Detection of Emotions during Learning with AutoTutor

Art Graesser (a-graesser@memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis. TN 38152 USA

Bethany McDaniel (btmcdanl@memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis. TN 38152 USA

Patrick Chipman (pchipman@memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis. TN 38152 USA

Amy Witherspoon
(awthrspn@mail.psyc.memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis. TN 38152 USA

Sidney D'Mello (sdmello@memphis.edu)
Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152 USA

Barry Gholson (b.gholson@memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis. TN 38152 USA

Abstract

The relationship between emotions and learning was investigated by tracking the affective states that college students experienced while interacting with AutoTutor, an intelligent tutoring system with conversational dialogue. An emotionally responsive tutor would presumably facilitate learning, but this would only occur if learner emotions can be accurately identified. After a learning session with AutoTutor, the affective states of the learner were classified by the learner, a peer, and judges trained on Ekman's Facial Action Coding system. The classification of the trained judges was more reliable and matched the learners much better than the low scores of untrained peers. This result suggests that peer tutors may be limited in detecting the affective states of peer learners. Classification accuracy was poor at constant intervals of polling (every 20 seconds) but much higher when individuals declared that an affect state had been experienced.

Keywords: Emotion; Instruction and teaching; Human computer interaction; AutoTutor; Affective states

Introduction

Connections between complex learning and emotions have received increasing attention in the fields of psychology (Carver, 2004; Deci & Ryan, 2002; Dweck, 2002), education (Lepper & Henderlong, 2000; Linnenbrink & Pintrich, 2004; Meyer & Turner, 2002), neuroscience (Damasio, 2003), and computer science (Kort, Reilly, & Picard, 2001; Picard, 1997). A deep understanding of such affect-learning connections is needed in order to design engaging educational artifacts that range from responsive intelligent tutoring systems on technical material (DeVicente & Pain, 2002; Graesser, Person, Lu, Jeon, & McDaniel, 2005; Guhe, Gray, Schoelles, & Ji, 2004; Litman & Silliman, 2004) to entertaining media and games (Conati, 2002; Gee, 2003; Vorderer, 2003).

There have been several theories that link cognition and affect very generally (Bower, 1981; Mandler, 1984; Ortony, Clore, & Collins, 1988; Russell, 2003; Stein & Levine, 1991). While these theories convey general links between cognition and emotions, they do not directly explain and predict the sort of emotions that occur during complex learning, such as attempts to master physics, biology, or computer literacy. Researchers in many different fields are familiar with Ekman's work on the detection of emotions from facial expressions (Ekman & Friesen, 1978). However, the emotions that Ekman intensely investigated (e.g., sadness, happiness, anger, fear, disgust, surprise) have minimal relevance to learning per se (Kort et al., 2001). The pervasive affective states during complex learning include confusion, frustration, boredom, flow/engagement, interest, and being stuck (Craig, Graesser, Sullins, & Gholson, 2004; Csikszentmihalyi, 1990).

There are a number of ways in which tutors (and other types of learning environments) might adaptively respond to the learner's affective states in the course of enhancing learning (D'Mello, Craig, Sullins, & Graesser, in press; Graesser et al., 2005; Lepper & Woolverton, 2002). If the learner is frustrated, for example, the tutor can give hints to advance the learner in constructing knowledge or can make supportive empathetic comments to enhance motivation. If the learner is bored, the tutor needs to present more engaging or challenging problems for the learner to work on. The tutor would probably want to lay low and stay out the learner's way 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 flow experience is believed to occur when the learning rate is high and the learner has achieved a high level of mastery at the *region of proximal learning* (Metcalfe & Kornell, 2005).

The affective state of confusion is particularly interesting because it is believed to play an important role in learning (Graesser et al., 2005; Guhe et al., 2004) and has a large correlation with learning gains (Craig et al., 2004). 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, Pye-Cooper, & Whitten, 2005; Graesser & Olde, 2003; Piaget,

1952). Cognitive equilibrium is restored after thought, reflection, problem solving and other effortful cognitive activities. When the learner is confused, there might be a variety of paths for the tutor to pursue. The tutor might want to allow the learner to continue being confused during the cognitive disequilibrium (and the affiliated increased physiological arousal that accompanies all affective states); the learner's self-regulated thoughts might hopefully restore equilibrium. Alternatively, after some period of time waiting for the learner to progress, the tutor might give indirect hints to nudge the learner into more productive trajectories of thought.

Goleman (1995) stated in his book, *Emotional Intelligence*, that expert teachers are able to recognize a student's emotional state and respond in an appropriate manner that has a positive impact on the learning process. This is an important claim, but would be seriously limited if teachers are unable to detect the affective states of the learner. One important question needs to be addressed by all theoretical frameworks and pedagogical practices that relate emotions and learning: How are affective states detected and classified? That is, how are the emotions recognized by tutors, peers, automated sensing devices, and the learners themselves? This question motivated the present study.

One preliminary step in answering the fundamental question of how affective states are classified is to investigate a simple measurement question: How reliably can emotions be classified by the learners themselves versus peers versus trained judges (experts)? An emotionally sensitive learning environment, whether it be human or computer, requires some degree of accuracy in classifying the learners' affect states. The emotion classifier need not be perfect, but it must have some modicum of accuracy. The present study tracked the affective states that college students experience while interacting with AutoTutor, an intelligent tutoring system that helps students learn by holding a conversation in natural language (Graesser, Chipman, Haynes, & Olney, 2005; Graesser et al., 2004; Graesser, Person, & Harter, 2001; VanLehn et al., in press).

AutoTutor was designed to simulate human tutors while it converses with students in natural language. 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 usually only one word to two sentences in length. Therefore, AutoTutor uses a series of pumps ("What else?", "uh huh"), prompts for the learner to express a specific word, hints, assertions, and feedback to elicit responses from the learner that lead to a complete answer of the question. Before the learner is able to give AutoTutor a paragraph of correct information, there can be between 30 to 200 student and tutor turns, about the length of a dialogue with a human tutor.

The present study investigated the extent to which trained judges and untrained peers can accurately identify the affective states of learners who interact with AutoTutor. This immediate objective feeds into the long-term goal of building a version of AutoTutor that identifies and responds adaptively to the affective states of the learner. AutoTutor will never be able to adapt to the learner's emotions if it cannot detect the learner's emotions. Peer tutors and expert tutors similarly will be unable to adapt to the learner's emotions if they cannot identify such affective states.

Methods

Participants

The participants were 28 undergraduates at the University of Memphis who participated for extra course credit.

Materials and Procedure

The experiment was divided into two sessions. Session 1 took two hours and consisted of a pretest, interaction with AutoTutor, a posttest, and judgments of emotions they experienced while interacting with AutoTutor (*self* judgments, see below). Session 2 lasted one hour and consisted of judgments of the emotions of a peer while the peer had interacted with AutoTutor *(peer* judgments).

AutoTutor Participants interacted with AutoTutor for 32 minutes on one of three randomly assigned topics in computer literacy: hardware, Internet, or operating systems (see Graesser et al., 2001 for detailed information about AutoTutor). Each of these topics had 12 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?*. Although each question required 3-7 sentences in an ideal answer, learners rarely give the complete answer in a single turn. A conversation occurs with multiple turns that take a few minutes. A typical learner could only answer 3-5 questions within the allotted 32 minutes.

The AutoTutor interface had 4 windows, as shown in Figure 1. Window 1 (top of screen) was the main question that stayed on the computer screen throughout the conversation that involved answering the question. Window 2 (bottom of screen) was affiliated with the learner's answer in any one turn and echoed whatever the learner typed in via keyboard. Window 3 (left middle) was an animated conversational agent that spoke the content of AutoTutor's turns. The talking head had facial expressions and some rudimentary gestures. Window 4 (right middle) was either blank or had auxiliary diagrams.

Figure 1: Interface of AutoTutor
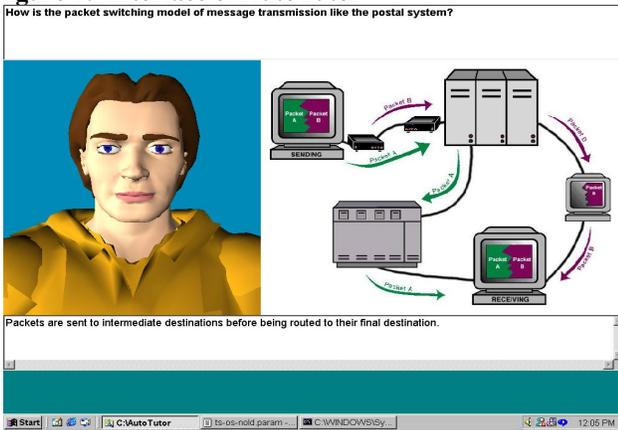
Each turn of AutoTutor in the conversational dialogue had three information slots (i.e., units, constituents). The first slot of most turns was feedback on the quality of the learner's last turn. This feedback was either positive (*very good, yeah*), neutral (*uh huh, I see*), or negative (*not quite, not really*). The second slot advanced the conversation 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. Discourse markers (*and also, okay, well*) connected the utterances of these three slots of information. The conversations managed by AutoTutor are sufficiently smooth that learners can get through the session with minimal difficulties.

Judging Affective States Four sets of emotion judgments were made for the observed affective states of each AutoTutor session. First, for the *self* judgments, the learner watched his or her own session with AutoTutor immediately after having interacted with AutoTutor. Second, for the *peer* judgments, each learner came back a week later to watch and judge another learner's session on the same topic in computer literacy. Finally, there were two trained judges: undergraduate research assistants who were trained extensively on tutorial dialogue characteristics and how to detect facial action units according to Paul Ekman's Facial Action Coding System (Ekman & Friesen, 1978). The two trained judges judged all sessions separately.

A list of the affective states and definitions was provided for the learners, peers, and two trained judges. The states were boredom, confusion, flow, frustration, delight, neutral and surprise, the emotions that were most frequently experienced in a previous study of AutoTutor (Craig et al., 2004). 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.

The judgments for a learner's tutoring session proceeded by playing a video of the face along with a screen capture video of interactions with AutoTutor. Judges were instructed to make judgments on what affective states were present in each 20-second interval at which the video automatically stopped. There was a checklist of emotions for them to mark, along with an "other" category for them to provide additional emotions that they viewed as relevant. They were also instructed to indicate any affective states that were present in between the 20-second stops. If the participant was experiencing more than one affective state in a 20-second block, judges were instructed to mark each state and indicate which was most pronounced.

In summary, each video of the tutorial interaction was judged by the self (the learner), a peer (another learner), and two trained judges.

Results

Interjudge reliability in judging emotions was computed using Cohen's kappa for all possible pairs of judges: self, peer, trained judge1, and trained judge2. Altogether, there were six possible pairs (see Table 1). The reliability scores were based on the first-choice affect state the learner gave. The observations included those judgments at the 20-second interval polling (approximately 2500 observations) and those in-between observations in which learners stated that they had an emotion in between two successive pollings (between 78 and 180 observations for each of the 6 pairs in Table 1). Cohen's kappa scores were computed separately for each of the 28 learners. Statistical analyses were performed on these kappa scores when comparing agreement of the 6 pairs of judges in Table 1.

The scores in Table 1 revealed that the trained judges had the highest agreement, the self-peer pair had near zero agreement, and the other pairs of judges were in between. An ANOVA was performed on the left column of scores that included *all* observations, namely those at fixed 20-second intervals plus those at voluntary timestamps. The results reveal that there were significant differences in kappa scores among the six pairs, $F(5, 135) = 33.34$, $MSe = .008$, $p < .01$. Fisher LSD post hoc tests revealed that the self-peer pair had the lowest inter-judge reliability scores ($p < .05$) when compared to the other five pairs. The two trained judges had significantly higher kappa scores than the other five pairs. These results support the conclusion that peers are not particularly good at detecting learner emotions. Another conclusion is that training on Ekman's facial action coding system and tutorial dialogue can enhance the reliability and accuracy of judgments of affective states.

Table 1: Kappa scores for judgments of affective states at all points, 20-second intervals, and voluntary timestamps.

Pair of Judges	All	20-second	Voluntary
Self/Peer	0.08	0.06	0.12
Self/Judge1	0.14	0.11	0.31
Self/Judge2	0.16	0.13	0.24
Peer/Judge1	0.14	0.11	0.36
Peer/Judge2	0.18	0.15	0.37
Judge1/Judge2	0.36	0.31	0.71

Further analyses were performed after segregating judgments that were made at the regularly polled timestamps (every 20 seconds) and those which were made at voluntary timestamps in between the automatic 20-second stop points. For example, if a judge made a judgment at 4 minutes, when the video playback automatically paused, that judgment would be in the "regularly polled sample" group. If the same judge manually paused the video and made a judgment at 4 minutes and 16 seconds, that particular judgment would be in the "voluntary" judgment sample. There were substantially fewer observations in the voluntary sample than the regularly polled sample because judges were not required to stop and make judgments in between. The voluntary sample presumably had more salient affective states than the regularly polled sample, so agreement should be higher.

The inter-judge reliability increased considerably for all pairs of judges when computed only on those observations that were voluntary judgments. The highest inter-judge score was between the trained judges (kappa = 0.71) whereas the lowest was between the self and peer (kappa = 0.12). The kappa for self and peer did increase for voluntary timestamps, but the voluntary kappa for self and peer was not appreciably above the kappa for all judgments (.12 versus .08). When considering only those judgments made at the 20-second interval stops, inter-judge reliability was substantially lower and closely corresponded to the kappa scores for all judgments.

The judgments made in the voluntary sample involved more animated emotions (and theoretically higher physiological arousal) compared to the more subtle emotions at the 20-second intervals. An analysis was performed on the proportions of emotion categories at the 20-second intervals. We examined the proportion of judgments that were made for each of the affect categories, averaging over the 4 judges. The most common affective state was neutral (.369), followed by confusion (.212), flow (.188), and boredom (.167); the remaining states of delight, frustration and surprise totaled .065 of the observations. The more salient voluntary points had a rather different distribution. The most prominent affect state was confusion (.377), followed by delight (.192) and frustration (.191), whereas the remaining affective states comprised .240 of the observations (boredom, surprise, flow, and neutral, in descending order). Most of the time learners are either in a neutral state or in a subtle affective state (boredom or flow).

Discussion

An emotion-sensitive AutoTutor would presumably promote both learning gains and more engagement in the learner. AutoTutor should have different strategies and dialogue moves when the learner is confused or frustrated than when the learner is bored. However, both human and automated tutors can be emotionally adaptive only if the emotions of the learner can be detected. The accuracy of the detection need not be perfect, but it should be approximately on target.

The results of this study support a number of conclusions about emotion detection. First, trained judges who are experienced in coding facial actions and tutorial dialogue provide affective judgments that are more reliable and that match the learner's self reports better than the judgments of untrained peers. Second, the judgments by peers have very little correspondence to the self reports of learners. Peers apparently are not good judges of the emotions of learners. Third, an emotion labeling task is more difficult if judges are asked to make emotion judgments at regularly polled timestamps, rather than being able to stop a video display to make spontaneous judgments. The states at regular timestamps are much less salient so there is minimal information for judges to base their judgments, compared with those points when affective states are voluntarily spotted. Training on facial expressions makes judges more mindful of relevant facial features and transient facial movements, but judges can do this only if the expressions have enough information to fortify these judgments.

Many advocates of peer tutoring have extolled the virtues of having peers tutor each other. One potential advantage of peer tutoring is that there is no appreciable status difference between peers, compared to when a teacher tutors a student or an older tutor helps a younger learner (Rogoff, 1990). The results of the present study suggest, however, that there may be a drawback of peer tutoring. Peer tutors apparently are not very good at classifying emotions of learners. It takes expertise in tutoring or emotion detection before accurate detection of learner emotions can be achieved. This requirement of expertise is apparently quite important because, according to Lepper and Woolverton (2002), roughly half of expert tutors' interactions with the student are focused on affective elements. Our trained judges were simply trained on Ekman's facial action coding system and characteristics of tutorial dialogue. We are uncertain at this point whether it is the detection of facial expressions that is important in tutoring or a seasoned experience with domain knowledge and pedagogy. Future research is needed to resolve this.

It is unclear what exactly should be the gold standard for deciding what emotions a learner is truly having. Should it be the learner or the expert? We are uncertain about

the answer to this question, but it is conceivable that some emotions may best be classified by learners and others by experts. Perhaps a composite score that considers both viewpoints would be most defensible.

Whatever the gold standard might be, there is the challenge of identifying what sensing devices and automated affect classifiers we should integrate with AutoTutor. An automated affect classifier is of course needed to make AutoTutor responsive to learner emotions. We have previously reported some studies that collected verbal expressions of emotions (an *emote-aloud* protocol) from college students while interacting with AutoTutor. These learners say out loud whatever emotions come to mind while interacting with the system. We have simultaneously recorded the dialogue history and facial action units while they learn and emote aloud. There are systematic relations between these sensing channels and particular emotions. For example, verbalized emotions are prevalent after AutoTutor's feedback (positive, neutral, negative), the directness of AutoTutor's dialogue moves (hints are less direct than assertions), and the quality of learner's contributions (D'Mello, Craig, Sullins, & Graesser, in press). Particular facial expressions are correlated with particular emotions (D'Mello et al., 2005). Frustration is associated with outer brow raise, inner brow raise, and the dimpler whereas confusion is associated with brow lowerer, lid tightener, and lip corner puller. Posture may be correlated with interest (Mota & Picard, 2003). If we record speech, then affective states may be induced from a combination of lexical, acoustical, and prosodic features (Litman & Forbus-Reilly, 2004). We believe that most of these features from the various modalities can be detected in real time automatically on computers. Whether an automated affect detector can be achieved awaits future research and technological development.

Acknowledgements

References

Bower. G.H. (1981). Mood and memory. *American Psychologist, 36,* 129-148.

Carver, C. S. (2004). Negative affects deriving from the behavioural approach system. *Emotion, 4,* 3-22.

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

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.

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

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Orlando, FL:Harcourt.

De Vicente, A., & Pain, H. (2002). Informing the detection of students' motivational state : An empirical study. In S.A. Cerri, G. Gouarderes, and F. Paraguacu (Eds *Proceedings of the sixth international conference on intelligent tutoring systems* (pp.933-943). Berlin, Germany: Springer.

Deci, E. L., & Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R.,& Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International conference on Intelligent User Interfaces* (pp. 7-13) New York: AMC Press.

D'Mello, S.K., Craig, S.D., & Graesser, A.C. (in press). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*.

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.

Ekman, P, & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

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

Gee, J.P. (2003). *What video games have to teach us about language and literacy*. New York: Macmillan.

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.

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. & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95,* 524–536.

Graesser, A. C., Person N., Harter, D. & The Tutoring Research Group (2001) Teaching tactics and dialog in AutoTutor, *International Journal of Artificial Intelligence in Education, 12,* 257–279.

Graesser, A.C., Person, N., Lu, Z., Jeon, M.G., & McDaniel, B. (2005). Learning while holding a conversation with a computer. In L. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 143-167). Greenwich, CT: Information Age Publishing.

Goleman, D. (1995). *Emotional intelligence* (New York, Bantam Books).

Guhe, M., Gray, W. D., Schoelles, M. J., & Ji, Q. (2004, July). *Towards an affective cognitive architecture.* Poster session presented at the Cognitive Science Conference, Chicago, IL.

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 IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges* (pp.43-48). Madison, Wisconsin: IEEE Computer Society.

Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257-307). San Diego, CA: Academic Press.

Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.

Linnenbrink, E. A., & Pintrich, P. A. (2004). Role of affect in change processing in academic contexts. In D. Y. Dai, & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 57-88). Mahwah, New Jersey: Lawrence Erlbaum Associates.

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.

Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. In *proceedings of the human language technology conference: 3rd meeting of the North American chapter of the association of computational linguistics* (pp. 52-54). Edmonton, Canada: ACL.

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York: Norton.

Metcalfe, J., & Kornell, N. (2005). A region or proximal of learning model of study time allocation. *Journal of Memory and Language, 52,* 463-477.

Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist, 37,* 107-114.

Mota, S. & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. *Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI,* June, 2003.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. New York: Cambridge University Press.

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

Picard, R. W. (1997). *Affective computing*. Cambridge, Mass: MIT Press.

Rogoff, B. (1990). *Apprenticeship in thinking.* New York: Oxford University Press.

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

Stein, N. L., & Levine, L. J. (1991). Making sense out of emotion. In W. Kessen, A. Ortony, & F. Kraik (Eds.), *Memories, thoughts, and emotions: Essays in honor of George Mandler* (pp. 295-322). Hillsdale, NJ: Erlbaum.

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

Vorderer, P. (2003). Entertainment theory. In B. Jennings, D. Roskos-Ewoldsen, & J. Cantor (Eds.), *Communication and emotion: Essays in honor of Dolf Zillmann* (pp. 131-153). Mahwah, NJ: Erlbaum.