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

Arthur C. Graesser, Sidney K. D'Mello, & Amber C. Strain
University of Memphis
400 Innovation Drive, Memphis, TN, 38152-3230
901-678-4857
[a-graesser; sdmello; dchuncey] @memphis.edu

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 novice human tutors in addition to incorporating ideal pedagogical strategies of expert tutors. The agents are not perfect conversation partners and comprehenders of language, but the conversations are surprisingly coherent and also help students learn.

Keywords
Agents, tutoring, discourse, learning technologies, emotions

1. INTRODUCTION
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 [7] [22] [51]. 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 complain 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 deeply understand what students know and do not implement sophisticated strategies to help them learn [22] [31]. 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 quite effective in helping students learn. Indeed, they are 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 [6], and vary from .2 to 2.0 for accomplished human tutors [4] [56]. Collaborative peer tutoring even shows an effect size advantage of 0.2 to 0.9 sigma [36] [43] [55]

2. WHAT DO HUMAN TUTORS DO?
Given that human tutors are effective, what is it they do to help students learn? It should be noted that we use the term *tutor* to refer to novice human tutors. These tutors are untrained in tutoring skills and have moderate domain knowledge. They were peer tutors, cross-age tutors, or paraprofessionals, but rarely are accomplished professionals or expert tutors.

An analysis of the collaborate dialogue patterns of these tutors has indicated that they 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 answer with 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."

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

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

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

(7) *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.

(8) *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 [5] 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 Cade [22] 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 that 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 [23] [31]. 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 is an Intelligent Tutoring System (ITS) that was designed to achieve to simulate human tutoring and to implement more ideal tutoring mechanisms.

3. AUTOTUTOR

AutoTutor [24] [25] 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 modeling and optimizing the 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* [42] *Tactical Language and Culture Training System* [37], *Why-Atlas* [56], *Operation ARIES!* [46], and *iSTART* [44]. 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 [30] [53]. 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, it keeps 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 [40] and symbolic interpretation algorithms [52] 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 $20 \times 5 = 100$ 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.

When one expectation is covered, AutoTutor selects the next expectation to target 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. 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 maximize the match score if the hint was answered correctly. Stated differently, a correct answer to the hint would maximally increase the coverage of the expectation. However, hints might not work out in every case, 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 initial 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 [30]. There is a continuum from the student to the tutor supplying information as the system moves from pumps, to hints, to prompts, to assertions. The correlations with student knowledge reflected this continuum perfectly, with values of .49, .24, -.19, and -.40, respectively, when the relative frequencies of dialogue moves are correlated with the students' prior knowledge of physics.

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, [31] carefully documented that human tutors rarely implement such sophisticated strategies, even expert tutors [22]. 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 [26] [56]. 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). Van Lehn et al. [56] 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 [50]. 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 make 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, discourse tangents, and off-topic contributions of students.

4. THE PROGENY OF AUTOTUTOR

Versions of AutoTutor and its derivatives have evolved since its inception in 1997. The different versions were designed to incorporate particular pedagogical goals and cover different topics. So far, the topics have covered computer literacy, physics, biology, tactical planning, and critical thinking. In most versions of AutoTutor, the students type in their contributions via keyboard, whereas recent versions allow spoken input [12] [15]. Available data have revealed that it doesn't matter whether the student input is typed versus spoken as long as the speech recognition is reasonably accurate.

The interface of AutoTutor has a number of different windows that show: the conversational agent, the main question to work on, the information entered by the student, the dialogue history (optional), and an image or interactive simulation (optional). Comparisons between the AutoTutor communication media have compared the full conversational agent versus spoken only, text only, and various combinations of these with facial expressions. There is a slight advantage for the full conversational agent, but the increment is surprisingly modest. The important factor lies in content: what is said at the right time for the right student.

One version of AutoTutor that has an interactive simulation (see Figure 1). This *AutoTutor-3D* version guides learners on using interactive simulations of physics microworlds [21] [35]. The student manipulates parameters of the situation (e.g., mass of objects, speed of objects, distance between objects) and then asks the system to simulate what will happen. Students are also prompted to describe what they see. Their actions and descriptions are evaluated with respect to covering the expectations or matching misconceptions. AutoTutor manages the dialogue with hints and suggestions that scaffold the learning process with dialogue. The interactive simulation has indeed been found to have added value over the conversational AutoTutor, but only those students who actually use the simulation environment at a sufficient frequency, ideally in a strategic fashion. Unfortunately, most students did not use the simulation environment with enough frequency or strategic approach to lead to learning gains. Students need to be trained and scaffolded on such self-regulated learning environments [27] [39].

Figure 1. AutoTutor with interactive simulation

AutoTutor-Lite [33] is a version of AutoTutor that that can be developed quickly and applied to a multitude of subject matters. There are authoring tools for researchers to develop content on a

practical production schedule. Imagine AutoTutor being up and running for a novel topic in 2-3 days and available on the web for the world to use. There is a talking head, like most versions of AutoTutor, that can be integrated with different Internet platforms. The major bottleneck lies in the authoring tools, although they are sufficient for individuals with minimal training to develop curriculum content. One downside of AutoTutor-Lite is that there is limited depth in the discourse interactions. However, for those who want to spend more time and build a deeper AutoTutor on new subject matters, the open source version of GnuTutor is available [47].

Researchers in the IIS at University of Memphis have developed or are developing other learning environments with agents that are derivatives of AutoTutor. Some of the prominent ones are listed below.

(1) DeepTutor (led by Vasile Rus) is a physics tutor with deep natural language processing that implements dialogue strategies that overcome the tutor illusions discussed earlier.

(2) GuruTutor (led by Andrew Olney) is a biology tutor that models expert human tutors. It covers a large biology curriculum and assessments, with a full-bodied agent that can strategically point to pictorial elements.

(3) HURA Advisor [34] has a narrator agent that guides the student how to learn about research methods by a large number of learning modules with multimedia.

(4) iDRIVE [20] has a tutor agent and student agent who interact by asking answering questions so that students can vicariously learn how to ask good questions.

(5) MetaTutor [1] has multiple agents train students how to learn about metacognition and self-regulated learning on topics in biology.

(6) Operation ARIES! [46] is a game environment where students learn about critical thinking about science methodology. There are multiple agents, including trialogues with a tutor and peer student interacting with a human. This system is being commercialized by Pearson Education.

(7) AutoMentor [54] simulates a mentor who helps students learn about urban planning in a multiparty serious game.

There are additional systems with agents developed at the University of Memphis that are indirectly inspired by AutoTutor, such as the iSTART system that trains students how to read deeper with self-explanation reading strategies [44] and the Writing Pal system that trains students how to write (led by Danielle McNamara).

All of these derivatives of AutoTutor are designed with explicit principles of learning and are tested on students to see whether there are improvements in learning, perceptions, and motivation. This is a very different mission that building fancy agents that have little or no grounding in the learning sciences.

5. AUTOTUTOR WITH EMOTIONS
A version of AutoTutor has 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 [2] [10]. Our recent emotion-sensitive AutoTutor (*Affective AutoTutor*) automatically detects learner emotions (confusion, frustration, and boredom) based on multiple channels of communication (discourse, gross body movements, and facial expressions [13]) and responds appropriately to the students' affect states by selecting appropriate discourse moves and displaying emotions in facial expressions and speech [14].

The role of emotions in complex learning has been explored in the context of human tutoring, classrooms, and other educational contexts [8] [41] [45] [48] [49]. Interestingly, the "universal" emotions that Ekamn [19] 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, curiosity, and (sometimes) delight and surprise [2] [10]. The affect state of anxiety also occurs when students are being evaluated.

The cognitive-affective state of confusion theoretically is expected to play an important role in learning and indeed does empirically have a positive correlation with learning gains [10] [29]. 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 restored after thought, reflection, problem solving and other effortful cognitive activities. 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 [18] [45]. 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.

Affective AutoTutor responds to different profiles of the students' emotional and cognitive states [14]. If the learner is frustrated, for example, the tutor makes supportive empathetic comments to enhance motivation. If the learner is bored, the tutor tries to eengage the learner with encouraging dialogue moves. The tutor continues business as usual when the learner is in a state of engagement or flow [9], i.e., when the learner is deeply engaged in learning the material. 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 Affective AutoTutor 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 [11] [13]. 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

5

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 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 Affective AutoTutor.

It is too early to make any firm conclusions about the impact of Affective AutoTutor on learning, but we have conducted some initial studies. We have compared the original AutoTutor without emotion tracking to an AutoTutor version that is emotionally supportive (Supportive Tutor). The supportive AutoTutor would have polite and encouraging positive feedback ("You're doing extremely well"; "This is difficult for most students. So keep trying"). These empathetic and encouraging responses always attribute the source of the students' emotion to the material instead of the students themselves. There is another version that attributes the students' negative emotions to the students themselves (Shakeup Tutor). For example, possible shakeup responses to confusion are, "This material has got *you* confused, but I think you have the right idea. Try this…" and "*You* are not as confused as you might think. I'm actually kind of impressed. Keep it up". Another difference between the two versions lies in the conservational style. While the Supportive AutoTutor is subdued and polite, the Shakeup tutor is edgier, flaunts social norms, and is witty

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 Shakeup tutor is motivating when boredom starts emerging for the more confident, high-knowledge learners.

6. CONCLUDING REMARKS

In closing, we have reached a point in history when computers can simulate (or emulate) human tutors on many levels of language, discourse, and pedagogy. The vision of humans communicating with computers in natural language has fascinated science fiction writers for decades. This vision appears to be shifting from science fiction to reality with advances in computational linguistics, corpus linguistics, artificial intelligence, information retrieval, data mining, affective computing, and discourse processing. These recent advances in automated tutorial dialogue would never have occurred without a deep interdisciplinary fusion among these diverse fields.

8. REFERENCES

[1] Azevedo, R., A. Witherspoon, A.C. Graesser, D.S. McNamara, A. Chauncey, E. Siler, Z. Cai, and M. Lintean, "MetaTutor: Analyzing self-regulated learning in a tutoring system for biology," In V. Dimitrova, R. Mizoguchi, B. Du Boulay, and A. C. Graesser (eds.), *Proc. of 14th Int. Conf. on Artificial Intelligence in Education.* (pp. 635-637). IOS Press, Amsterdam, The Netherlands (2009).

[2] Baker, R.S., S.K. D'Mello, M.T. Rodrigo, and A.C. Graesser, "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," *Int. J. Human-Computer Studies, 68,* 223-241 (2010).

[3] Bloom, B. S., "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring." *Ed. Researcher, 13,* 4-16 (1984).

[4] Chi, M. T. H., M. Roy, and R. G. M., Hausmann, "Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning," *Cognitive Science, 32,* 301-341 (2008).

[5] Clark, H. H., *Using language.* Cambridge University Press, Cambridge, (1996).

[6] Cohen, P. A., J. A. Kulik, and C. C. Kulik, "Educational outcomes of tutoring: A meta-analysis of findings." *Amer. Ed. Research Journal, 19,* 237-248 (1982).

[7] Collins, A., and R. Halverson, *"Rethinking education in the age of technology: The digital revolution and schooling in America,"* Teachers College Press, New York, NY.

[8] Conati C., "Probabilistic assessment of user's emotions in educational games," *J. Applied. Artificial Intell., 16,* 555-575 (2002).

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

[10] D'Mello, S.K., S.D. Craig, and A.C. Graesser, "Multimethod assessment of affective experience and expression during deep learning." *Int. J. Learning Technology, 4,* 165-187 (2009).

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

[12] D'Mello, S.K., N. Dowell, and A.C. Graesser, "Does it really matter whether students' contributions are spoken versus typed in an intelligent tutoring system with natural language?" *J. Exper. Psych: Applied, 17*(1), 1-17 (2011).

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

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

[15] D'Mello, S. K., B.G. King, P. Chipman, and A.C. Graesser, "Towards spoken human-computer tutorial dialogues," *Human-Computer Int., 25*(4), 289-323 (2010).

[16] D'Mello, S. K., B. Lehman, J. Sullins, R. Daigle, R. Combs, K. Vogt, et al. A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In J. Kay and V. Aleven (eds.), Proceedings of 10th Int. Conf. on Intell. Tutoring Systems, pp. 245-254, Springer, Berlin / Heidelberg, , (2010).

[17] D'Mello, S. K., R. W. Picard, and A.C. Graesser, "Towards an Affect-Sensitive AutoTutor," Special issue on Intelligent Educational Systems – IEEE Intelligent Systems, 22(4), 53-61, (2007)

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

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

[20] Gholson, B, R. Coles, and S.D. Craig, "Features of computerized multimedia environments that support vicarious learning processes." In M.S. Khine and I.M. Salah (eds.), *New science of learning: Cognition, computers, and collaboration in education,* Springer; New York, NY, pp. 53-78 (2010).

[21] Graesser, A. C., P. Chipman, B. Haynes, and A. Olney, "AutoTutor: An intelligent tutoring system with mixed-initiative dialogue." *IEEE Transactions on Education, 48*(4), 612-618 (2005).

[22] Graesser, A.C., S.K. D'Mello, and W. Cade, "Instruction based on tutoring." In R.E. Mayer and P.A. Alexander (eds), *Handbook of Research on Learning and Instruction,* Routledge, New York, NY, pp. 408-426 (2011).

[23] Graesser, A. C., S.K. D'Mello, and N Person, "Meta-knowledge in tutoring". In D. J. Hacker, J. Dunlosky, and A. C. Graesser (eds.) *Metacognition in educational theory and practice.* Erlbaum, Mahwah, NJ (2009).

[24] Graesser, A. C., M. Jeon, and D. Dufty, "Agent technologies designed to facilitate interactive knowledge construction," *Discourse Processes, 45*, 298–322 (2008).

[25] Graesser, A.C., S. Lu, G.T. Jackson H. Mitchell, M. Ventura, A. Olney, and M.M. Louwerse, AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers, 36*, 180-193 (2004).

[26] Graesser, A. C., S. Lu, B.A. Olde, E. Cooper-Pye, and S. Whitten, "Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down," *Memory and Cognition, 33*, 1235–1247 (2005).

[27] Graesser, A. C., and D.S. McNamara, "Technologies that support reading comprehension," In C. J. Dede, and J. Richards (eds.), *Digital teaching platforms.* Harvard Education Press, Cambridge, MA (in press).

[28] Graesser, A. C., D.S. McNamara, M.M. Louwerse, and Z. Cai, "Coh-Metrix: Analysis of text on cohesion and language," *Beh. Res. Methods, Instruments, and Computers, 36*, 193-202 (2004).

[29] Graesser, A., Y. Ozuru, and J. Sullins, "What is a good question?" In M. G. McKeown and L. Kucan (eds.), *Threads of coherence in research on the development of reading ability.* Guilford, New York, pp. 112-141. (2009).

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

[31] Graesser, A. C., N.K. Person, and J.P Magliano, "Collaborative dialogue patterns in naturalistic one-to-one tutoring," *Applied Cognitive Psych., 9*(6), 495-522 (1995).

[32] Graesser, A. C., and N.K. Person, "Question asking during tutoring," *Amer. Ed. Research J., 31*, 104-137 (1994).

[33] Hu, X., Z. Cai, L, Han, S.D Craig, T. Wang, and A.C. Graesser, "AutoTutor lite," *Proc. 2009 Conf. on AIED,* IOS Press, Amsterdam, The Netherlands, p.802 (2009).

[34] Hu, X., and A.C. Graesser, "Human use regulatory affairs advisor (HURAA): Learning about research ethics with intelligent learning modules," *Beh. Res. Methods, Instruments, & Computers, 36*(2), 241-249 (2004).

[35] Jackson, G. T., A. Olney, A.C. Graesser, and H.J. Kim, "AutoTutor 3-D simulations: Analyzing user's actions and learning trends," In R. Son (ed.), *Proc. of the 28th Ann. Meetings of the Cognitive Science Society,* Erlbaum, Mahwah, NJ, pp. 1557-1562 (2006).

[36] Johnson, D.W., and R.T Johnson, "Implementing cooperative learning." *Contemporary Education, 63*(3), 173–180 (1992).

[37] Johnson, L. W. and A. Valente, "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,* AAAI Press, Menlo Park, CA, pp. 1632-1639 (2008).

[38] Jurafsky, D., and J. Martin, *Speech and language processing.* Prentice Hall, Englewood, NJ (2008).

[39] Klahr, D., "Developing elementary science skills: Instructional effectiveness and path independence. *Cog. Development, 23*(4), 448-511 (2008).

[40] Landauer, T., D.S. McNamara, S. Dennis, and W. Kintsch, eds., *Handbook of Latent Semantic Analysis.* Erlbaum, Mahwah, NJ (2007).

[41] Lepper, M., and M. Woolverton, "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*, Academic Press, Orlando, FL, pp. 135-158 (2002).

[42] Litman, D.J, C.P. Rose, K. Forbes-Riley, K. VanLehn, D. Bhembe, and S. Silliman, "Spoken versus typed human and computer dialogue tutoring," *Int. J. Artificial Intell. in Ed, 16*, 145-170 (2006).

[43] Mathes, P. G., and L.S. Fuchs, "Peer tutoring in reading for students with mild disabilities: A best evidence synthesis," *School Psych. Rev., 23*, 59-80 (1994).

[44] McNamara, D.S., T. O'Reilly, M. Rowe, C. Boonthum, and I.B. Levinstein, "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*. Erlbaum, Mahwah, NJ, pp. 397–421 (2007).

[45] Meyer, D. K., and J.C. Turner, "Re-conceptualizing emotion and motivation to learn in classroom contexts," *Ed. Psych. Review, 18* (4), 377-390 (2006).

[46] Millis, K., C. Forsyth, H. Butler, P. Wallace, A. Graesser, and D. Halpern. "Operation ARIES! A serious game for teaching scientific inquiry," In M. Ma, A. Oikonomou and J. Lakhmi (eds*.) Serious games and edutainment applications*. Springer-Verlag, London, UK (*in press*).

[47] Olney, A. "GuruTutor: An open source intelligent tutoring system based on AutoTutor. *Proc. AAAI 2009 Fall Symposium on Cog. and Metacog Ed. Systems,* AAAI Press, Washington, DC, pp. 70-75, 2009.

[48] Ortony, A., G. L. Clore and A. Collins, *The cognitive structure of emotions,* Cambridge University Press (1988).

[49] Pekrun, R., "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," *Ed. Psych. Rev., 18*, 315-34 (2006).

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

[51] Resnick, L. B., "Nested learning systems for the thinking curriculum," *Ed. Researcher, 39*, 183-197 (2010).

[52] Rus, V., and A.C. Graesser, "Deeper natural language processing for evaluating student answers in intelligent tutoring systems." *Proc. of the Amer. Assoc. of Artificial Intell.*, AAAI, Menlo Park, CA (2006).

[53] Rus, V., P.M. McCarthy, D.S. McNamara, and A.C. Graesser, "A study of textual entailment," *Int. J. Artificial Intell. Tools, 17*, 659–685 (2008).

[54] Shaffer, D. W., and A.C. Graesser, "Using a quantitative model of participation in a community of practice to direct automated mentoring in an ill-formed domain. In C. Lynch, K. Ashley, T. Mitrovic, V. Dimitrova, N. Pinkwart, and V. Aleven (eds.), *Proc. of the 4th Int. Workshop on Intelligent Tutoring Systems and Ill-Defined Domains,* ITS 2010 pp. 61-68 (2010).

[55] Topping, K., "The effectiveness of peer tutoring in further and higher education: A typology and review of the literature," *Higher Education, 32*, 321-345 (1996).

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