowledge about a word, referent, or idea being discussed in the tutoring session. A good tutor is skeptical of the student's level of understanding so 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 student and tutor give each other is accurate. For example, tutors incorrectly believe the students' answers to their comprehension gauging questions (e.g., "Do you understand?").   It is the more knowledgeable students who tend to answer they do not understand. On the flip side, sometimes tutors are polite or encouraging so they do not give the student accurate feedback after the student gives low quality information.

(3) *Illusion of discourse alignment.* The unwarranted assumption that the student understands the discourse function, intention, and meaning of the tutor'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. The fact that a student expresses a single word or phrase does not mean that the student understands a complex idea.

(5) *Illusion of knowledge transfer.* The tutor's unwarranted assumption that the student understands whatever the tutor says and thereby knowledge is accurately transferred. Much of what the tutor expresses is not understood by the student so knowledge transfer is modest.

In summary, human tutors are far from perfect in performing student modelling and strategic elicitation. They also rarely implement the sophisticated tutoring techniques that are

extolled in the education and intelligent tutoring systems communities (Graesser, Person, & Magliano, 1995; Graesser, D'Mello & Person, 2009), such as Socratic questioning, modeling-scaffolding-fading, frontier learning, building on prerequisites, or diagnosis/remediation of deep misconceptions. These observations opened the door to the possibility of programming a computer to simulate the EMT dialogue process that is ubiquitously exhibited by human tutors. It might also be possible to move beyond what humans can do by performing more accurate student modeling and more intelligent strategic elicitation. AutoTutor was designed to achieve to simulate human tutoring and to implement more ideal tutoring mechanisms.

AutoTutor

AutoTutor (Graesser, Jeon, & Dufty, 2008; Graesser, Lu et al., 2004) was the first ITS with conversational agents developed by researchers in the Institute for Intelligent Systems at the University of Memphis. Students learned about topics in science and technology by holding conversations in natural language. There was an explicit attempt to simulate human tutorial dialogue in the design of AutoTutor. However, some versions of AutoTutor attempted to go beyond normal tutors by enhancing the accuracy of student modelling and optimizing elicitation of student contributions. These goals are somewhat different from other learning environments with conversational agents that have directly incorporated ideal learning principles, such as *ITSPOKE* (Litman et al., 2006), *Tactical Language and Culture Training System* (Johnson & Valente, 2008), *Why-Atlas* (VanLehn et al., 2007), *Operation ARIES!* (Millis et al., in press), and *iSTART* (McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). AutoTutor and these other conversation-based learning environments have collectively covered a variety of subject matters and skills, such as computer literacy, electronics, physics, circulatory systems, critical thinking about science, foreign language, cultural practices, and reading strategies.

Student contributions rarely match the expectations perfectly because natural language tends to be imprecise, fragmentary, vague, and ungrammatical. AutoTutor implements semantic match algorithms that can accommodate the scruffiness of natural language (Graesser, Penumatsa, et al., 2007; Rus, McCarthy, McNamara, & Graesser, 2008). These semantic match algorithms are computed on individual student turns, combinations of turns, or the cumulative sequence of turns that lead up to a particular point in the dialogue.

How does AutoTutor handle or improve student modelling compared with human tutors? As each student contribution is expressed over the turns, we keep track of the extent to which the content of student contribution C overlaps the meaning of each expectation E_i (and misconception M_j), with match scores that vary from 0 to 1. These matching operations are based on a combination of syntactic information (such as the order of matching content words) and semantic information, such as the similarity of content words weighted by word frequency, latent semantic analysis (Landauer et al., 2007), and symbolic interpretation algorithms (Rus & Graesser, 2006) that are beyond the scope of this article to address. If there are 20 student turns in a conversation and 5 expectations, then there would be 20x10 = 100 pattern match scores computed as the information accrues turn by turn. An expectation is considered covered by the student when the match score meets or exceeds some threshold T. The conversation finishes when all 5 expectations are covered with above threshold match scores. AutoTutor selects the next expectation to work by identifying the expectation with the highest match score, given it is not already covered (i.e., exceeding the threshold T). In this fashion, AutoTutor builds on what the student knows, a form of frontier learning or zone of proximal development (Vygotsky, 1978). Thus, the student model in AutoTutor at any one moment in time for problem P, is the vector of match scores for the set of expectations and misconceptions (E_i and M_j).

AutoTutor generates dialogue moves to fill in missing content and achieve *pattern completion*. More specifically AutoTutor periodically identifies a missing expectation during the course of the dialogue and posts the goal of covering the expectation (E_i). When a particular expectation is posted, AutoTutor tries to get the student to express it by generating hints and prompts that encourage the student to fill in missing ideas and words.

How does AutoTutor handle or improve strategic extraction of information from the student? AutoTutor selects particular prompts and hints that elicit answers that would optimize the likelihood of filling in the missing information and thereby boosts the match score above threshold. For example, suppose that the expectation (*The magnitudes of the forces exerted by two objects on each other are equal*) needs to be articulated in the answer. AutoTutor would start out by selecting the one hint, from the set of hints associated with that expectation, that would increment the match score the most if the hint were answered correctly. Stated differently, a correct answer to the hint would maximally increase the coverage of the expectation. However, hints might work out so AutoTutor then judiciously selects one or more prompts to get the student to articulate particular words. For example, the following family of candidate prompts is available for selection by AutoTutor to encourage the student to articulate particular content words in the expectation.

(a) The magnitudes of the forces exerted by two objects on each other are ____.

(b) The magnitudes of forces are equal for the two _____.

(c) The two vehicles exert on each other an equal magnitude of _____.

(d) The force of the two vehicles on each other are equal in _____.

If the student has failed to articulate one of the four content words (*equal, objects, force, magnitude*), then AutoTutor selects the corresponding prompt (a, b, c, and d, respectively). Or more generally, the prompt is selected if a correct completion optimally increases the coverage

of that expectation. If the student fails to articulate the expectation above threshold T after a series of hints and prompts, then AutoTutor resorts to asserting the expectation and moving on.

It follows from these computational procedures that there is a progressively more directed line of strategic extraction as AutoTutor tries to get the student to do the talking. AutoTutor starts out pumping at a general level (Tell me more, what else) and then selects a particular expectation to work on. AutoTutor then implements a [hint → prompt → assertion] cycle for each expectation until the expectation is covered (and immediately exiting from the cycle when it is covered). In this fashion, the selection of AutoTutor's dialogue moves is sensitive to the cognitive states of the learner. For example, students who have more knowledge and verbal abilities provide most of the information in the answer, so AutoTutor generates primarily pumps and hints. In contrast, students with low knowledge and/or verbal abilities need more prompts and assertions from AutoTutor (Graesser et al., 2007). There is a continuum from the student to the tutor supplying information as the system moves from pumps, to hints, to prompts, to assertions.

We believe that AutoTutor's tuning of the student model and the optimization of strategic elicitation is superior to what a human could ever accomplish. Humans simply cannot handle such precise computations. However, this advantage of the computer may be offset by more potentially sophisticated strategies of AutoTutor than the expectation plus misconception tailored dialogue. However, Graesser et al. (1995) carefully documented that human tutors rarely implement such sophisticated strategies, even expert tutors (Graesser, D'Mello, & Cade, 2012). Therefore, it is not entirely science fiction to propose that the computer tutors may exceed human tutors in improving learning.

How well does AutoTutor help students learn? AutoTutor has significantly helped students learn in dozens of experiments that target the areas of computer literacy and conceptual physics. The system shows learning gains of approximately 0.80 sigma (standard deviation units) compared with pretests or with a condition that has students read a textbook for an equivalent amount of time (Graesser, Lu et al., 2004; VanLehn et al., 2007). It is most effective for deeper conceptual levels of comprehension and reasoning (e.g., why, how, what-if), as opposed to shallow facts (e.g., who, what, when, where). VanLehn et al. (2007) reported that AutoTutor produced the same learning gains as expert human tutors when the humans interacted with the students in computer-mediated communication. Such results are very encouraging.

AutoTutor conversations are not always coherent but they do help the students learn and the dialogue is adequate for students to get through the sessions with minimal irregularities. It is difficult for third-person bystander judges to decide whether the content of a particular turn in the dialogue was generated by AutoTutor or by an expert human tutor of computer literacy (Person & Graesser, 2002). Person and Graesser randomly sampled AutoTutor turns and half of the time substituted content generated by human tutors at the sample points in the dialogue. The judges received written transcripts of the experimentally manipulated tutorial dialogues and decided whether each move was generated by a computer or a human. The judges could not discriminate whether particular turns were generated by humans or AutoTutor. This is a remarkable success in AutoTutor simulating human dialogue. However, observers would no doubt be able to decide whether a *sequence* of turns is a conversation with AutoTutor versus a human tutor.

The successes of AutoTutor in promoting learning and simulating human tutoring are of course very encouraging. However there are a number of shortcomings of AutoTutor that need

to be acknowledged. AutoTutor does sometimes makes errors in evaluating the quality of student contributions. This results in AutoTutor's short feedback being incorrect (e.g., negative instead of positive) and the tutors' hints or prompts being a bit off the mark (e.g., eliciting information that the student has already expressed). Sometimes AutoTutor makes errors in classifying student contributions to the correct speech act category, e.g., question, assertion, meta-comment ("I'm lost"), so AutoTutor's response is not relevant and coherent. AutoTutor cannot answer many of the student questions; some answers do not seem relevant so students are prone to stop asking questions. AutoTutor is limited in its mixed-initiative dialogue because it cannot handle changes in topics, tangents, and off-the-record contributions of students.

The Progeny of AutoTutor

Versions of AutoTutor and derivatives of AutoTutor have evolved since its inception in 1997. Table 1 presents a list of the systems with conversational agents that have been developed in my collaborations with colleagues in Memphis or at other universities, often with their taking the lead on these funded projects. The design of all of these systems is grounded in principles of learning that are endorsed by the cognitive and learning sciences (Graesser, Halpern, & Hakel, 2009). Most of these systems have facilitated learning when tested on students in middle school, high school, or college.

EMOTIONS WITH AUTOTUTOR

Emotions with Conversational Agents

Conversational agents have recently been designed to respond to student emotions in addition to their cognitive states.   An adequate understanding of affect-learning connections is essential to the design of engaging educational artifacts that range from responsive intelligent

tutoring systems on technical material to entertaining media and games. Therefore, our designs

of AutoTutor and other systems with agents have documented the emotions that learners

experience while using these advanced learning environments (Baker, D'Mello, Rodrigo, &

Graesser, 2010; D'Mello, Craig, & Graesser, 2009). Our recent emotion-sensitive AutoTutor

(*AutoTutor-ES*) automatically detects learner emotions based on multiple channels of discourse

(D'Mello & Graesser, 2010) and responds appropriately to the students' affect states by selecting

appropriate discourse moves and displaying emotions in facial expressions and speech (D'Mello

& Graesser, in press).

The role of emotions in complex learning has been explored in the context of human

tutoring, classrooms, and other educational contexts (Lepper & Woolverton, 2002; Meyer &

Turner, 2006; Pekrun, 2006) in addition to more general cognition activities (Bower, 1992;

Mandler, 1984; Ortony, Clore, & Collins, 1988). Interestingly, the "universal" emotions that

Ekman (1992) investigated (e.g., sadness, happiness, anger, fear, disgust, surprise) have minimal

relevance to learning-centered emotions, where the dominant affective states include confusion,

frustration, boredom, flow/engagement, delight, and surprise (Baker et al., 2010; D'Mello et al.,

2009). The affect state of anxiety also occurs when students are being evaluated.

The cognitive-affective state of confusion is particularly interesting because it

theoretically is expected to play an important role in learning and empirically has a positive

correlation with learning gains (D'Mello et al., 2009; Graesser et al., 2009). Confusion is

diagnostic of *cognitive disequilibrium*, a state that occurs when learners face obstacles to goals,

contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Festinger, 1957;

Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Piaget, 1952). Cognitive equilibrium is

restored after thought, reflection, problem solving and other effortful cognitive activities.

Cognitive disequilibrium is a critical juncture in the learning process that is sensitive to individual differences. Some students give up when experiencing confusion because they have a self-concept that they are not good at the subject matter or they prefer not to receive negative feedback (Dweck, 1999; Meyer & Turner, 2006). Other students treat confusion as a challenge to conquer and expend cognitive effort to restore equilibrium. The first type of student needs encouragement, hints, and prompts to get the student over the hurdle, whereas the second type would best be left to the student's own devices. An adaptive tutor would treat these students differently.

AutoTutor-ES responds to different profiles of the students' emotional and cognitive states (D'Mello & Graesser, in press). If the learner is frustrated, for example, the tutor gives hints or prompts to advance the learner in constructing knowledge and makes supportive empathetic comments to enhance motivation. If the learner is bored, the tutor presents more engaging material or challenging problems for the more knowledgeable learner. The tutor continues business as usual when the learner is in a state of flow (Csikszentmihaly, 1990), i.e., when the learner is so deeply engaged in learning the material that time and fatigue disappear. The emotions of delight and surprise are fleeting, so there is no need to respond to these states in any special way. AutoTutor's intervention when the student is confused is both critical and complex, as previously discussed. One speculation is that each student has a *zone of optimal confusion* that varies with the student's background knowledge and interest in the subject matter.

An automated emotion classifier is necessary for AutoTutor-ES to be responsive to learner emotions. We have developed and tested an automated emotion classifier for AutoTutor based on the dialogue history, facial action units, and position of student's body during tutoring (D'Mello, Dale, & Graesser, in press; D'Mello & Graesser, 2010). There are systematic

relations between these sensing channels and particular emotions. With respect to dialogue history, emotions are predicted by (a) the occurrence of AutoTutor's feedback, (b) the type of feedback (positive, neutral, negative), (c) the directness of AutoTutor's dialogue moves (e.g., hints are less direct than assertions), (d) the quality of learner's contributions, and (e) the phase of the tutoring session (early versus late). Regarding the nonverbal channels, emotions are correlated with particular facial expressions, posture, and face-posture-dialogue combinations. Confusion, surprise, and delight are most directly manifested on facial expressions, whereas frustration is best predicted by dialogue history, and posture dynamics are needed to discriminate boredom, engagement/flow, and neutral states. AutoTutor's body pressure measurement system has revealed that bored students either fidget or have a large distance between their face and the screen. The features from the various modalities can be detected in real time automatically on computers, so we have integrated these sensing technologies with AutoTutor-ES.

It is too early to make any firm conclusions about the impact of AutoTutor-ES on learning, but we have conducted some studies. We have compared the original AutoTutor without emotion tracking to an AutoTutor version that is emotionally supportive. The supportive AutoTutor would have polite and encouraging positive feedback ("You're doing extremely well") or negative feedback ("This is difficult for most students"). There is another version that tries to shake up the emotions of the student by being playfully rude and telling the student what emotion the student is having ("I see that you are frustrated"). Instead of giving earnest feedback, the rude AutoTutor gives positive feedback that is sarcastic (e.g., "Aren't you the little genius") and negative feedback that is derogatory (e.g., "I thought you were bright, but I sure pegged you wrong"). The simple substitution of this feedback dramatically changes AutoTutor's personality. The rude tutor is very engaging for some students whereas other students would prefer to interact with the polite

supportive tutor. The data we have collected reveals that the impact on learning appears to depend on the phase of tutoring and the student's level of mastery. An emotion-sensitive AutoTutor had either no impact or a negative impact on learning during early phases of the tutoring session. During the later stages, the polite supportive AutoTutor improved learning, but only for the low knowledge students. Although more studies need to be conducted, it is tempting to speculate that emotional displays by AutoTutor may not be beneficial during the early phases of an interaction when the student and agent are "bonding," that a supportive polite tutor is appropriate at later phases for students who have low knowledge and abilities, and that the playful rude tutor is motivating when boredom starts emerging for the more confident, high-knowledge learners.

Emotions are of course central to the design of educational games (Conati, 2002; McNamara, Jackson, & Graesser, in press; Millis et al., in press; Moreno & Mayer, in press; Shaffer, 2006). Educational games ideally are capable of turning work into play by minimizing boredom, optimizing engagement/flow, presenting challenges that reside within the optimal zone of confusion, preventing persistent frustration, and engineering delight and pleasant surprises.

We have reached a point in history when computers can simulate (or emulate) many aspects of discourse comprehension, generation, and interaction. The vision of humans communicating with computers in natural language has fascinated science fiction writers for decades. This vision has been shifting from science fiction to reality with advances in computational linguistics, corpus linguistics, artificial intelligence, information retrieval, data mining, and discourse processing (Graesser, Gernsbacher, & Goldman, 2003; Jurafsky & Martin, 2008; Landauer, McNamara, Dennis, & Kintsch, 2007; Shermis & Burstein, 2003). There are

computer tutors that hold conversations in natural language and help students learn subject

matters almost as well as human tutors (Graesser, Jeon, & Dufty, 2008). There are systems that

can grade student essays as well as experts in English composition (Shermis & Burstein, 2003).

Computer systems can detect the emotions of learners on the basis of dialogue history, facial

expressions, and body posture with accuracy scores on par with humans trained to detect

emotions (D'Mello & Graesser, 2010). A system called *Linguistic Inquiry Word Count* (LIWC,

Pennebaker, Booth, & Francis, 2007) can identify the personalities, social status, and other

psychological characteristics of writers by classifying the words they use on dozens of

psychological categories. Discourse patterns can unveil the characteristics and status of

individuals that vary from political leaders to terrorists (Hancock et al., 2010). The fusion of

psychology with computer science, linguistics, and other fields has not only advanced the

science of discourse, but has also provided learning environments and useful text analysis tools.

References

Baker, R.S., D'Mello, S.K., Rodrigo, M.T., & Graesser, A.C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68,* 223-241.

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.

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,* 301-341.

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.

Conati C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence,* 16*,* 555-575.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper-Row.

D'Mello, S.K., Craig, S.D., & Graesser, A.C. (2009). Multi-method assessment of affective experience and expression during deep learning. *International Journal of Learning Technology, 4*, 165-187.

D'Mello, S. K., Dale, R. A., & Graesser, A. C. (in press). Disequilibrium in the mind, disharmony in the body. *Cognition & Emotion*.

D'Mello, S., Dowell, N., & Graesser, A.C. (in press). Does it really matter whether students' contributions are spoken versus typed in an intelligent tutoring system with natural language? *Journal of Experimental Psychology: Applied.*

D'Mello, S., & Graesser, A.C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20*, 147-187.

D'Mello, S., & Graesser, A.C. (in press). Emotions during learning with AutoTutor. In P.J. Durlach and A. Lesgold (Eds.), *Adaptive technologies for training and education.* Cambridge: Cambridge University Press.

Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality, and development.* Philadelphia, PA: The Psychology Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*, 169-200.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Graesser, A. C., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly, 70*, 3–22.

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

Graesser, A.C., Halpern, D.F., & Hakel, M. (2008). 25 Principles of Learning. Washington, DC: Taskforce on Lifelong Learning at Work and at Home. Retrieved from http://psyc.memphis.edu/learning/whatweknow/index.shtml

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. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*.

Graesser, A.C., McNamara, D.S., & Kulikowich, J. (in press). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher.*

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193-202.

Graesser, A., Ozuru, Y., & Sullins, J. (2009). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (pp. 112-141). NY: Guilford.

Graesser, A. C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed initiative dialogue in natural language. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 243–262). Mahwah, NJ: Erlbaum.

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

Hancock, J.T., Beaver, D.I., Chung, C.K., Frazee, J., Pennebaker, J.W., Graesser, A., & Cai, Z. (2010). Social language processing: A framework for analyzing the communication of terrorists and authoritarian regimes. *Behavioral Sciences of Terrorism and Political Aggression, 2,* 108-132.

Johnson, L. W. & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In M. Goker and K. Haigh (Eds.), *Proceedings of the Twentieth Conference on Innovative Applications of Artificial Intelligence* (pp. 1632-1639). Menlo Park, CA: AAAI Press.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Englewood, NJ: Prentice Hall.

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

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

Lepper, M., & 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

Litman, D.J, Rose, C.P., Forbes-Riley, K., VanLehn, K., Bhembe, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education*, *16*, 145-170.

McNamara, D.S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I.B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies (pp. 397–421)*. Mahwah, NJ: Erlbaum.

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, Forsyth, C., Butler, H., Wallace, P., Graesser, A.,& Halpern, D. (*in press*) Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou & J. Lakhmi (Eds*.) Serious* g*ames and edutainment applications*. London, UK: Springer-Verlag.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315-341.

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

Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.

Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher, 39*, 183-197.

Rus, V., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2008). A study of textual entailment. *International Journal on Artificial Intelligence Tools*, *17*, 659–685.

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

Shermis, M.D., & Burstein, J. (2003)(Eds.), *Automated essay scoring: A cross-disciplinary perspective.* Hillsdale, NJ: Erlbaum.

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

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

Table 1: *Learning Environments with Conversational Agents Developed by Graesser and Collaborators.*

SYSTEM	BRIEF DESCRIPTION	LEADER/ COLLAB
AutoTutor	Conversational tutor on computer literacy and physics	
AutoTutor-3D	Physics with embedded interactive simulation in 3D world	Tanner Jackson
AutoTutor-Lite	Simplified discourse applied to powerpoint on any topic	Xiangen Hu
AutoTutor-ES	AutoTutor being sensitive to learners' emotions	Sidney D'Mello
AutoMentor	Multiparty serious game with mentor on urban planning	David Shaffer
DeepTutor	Physics tutor with deep natural language processing	Vasile Rus
GnuTutor	Open source version of AutoTutor on any topic	Andrew Olney
GuruTutor	Biology tutor with deep natural language and pointing	Andrew Olney
HURAA Advisor	Web tutor on ethical treatment of subjects in experiments	Xiangen Hu
iDRIVE	Learning to ask deep questions on science topics	Barry Gholson & Sco
iSTART, iSTART-ME	Learning to generate self-explanations while reading text	Danielle McNamara
MetaTutor	Learning skills of self-regulated learning & metacognition	Roger Azevedo
Operation ARIES!	Critical reasoning about scientific methods	Keith Millis, Diane H

		Zhiqiang Cai
Writing-Pal	Learning to write argumentative essays	Danielle McNamara