

Exploring Relationships Between Affect and Learning with AutoTutor

Arthur GRAESSER, Sidney D'MELLO, Patrick CHIPMAN,
Brandon KING, and Bethany McDANIEL
Institute for Intelligent Systems, The University of Memphis
365 Innovation Drive, University of Memphis, Memphis, TN, 38152, USA

Abstract. The relationship between emotions and learning was investigated by tracking the emotions that college students experienced while learning about computer literacy with AutoTutor. AutoTutor is an animated pedagogical agent that holds a conversation in natural language, with spoken contributions by the learner. Thirty students completed a multiple-choice pre-test, a 35-minute training session, and a multiple-choice post-test. After completing the post-test, the students reviewed the tutorial interaction and were stopped at strategically sampled points for emotion judgments. At these points they judged what emotions they experienced on the basis of the dialogue history and their facial expressions. The emotions they judged were boredom, flow (engagement), frustration, confusion, delight, surprise, and neutral. A multiple regression analysis revealed that post-test scores were significantly predicted by pre-test scores and confusion, but not by any of the other emotions. The results support a theoretical framework that emphasizes the role of cognitive disequilibrium (perplexity) in promoting learning at deeper levels. The long-term goal of this project is to have a version of AutoTutor that is sensitive to the affective and cognitive states of the learner, as they are detected from spoken dialogue and other non-verbal channels (i.e., facial expression, body position)

Keywords: Tutorial dialogue, emotions, learner modeling

Theoretical frameworks that predict systematic relationships between emotions and complex learning are just beginning to emerge in the fields of psychology, neuroscience, education, and computer science. Three theories that have emerged highlight the contributions of academic risk taking, flow, and goals (Meyer & Turner, 2006). The academic risk theory contrasts (a) adventuresome learners who want to be challenged with difficult tasks, take risks of failure, and manage negative emotions when they occur, with (b) cautious learners who tackle easier tasks, take fewer risks, and minimize failure and its resulting negative emotions. According to flow theory, learners are in a state of flow (Csikszentmihaly, 1990) when they are so deeply engaged in learning the material that time and fatigue disappear. Goal theory emphasizes the role of goals in predicting emotions, so outcomes that achieve challenging goals result in positive emotions whereas outcomes that jeopardize goal accomplishment result in negative emotions (Dweck, 2002; Stein & Hernandez, in press). Obstacles to goals are particularly diagnostic of both learning and emotions.

The affective state of confusion correlates with learning gains perhaps because it is a direct reflection of deep thinking (Craig, Graesser, Sullins, & Gholson, 2004). Confusion is diagnostic of *cognitive disequilibrium*, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). Cognitive equilibrium is restored after thought, reflection, problem solving and other effortful deliberations. Of course, it is important to differentiate the state of being productively confused, which leads to learning and positive emotions, from being hopelessly confused, which has no pedagogical value.

An affect-sensitive tutor would presumably enhance intelligent learning environments (D'Mello, Picard, & Graesser, in press; Graesser, Jackson, & McDaniel, 2007; Lepper & Henderlong, 2000). If the learner is frustrated, for example, the tutor would need to generate hints to advance the learner in constructing knowledge, and make supportive empathetic comments to enhance motivation. If the learner is bored, the tutor would need to present more engaging or challenging problems for the learner to work on. If the learner is confused, the tutor would need to guide the learner in productive learning trajectories that oscillate between cognitive disequilibrium and equilibrium. The ideal management of confusion no doubt depends on characteristics of the learner, e.g., high versus low domain knowledge, risky versus cautious, patient versus impulsive. At this point in the science, however, there is insufficient empirical research on relations between tutoring moves and affective states of the learner. In order to systematically investigate these relationships, we are in the process of developing a version of AutoTutor that is sensitive to both the cognitive and affective states of learners (D'Mello, Picard, & Graesser, in press; Graesser, Jackson, & McDaniel, 2007).

AutoTutor is an intelligent tutoring system that helps students learn by holding a conversation in natural language (Graesser, Chipman, Haynes, & Olney, 2005). Assessments of AutoTutor on learning gains have shown effect sizes of approximately .8 standard deviation units in the areas of computer literacy (Graesser et al., 2004) and Newtonian physics (VanLehn, Graesser et al., 2007). AutoTutor begins by presenting a challenging

question to the learner that requires about a paragraph of information to answer correctly. The typical response from the learner, however, is only one word to two sentences in length. Therefore, AutoTutor uses a series of pumps (“What else?”, “uh huh”) to request additional information and provides prompts for the learner to express specific words. AutoTutor also uses hints, assertions, and feedback to elicit responses from the learner that lead to a complete answer of the question.

The present study investigated the relationship between learning and the emotions that college students experience while interacting with AutoTutor on the topic of computer literacy. College students received a pre-test on the subject matter, interacted with AutoTutor for 35 minutes, completed a post-test, and reviewed the tutorial interaction to identify their emotions at strategically sampled points in the dialogue. The emotions they judged were those that were found to be prominent in our previous studies that examined emotions while learning with AutoTutor (Craig et al., 2004; D’Mello, Craig, Sullins, & Graesser, 2006) and other environments (Burlinson & Picard, 2004; Graesser et al., 2006; Kort, Reilly, & Picard, 2001). These included boredom, flow (engagement), frustration, confusion, delight, surprise, and neutral. We investigated how these emotions were correlated with pre-test and post-test scores. Whereas our previous studies of AutoTutor had learners enter their contributions through keyboard, this is the first AutoTutor study that had students make their contributions through speech. Properties of speech should be diagnostic of the learners’ emotions (Litman & Forbes-Riley, 2004).

2. Methods

2.1. Participants, Materials, and Procedure

The participants were 30 undergraduates at the University of Memphis who participated for course credit. The experiment consisted of a pre-test, an interaction with AutoTutor, a post-test, and judgments of emotions the learner experienced while interacting with AutoTutor. The participants were tutored with AutoTutor on one of three major computer literacy topics: hardware, operating systems, or the Internet. The pre-test and post-tests consisted of multiple choice questions that had been used in previous research on AutoTutor (Graesser, Lu et al., 2004). The questions tapped deep levels of reasoning, causality, and explanations. An example question is: “You buy a new graphics program but it will not run on your computer, so what is the best solution? (a) partition your hard drive, (b) delete unnecessary files, (c) increase RAM (correct answer), (d) return the program to the store.” There were 6 pre-test questions and 6 post-test questions associated with each of the major topics (hardware, operating systems, internet); on any given test of 10 questions, the participants answered the 6 questions associated with a topic assigned to them, plus 2 questions from each of the other two topics. However, only the 6 topic-related questions were analyzed in the present study. There were 2 versions of the 10 test questions (A and B) per major topic, which were counterbalanced across students in their assignment to pre-test and post-test. Performance in these multiple choice tests was simply the proportion of questions answered correctly.

2.2 AutoTutor

Participants interacted with AutoTutor for 35 minutes on one of three randomly assigned topics in computer literacy. A detailed discussion of the architecture, strategies, and effectiveness of AutoTutor are provided in previous publications, as cited above. Each major topic had 6 tutoring questions that required about a paragraph of information (3-7 sentences) in an ideal answer. The questions required answers that involved inferences and deep reasoning, such as *why*, *how*, *what-if*, *what if not*, *how is X similar to Y*? An example question about the internet is “How is a packet switching model of message transmission like the postal system?” A typical exchange had 50-100 turns to answer a single main question. One of the pedagogical philosophies behind AutoTutor was to get the student to articulate answer information, as opposed to AutoTutor being a mere information delivery system. This requires a careful management of the dialogue to get the learner to do the talking. We used the commercially available Dragon NaturallySpeaking™ (version 6) speech recognition system for speech-to-text translation.

Each turn of AutoTutor in the conversational dialogue had three information slots (i.e., units, constituents). The first slot of most turns was feedback on the quality of the learner’s last turn. This feedback was either positive (*very good, yeah*), neutral (*uh huh, I see*), or negative (*not quite, not really*). The second slot advanced the interaction with either prompts for specific information, hints, assertions with correct information, corrections of misconceptions, or answers to student questions. The third slot was a cue for the floor to shift from AutoTutor as the speaker to the learner. For example, AutoTutor ended each turn with a question or a gesture to cue the learner to do the talking. Discourse markers (*and also, okay, well*) connected the utterances of these three

slots of information. The conversations managed by AutoTutor were sufficiently smooth that learners could get through the session with minimal difficulties (Person & Graesser, 2002).

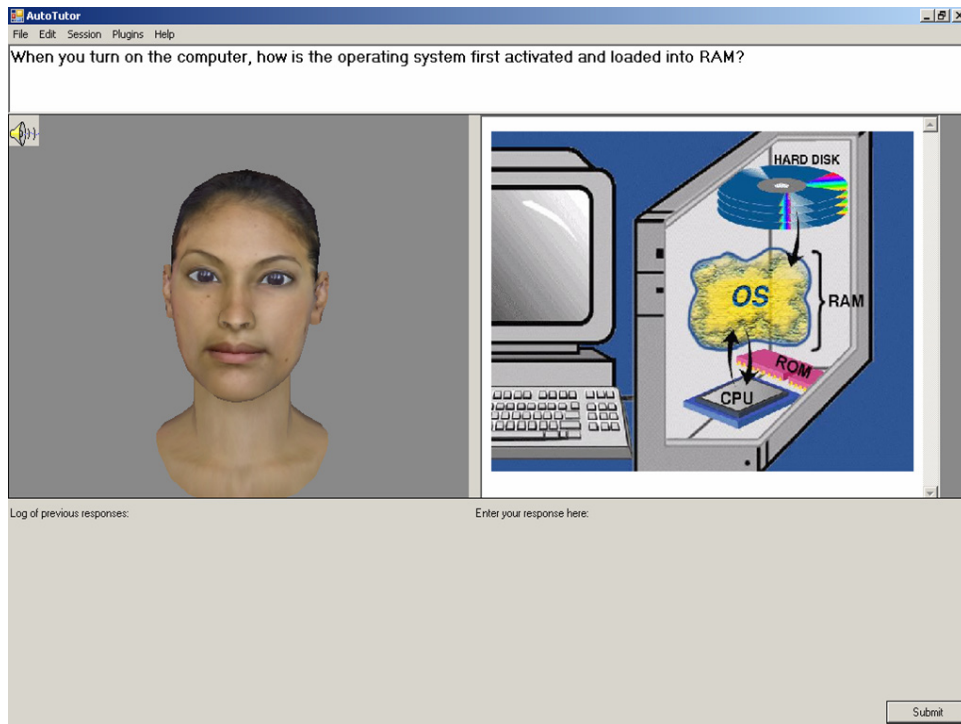


Figure 1. AutoTutor Interface

The AutoTutor interface had 3 major windows, as shown in Figure 1. Window 1 (top of screen) was the main question that stayed on the computer screen until the completion of that question. Window 2 (left middle) was an animated conversational agent that spoke the content of AutoTutor's turns. Window 3 (right middle) was either blank or had auxiliary diagrams. In addition to these interface components, there were 2 buttons on the keyboard that the learner pressed to start speaking and stop speaking.

2.3 Judging Affective States

Each learner watched his or her own session with AutoTutor after interacting with the tutor and completing the post-test. The judgments for a learner's tutoring session proceeded by playing a video of the learner's face that was captured during the tutorial session. The video also included an audio track that consisted of AutoTutor's speech along with the verbalized contributions of the learner. The participants were instructed to make judgments on what affective states were present at three different points during the tutorial dialogue: (1) immediately after AutoTutor gave the short feedback (positive, neutral, negative) during a turn, (2) immediately before the learner started expressing his or her spoken turn, and (3) other randomly selected points in the dialogue. The participants could also go back in between these points and make an emotion judgment. The data collection program provided a checklist of emotions for them to mark at these points. The participant was instructed to mark the affect state that was most pronounced at each point.

A list of the affective states and definitions was provided for the learners. The states were boredom, confusion, flow, frustration, delight, surprise, and neutral. These were the affective states that were most frequently experienced in previous studies of AutoTutor (Craig et al., 2004; Graesser et al., 2006). Boredom was defined as being weary or restless through lack of interest. Confusion was defined as a noticeable lack of understanding, whereas flow was a state of interest that results from involvement in an activity. Frustration was defined as dissatisfaction or annoyance. Delight was a high degree of satisfaction. Surprise was wonder or amazement, especially from the unexpected. Neutral was defined as no apparent emotion or feeling. Some of our colleagues may view some of these emotions (i.e., affect states) as cognitive states, whereas other researchers would classify them as either emotions or affect states (see Barrett, 2006; Stein & Hernandez, in press; Meyer & Turner, 2006). Our position agrees with the latter group because these states are accompanied by enhanced physiological arousal (compared with neutral) and affect-cognition amalgamations are particularly relevant to complex learning.

3. Results

A number of scores were computed for each of the 30 college students and averaged across the students. These included pre-test scores, post-test scores, and the proportion of first-choice emotion judgments in each of the 7 emotion categories. The pre-test and post-test scores varied from 0 to 1; a score of 0.25 would be chance guessing in these tests with a 4-alternative multiple-choice format. The first-choice emotion judgments would add to 1.0, given that the selection of an emotion was required in each sampled observation.

Table 1: Means, SD's, and correlations of emotions, pre-test, and post-test scores.

Measure	Mean	Standard Deviation	Correlation with pre-test	Correlation with post-test
Post-test score	.483	.193	.398*	
Pre-test score	.361	.244		
Confusion	.188	.098	.232	.490*
Frustration	.129	.086	-.019	.040
Flow (engagement)	.232	.199	-.163	-.121
Boredom	.190	.158	.105	-.135
Delight	.051	.068	-.063	.095
Surprise	.021	.023	-.112	.189
Neutral	.189	.132	.012	-.126

* Correlation is statistically significant at $p < .05$

The scores in Table 1 include means and standard deviations of the test scores and the emotion judgments. The final two columns include Pearson correlations between the emotions judgments and the pre-test and post-test scores. As one would hope, the post-test scores were significantly higher than the pre-test scores, $t(29) = 2.75, p < .05, S_e = .044$. So clearly, learning occurred during the 35 minute interaction with AutoTutor, showing an effect size of 0.50 sigma. Post-test scores were significantly correlated with pre-test scores, $r = .398, p < .05$.

The most frequent emotions were confusion, frustration, flow, and boredom, whereas delight and surprise were comparatively rare. These results were confirmed with a repeated measures ANOVA, $F(6, 174) = 10.951, MSe = .017, p < .001$, and Bonferroni post-hoc tests. This result is compatible with a previous study that collected emotion judgments from 28 college students who interacted with AutoTutor by typing in their contributions (Graesser et al., 2006) rather than giving spoken contributions, as was done in the present study. Therefore, the emotions are similar whether they are experienced with keyboard or spoken input.

Perhaps the most illuminating finding is that post-test scores were significantly predicted by confusion ($r = .490$), but not the other emotion categories. We performed a multiple regression analysis that assessed the extent to which post-test scores are predicted by pre-test scores and confusion. The regression equation was significant, $F(2, 27) = 6.50, p < .05, R^2 = .326$. Confusion was a significant predictor ($p < .05$, two-tailed test, $\beta = .421$), as was also the pre-test scores ($p < .05$, one-tailed test, $\beta = .300$). This significant effect of confusion replicates a previous study by Craig et al. (2004) who had trained judges observe student-AutoTutor interactions and record emotions every 5 minutes. Confusion is a signal that the learner is experiencing cognitive disequilibrium and thinking.

We conducted an analysis on how the emotions were related. The absolute values of the correlations were designated as a measure of emotion similarity. Absolute values (that ignore the sign, + or -) were deemed appropriate because emotional opposites (e.g., flow versus boredom) share dimensions and are hardly equivalent to unrelatedness. A Multiple Dimensional Scaling (MDS) analysis was performed on the similarity matrix of the 7 emotions. A 2-dimensional solution, which was an excellent fit to the data, is shown in Figure 2. The spatial distribution of emotions in the MDS analysis has an interpretable pattern that may have some correspondence to Barrett's (2006) valence-intensity model, although it is hardly definitive whether the dimensions are equivalent. A *valence* dimension is on the X-axis, with positive, pleasant emotions on the left, and negative, unpleasant emotions on the right. An *activity* (or arousal) dimension is on the Y-axis, with more facial activities and physiological arousal on the top and less facial activity and physiological arousal on the bottom. The delight emotion seems to be an outlier, near neutral, but it should be acknowledged that the incidence of delight was

infrequent and perhaps unstable.

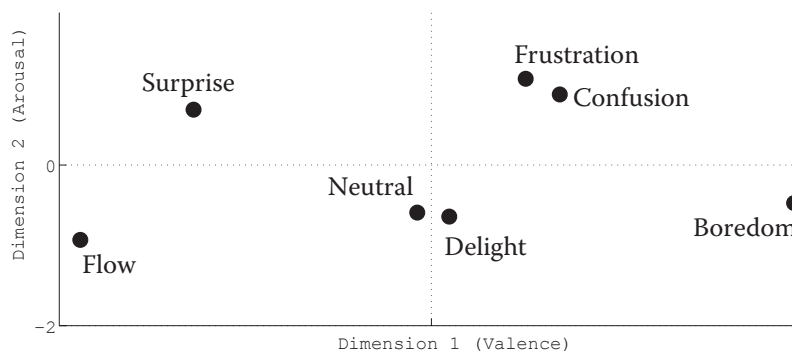


Figure 2. MDS of the Seven Affective States

4. Discussion

This is yet another study that substantiates the importance of confusion (perplexity) in complex learning (Craig et al., 2004; Graesser et al., 2005; Guhe, Gray, Schoelles, & Ji, 2004). When the learner is confused, they are in the state of cognitive disequilibrium, heightened physiological arousal, and more intense thought. Confusion has pronounced manifestations on the face (with activities in Ekman's facial action units of brow lowerer, lid tightener, and lip corner puller) and the learner's posture (D'Mello et al., in press). In contrast, post-test scores were not significantly predicted by the other emotions (flow, boredom, frustration, delight, surprise) that were retrospectively identified by the learners when they viewed a recording of their learning experience. These other emotions presumably play a more prominent role in other learning environments and other populations of learners. More fine-grained analyses that operate at the item level instead of the subject level (like the ones presented here) will be conducted in order to explore relationships between the other affective states and learning. However, as it stands, it is confusion that reigns supreme when college students learn about computer literacy with AutoTutor.

The fundamental assumption behind our meritorious claims related to the increased levels of confusion having a positive impact on learning is that the learners' confusion is caused by a lack of understanding and an abundance of misconceptions with respect to the tutoring content (i.e. topics in Computer Literacy). However, critics might rightly challenge this assumption by stating that other external factors may be the underlying source of the confusion. Perhaps the most prominent source of confusion is inaccuracies associated with the automatic speech recognition (ASR) process, i.e. speech-to-text translation of the learner's contributions. Impoverished speech recognition quality implies that AutoTutor does not fully comprehend the learner's responses, which would result in increases levels of confusion and frustration. We directly tested this hypothesis in a follow-up study in which we measured error rates of the ASR system by manually transcribing a sample of the learner's spoken contributions (D'Mello et al., in review). Correlational analyses indicated that there were no statistically significant relationships between ASR errors and the proportion of confusion experienced, thereby alleviating ASR errors as a potential source of learner confusion.

Other potential sources of confusion are the quality of text to speech synthesis produced by the AutoTutor agent (see Figure 1), the conversational smoothness of AutoTutor's dialogue moves, and other issues with the graphical user interface of the program (see Figure 1). However, unlike ASR errors, which vary by each participant, these factors are constant among all participants. Furthermore, the quality of these components has been verified by several quality tests during the process of implementing the AutoTutor system. However, it is still likely that speech synthesis errors and conversational inadequacies were more prominent when learner's experienced confusion as opposed to the other emotions. Therefore, we plan on conducting a follow-up study which would involve measuring the quality of speech produced by the agent and the conversational smoothness of AutoTutor's dialogue during intervals where confusion was reported. We would then contrast these measures with instances where the learners' declared their emotional state as being neutral.

The present study has also explored how the emotions cluster when the learners make decisions on the emotions during their learning experience. A 2-dimensional valence-activity model appears to account for the clustering (see Figure 2). Analogous 2-dimensional models have been proposed by Kort et al. (2001) in their model of emotions during learning and by Barrett (2006) and Russell (2003) in their theories of emotion that extend beyond learning per se. However, the dimensions of our 2-dimensional MDS solution are probably not equivalent to those advocated by the other researchers, so it is appropriate to hedge in pointing out such correspondences.

We acknowledge that this study is not without limitations. First, learners' self reports were included as the

only measure of their affective state. Since affect is a conceptual variable, or construct, current standards of establishing construct validity involve correlating measurements of affect from multiple methods. Therefore, in order to establish construct validity in measuring the emotions of the learner we are in the process of repeating the aforementioned affect judgment procedure with additional judges that have been trained on affect measurement with Ekman's Facial Action Coding System. Ideally the judges should also have some experience with the AutoTutor learning environment so that they can include contextual knowledge with facial expressions in making their affect judgments. Another limitation of the current study was that the self judgments of emotions were conducted as an offline measure, after the interaction session with AutoTutor. This approach can be contrasted with online measures of learner affect that occur during the tutoring session. We have conducted one study that utilized an emote-aloud procedure where participants verbalized their emotions at the time of experience during interactions with AutoTutor (D'Mello et al., 2006). Here we resorted to an offline measure due to several documented limitations with the emote-aloud protocol as well as in order to preserve internal validity for the follow up study which involves trained judges making offline ratings of the learner's affect. Simply put, in order to make unbiased comparisons between the self reported affect judgments with those provided by trained judges, the judgment procedure needs to be identical.

Our next step is to build an emotion-sensitive AutoTutor that will promote both learning gains and more engagement in the learner. An automated emotion classifier is necessary for AutoTutor to be responsive to learner emotions. We have previously reported some studies that collect the dialogue history, facial action units, position of their body, and other sensory channels of students as they learn and emote aloud (D'Mello et al., 2006). There are systematic relations between these sensing channels and particular emotions. For example, verbalized emotions are predicted by (a) the occurrence of AutoTutor's feedback, (b) relations to the type of feedback (positive, neutral, negative), (c) the directness of AutoTutor's dialogue moves (e.g., hints are less direct than assertions), and (d) the quality of learner's contributions (D'Mello et al., 2006). Regarding the nonverbal channels, emotions are correlated with particular facial expressions (Ekman, 2003; El Kaliouby & Robinson, 2005), posture (D'Mello et al., in press; Mota & Picard, 2003), and face-posture combinations. When speech is recorded, affective states may be induced from a combination of lexical, acoustical, and prosodic features (Litman & Forbes-Riley, 2004). The features from the various modalities can be detected in real time automatically on computers, so we are currently integrating these technologies with AutoTutor.

AutoTutor should have different strategies and dialogue moves when the learner is confused, frustrated, bored, versus experiencing flow – the four most frequent emotions we have found in our work with AutoTutor. Since confusion is tightly linked to learning, it will be important to have the dialogue moves manage the learner's confusion productively. Some learners tend to give up when they are confused because they attribute their confusion to having the trait of low ability (Dweck, 2002; Meyer & Turner, 2006); these learners need to be encouraged and also informed that working on the problem will be fruitful and that confusion is a sign of thoughtful progress. Other learners become motivated when they are confused because it is a signal that they are being challenged and they have confidence in their ability to conquer the challenge. AutoTutor's mechanisms will need to be sensitive to cognitive and motivational characteristics of the learners in addition to their emotional state.

The emotion-sensitive AutoTutor will automatically detect the learner's emotions by verbal and nonverbal channels of communication. We have already designed a computational architecture and algorithms to automatically sense the four major emotions (confusion, frustration, boredom, and flow) on the basis of the dialogue history, facial expressions, and body posture (D'Mello et al., in press; Graesser et al., in press). Whether this new emotion-sensitive AutoTutor will help learning awaits future research.

Acknowledgements

This research was supported by the National Science Foundation (ITR 0325428) and the Institute of Education Sciences (R305H050169). Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF and IES.

References

1. Barrett, L. F. (2006). Emotions as natural kinds? *Perspectives on Psychological Science, 1*, 28-58.
2. Burleson, W., & Picard, R. W. (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems: Workshop on social and emotional intelligence in learning environments*. Maceio-Alagoas, Brazil: ITS.

3. Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29, 241-250.
4. D'Mello, S. K., King, B., Stolarski, M., Chipman, P., & Graesser, A. (in review). The Effects of Speech Recognition Errors on Learner's Contributions Knowledge, Emotions, and Interaction Experience. *Workshop on Speech and Language Technology in Education (SlaTE2007)*.
5. D'Mello, S. K., Picard, R., & Graesser, A. C. (in press). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*.
6. D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16, 3-28.
7. Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
8. Ekman, P. (2003). *Emotions revealed*. New York: Times Books.
9. El Kaliouby, R., & Robinson, P. (2005). Generalization of a vision-based computational model of mind-reading. In J. Tao, T. Tan, & R.W. Picard (Eds.), *Proceedings of Affective Computing and Intelligent Actions* (pp. 582-589). Berlin: Springer.
10. Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48, 612-618.
11. Graesser, A. C., Jackson, G. T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology*, 47, 19-22.
12. 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.
13. 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.
14. Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., Gholson, B. (2006). Detection of emotions during learning with AutoTutor. In R. Son (Ed.), *Proceedings of the 28th Annual Meetings of the Cognitive Science Society* (pp. 285-290). Mahwah, NJ: Erlbaum.
15. Guhe, M., Gray, W. D., Schoelles, M. J., & Ji, Q. (2004). Towards an affective cognitive architecture. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1565). Hillsdale, NJ: Erlbaum.
16. Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy - building a learning companion. In T. Okamoto, R. Hartley, Kinshuk, & J. P. Klus (Eds.), *Proceedings of the IEEE International Conference on Advanced Learning Technology: Issues, Achievements, and Challenges* (pp.43-48). Madison, WI: IEEE Computer Society.
17. Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257-307). San Diego, CA: Academic Press.
18. Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics* (pp. 352-359). East Stroudsburg, PA: Association for Computational Linguistics.
19. Meyer, D. K., & Turner, J. C. (2006). Reconceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18, 377-390.
20. Mota, S., & Picard, R. W. (2003, June). *Automated posture analysis for detecting learner's interest level*. Paper presented at the Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI.
21. 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. Gouardères, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems 2002* (pp. 821-830). Berlin: Springer.
22. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.
23. Stein, N. L., & Hernandez, M.W. (in press). Assessing understanding and appraisals during emotional experience: The development and use of the Narcoder. In J. A. Coan & J. J. Allen (Eds.), *Handbook of emotion elicitation and assessment*. New York: Oxford University Press.

24. 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.