Malleability of Students' Perceptions of an Affect-Sensitive Tutor and its Influence on Learning

Sidney D'Mello[1] and Art Graesser[2]
[1]University of Notre Dame, Notre Dame, IN 46556, USA
sdmello@nd.edu
[2]University of Memphis, Memphis, TN 38152, USA
graesser@memphis.edu

Abstract

We evaluated an affect-sensitive version of AutoTutor, a dialogue based ITS that simulates human tutors. While the original AutoTutor is sensitive to students' cognitive states, the affect-sensitive tutor (Supportive tutor) also responds to students' affective states (boredom, confusion, and frustration) with empathetic, encouraging, and motivational dialogue moves that are accompanied by appropriate emotional expressions. We conducted an experiment that compared the Supportive and Regular (non-affective) tutors over two 30-minute learning sessions with respect to perceived effectiveness, fidelity of cognitive and emotional feedback, engagement, and enjoyment. The results indicated that, irrespective of tutor, students' ratings of engagement, enjoyment, and perceived learning decreased across sessions, but these ratings were not correlated with actual learning gains. In contrast, students' perceptions of how closely the computer tutors resembled human tutors increased across learning sessions, was related to the quality of tutor feedback, the increase was greater for the Supportive tutor, and was a powerful predictor of learning. Implications of our findings for the design of affect-sensitive ITSs are discussed.

Introduction

It is widely acknowledged that emotions play a critical role during learning. Emotions such as boredom, confusion, frustration, and anxiety naturally arise during learning and are dependent upon the learning context (e.g., preparing for a high-stakes test vs. solving homework problems), the learning task (e.g., reading comprehension, problem solving), the student's mood at the start of the session (e.g., excited or lethargic), and a host of individual differences in motivation, goal orientation, self-efficacy, prior knowledge, and ability (Calvo & D'Mello, 2011; Meyer & Turner, 2006; Pekrun, in press). Importantly, student emotions are more than mere by-products of learning and are predictive of a number of learning outcomes including learning gains, self-efficacy, interest in educational activities, attrition, and dropout (Daniels, et al., 2009; Schutz & Pekrun, 2007).

The recent emergence of research documenting the importance of affect during learning has important implications for Intelligent Tutoring Systems (ITSs). ITSs have proven to be extremely effective in promoting learning gains to the extent that some ITSs are almost as effective as accomplished human tutors (Corbett, 2001; VanLehn, et al., 2007). However, there is still room for improvement. ITSs have come a long way towards modeling and responding to students' cognitive states, but the link between emotions and learning suggests that they should be affective processors as well (Issroff & del Soldato, 1996). An affect-sensitive ITS would incorporate assessments of the students' cognitive and affective states into its pedagogical and motivational strategies in order to keep students engaged, boost self-confidence, and maximize learning.

The last decade has witnessed an impressive array of research focusing on building fully-automated affect-sensitive ITSs (Conati & Maclaren, 2009; D'Mello, et al., 2010; Forbes-Riley & Litman, 2009; Sabourin, Mott, & Lester, 2011; Woolf, et al., 2009). The recent emergence of some functional affect-sensitive ITSs begs the question of whether these systems live up to their promises of (1) increasing learning gains over traditional ITSs and (2) delivering more usable, engaging, and enjoyable learning experiences.

We have made some progress towards answering these questions within the context of a recently developed affect-sensitive version of AutoTutor, an ITS with conversational dialogues (Graesser, Chipman, Haynes, & Olney, 2005). The original AutoTutor has a set of fuzzy production rules that are sensitive to the cognitive, but not to the affective states of the learner. The affect-sensitive AutoTutor (called the Supportive AutoTutor) has a set of production rules that map dynamic assessments of learners' cognitive and affective states with tutor actions to address the presence of boredom, confusion, and frustration with empathetic, encouraging, and motivational dialogue moves and emotional displays (D'Mello, Craig, Fike, & Graesser, 2009). The working hypothesis is that the Supportive AutoTutor will yield more enjoyable, engaging, and effective (in terms of learning) interactions than the non-affective tutor (called the Regular AutoTutor).

The question of whether affect-sensitivity improves learning (Question 1 from above) was tested in a previous study (D'Mello, et al., 2010) where 84 students completed two 30-minute training sessions with either the Regular AutoTutor or the Supportive AutoTutor. The results indicated that the Supportive tutor helped learning for low-domain knowledge students during the second 30-minute learning session. There was also a case where the Supportive tutor was less effective at promoting learning, particularly for high-domain knowledge students during the first 30-minute session. Importantly, learning gains increased from Session 1 to Session 2 with the Supportive tutor whereas they plateaued with the Regular tutor.

In addition to learning gains there is the important question of whether affect-sensitivity yields more engaging and enjoyable learning experiences (Question 2 from above). This is a critical question because a student's impressions of a learning technology have important consequences on task persistence and long-term engagement. Although data on students' subjective evaluations of each 30-minute learning session was collected in the previous study (D'Mello, et al., 2010), this data was not analyzed and reported. The present paper focuses on comparing the Supportive and Regular tutors along the dimensions of perceived effectiveness, fidelity of cognitive and emotional feedback, engagement, and enjoyment.

A distinctive goal of the paper is to analyze changes in students' perceptions of the tutor. This is an important issue because students' impressions of a learning technology are expected to be highly malleable rather than rigid. For example, a student might initially consider the Supportive AutoTutor to be novel and enjoyable but these positive impressions might quickly fade over a prolonged learning session as the novelty wears off. Alternatively, a student might initially find the empathetic and encouraging responses of the Supportive tutor to be unexpected and strange, but these impressions might change as the student forges a bond with the tutor. The present paper addresses these contrasting positions by how students' impressions of the tutors changes over a 60-minute learning session.

Versions of AutoTutor

Regular AutoTutor

AutoTutor is a dialogue-based ITS for Newtonian physics, computer literacy, and critical thinking (Graesser, et al., 2005). AutoTutor's dialogues are organized around difficult questions and problems (called main questions) that require reasoning and explanations in the answers. When presented with these questions, students typically respond with answers that are only one word to two sentences in length. In order to guide students in their construction of an improved answer, AutoTutor actively monitors students' knowledge states and engages them in a turn-based dialogue. AutoTutor adaptively manages the dialogue by providing feedback on the student's answers (e.g. "good job", "not quite"), pumping the learner for more information (e.g. "What else"), giving hints (e.g. "What about X"), prompts (e.g. "X is a type of what "), correcting misconceptions, answering questions, and summarizing topics. AutoTutor's dialogue moves are delivered by an animated pedagogical agent.

Learning gains produced by AutoTutor have ranged from 0.4-1.5 sigma (a mean of 0.8), depending on the learning measure, the comparison condition, the subject matter, and the version of AutoTutor (Graesser, et al., 2005; VanLehn, et al., 2007). A 1 sigma effect size is approximately a one letter grade increase in learning.

Supportive AutoTutor

The Supportive tutor focused on detecting and responding to boredom, frustration, and confusion, which were the major emotions observed during interactions with AutoTutor (D'Mello & Graesser, in press)..

Detecting Affect. The affect detection system monitors conversational cues, gross body language, and facial features to detect boredom, confusion, frustration, and neutral (no affect) (see Figure 1). Automated systems that detect these emotions have been integrated into AutoTutor and have been extensively described and evaluated in previous publications (D'Mello & Graesser, 2010).

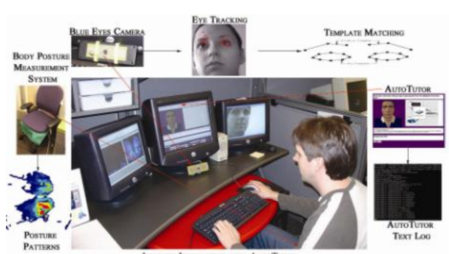

Figure 1. Affect sensing during learning

Responding to Affect. The affect-sensitive production rules that guide the tutor's responses to sensed negative affect are motivated by attribution theory (Weiner, 1986), cognitive disequilibrium during learning (Graesser & Olde, 2003), politeness (Brown & Levinson, 1987), and empathy (Lepper & Chabay, 1988). In addition to theoretical considerations, the assistance of experts in tutoring was enlisted to help create the set of production rules.

The production rules were designed to map dynamic assessments of the students' cognitive and affective states with appropriate tutor actions. An emotion generator was also needed for the Supportive AutoTutor because the system was expected to respond with suitable emotions. Therefore, the agent needed to speak with intonation that was properly integrated with facial expressions that displayed emotions. For example, an enthusiastic nod accompanied positive feedback after the student provided a correct response. There was a shaking of the head when the student response was low quality and a skeptical look when the tutor detected that the student was hedging (see Figure 2). A small set of emotion displays like these examples went a long way in conveying the tutor's emotions.

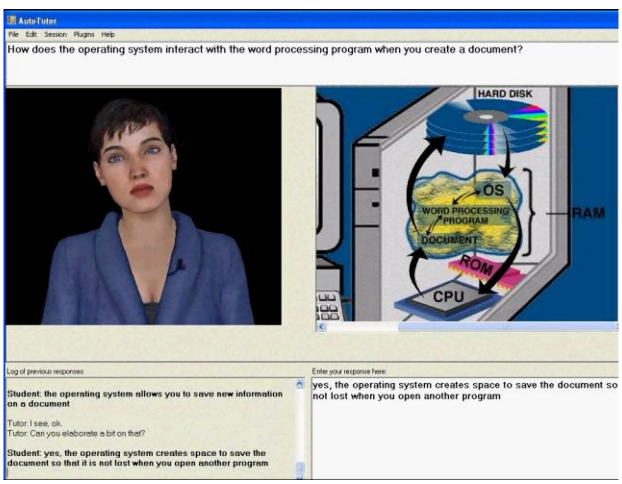

Figure 2. Affect-sensitive AutoTutor

In addition to the emotional feedback, there was also an empathetic expression conveyed in words, facial expressions, and motion when supportive encouragement was needed. As an example, consider a student who has been performing well overall, but the most recent contribution was not very good. If the student's current state is classified as boredom, with a high probability, and the previous state was classified as frustration, then the Supportive AutoTutor might say the following: "Maybe this topic is getting old. I'll help you finish so we can try something new." The exact phrase would be randomly selected from a list of phrases designed to indirectly address the student's boredom and to try to shift the topic before the student becomes disengaged from the learning experience. In this sense, the rules were context sensitive and are dynamically adaptive to each individual learner (see D'Mello et al., 2009 for more information). In contrast, the Regular tutor delivered the content of the feedback without any emotional display and did not make any supportive, empathetic, or encouraging statements.

Method

Participants and Design

84 participants (called students) from a mid-south university in the US participated for course credit. The experiment had a between-subjects design in which students were randomly assigned to either the Regular or the Supportive AutoTutor. Students completed two training sessions with the *same* version of AutoTutor but on two *different* computer literacy topics (hardware, operating systems, the Internet). The order in which topics were covered was counterbalanced across students with a Latin Square.

Content Covered in AutoTutor Sessions

Students completed three challenging computer literacy questions in each tutoring session. Each problem required approximately three to seven sentences of information for a correct answer. The questions required answers that involved inferences and deep reasoning, such as *why, how, what-if, what if not,* and *how is X similar to Y?*. An example question is: "How can John's computer have a virus but still boot to the point where the operating system starts?"

Measures

Knowledge Tests. Students were tested on their knowledge of computer literacy topics both before and after the tutorial session (pretest and posttest, respectively). Each test had 8 questions on each topic, thereby yielding 24 questions in all. The items were designed to assess deep levels of knowledge (e.g., "How does the computer assure that other stored information is not overwritten when a save command is given?") rather than recall of shallow facts (e.g. "What does RAM stand for?").

Post Interaction Questionnaire. Students provided subjective evaluations of the tutors and the tutorial session by completing a 10-item questionnaire after each tutorial session. The questionnaire contained the following items:

1. I felt that I learned new information from AutoTutor
2. Understanding the material was important to me
3. While I was covering the material I tried to make everything fit together
4. AutoTutor showed emotion
5. AutoTutor's emotions were natural

6. The feedback from AutoTutor was appropriate with respect to my progress
7. I felt that my interaction with AutoTutor was comparable to an interaction with a human tutor
8. I felt engaged during the session
9. I enjoyed interacting with AutoTutor
10. I felt that AutoTutor was difficult to use and work with.

Students responded to each item by choosing one of six alternatives: strongly disagree, disagree, somewhat disagree, somewhat agree, agree, and strongly agree.

Procedure

Students were tested individually during a 1.5 to 2 hour session. First, students completed an informed consent, followed by the pretest. Next, the general features of AutoTutor's dialogue and pedagogical strategies were described to the students. On the basis of random assignment, students interacted with either the Supportive or the Regular AutoTutor. They were tutored on one computer literacy topic until three main questions were successfully answered or the 30-minute training period had elapsed (Session 1). They then completed the Post Interaction Questionnaire. Next students interacted with the *same* version of AutoTutor on *another* computer literacy topic until three main questions were successfully answered or the 30-minute training period had elapsed (Session 2). The Post Interaction Questionnaire was also completed after this session. Finally, students completed the posttest and were debriefed.

Results and Discussion

Pretest and posttest scores were computed as the proportion of questions answered correctly. Proportional learning gains were computed as (posttest–pretest)/(1-pretest). Separate proportional scores were computed for Session 1 and Session 2. We also computed proportional learning gains for the topic for which participants received no tutoring to assess testing effects and knowledge transfer (not discussed here).

Students' responses to each of the 10 items on the Post Interaction Questionnaire were assigned a score of 1 (strongly disagree) to 6 (strongly agree). Outliers were identified as values exceeding two standard deviations away from the mean and were removed.

Students' Perceptions after Sessions 1 and 2

Descriptive statistics for students' responses to the 10-items of the Post Interaction Questionnaire for Sessions 1 and 2 are presented in Table 1. Independent samples t-tests comparing each item across tutors yielded only two significant differences (highlighted in bold).

First, students perceived the Supportive AutoTutor as showing more emotions (Item 4) compared to the Regular Tutor for Session 1, $t(82) = 2.96$, $p = .004$, $d = .65$, but not for Session 2 ($p = .476$). Second, students considered the Supportive tutor to more closely resemble a human tutor for Session 2, $t(82) = 2.28$, $p = .025$, $d = .50$, but not for Session 1.

These results suggest that differences among tutors was minimal, some factors changed over time. We performed additional analyses to quantify the extent to which students' perceptions changed over time and whether these perception changes were related to learning gains.

Table 1. Means and standard deviations *M(SD)*

	Session 1		Session 2	
Item	**Reg**	**Sup**	**Reg**	**Sup**
1.	4.4 (1.0)	4.3 (1.1)	2.6 (1.4)	2.4 (1.3)
2.	3.5 (1.3)	3.6 (1.4)	3.2 (1.6)	3.1 (1.5)
3.	4.5 (0.9)	4.6 (0.8)	4.2 (1.4)	4.1 (1.3)
4.	**2.9 (1.4)**	**3.8 (1.5)**	4.3 (0.9)	4.1 (1.2)
5.	2.9 (1.4)	3.2 (1.4)	3.2 (1.5)	3.4 (1.6)
6.	4.1 (1.0)	4.3 (1.1)	4.5 (1.0)	4.3 (1.0)
7.	2.6 (1.2)	2.7 (1.3)	**3.0 (1.5)**	**3.8 (1.5)**
8.	3.2 (1.4)	3.5 (1.3)	2.7 (1.4)	2.6 (1.3)
9.	2.8 (1.4)	3.3 (1.6)	2.7 (1.4)	2.8 (1.3)
10.	3.2 (1.5)	3.0 (1.5)	3.5 (1.6)	3.1 (1.4)

Note. Reg. = Regular AutoTutor; Sup = Supportive AutoTutor

Changes in Students' Perceptions across Sessions

We investigated whether there were changes in students' perceptions by computing delta scores for each item (i.e., Session 2 – Session 1). Positive scores for an item indicate that there was an increase in students' ratings for that particular item while negative scores reflect a decrease.

Comparing delta scores to zero. We performed one-sample t-tests on the delta scores associated with each tutor in order to identify measures that significantly diverged from 0 (a delta score of 0 is indicative of no change). The results were remarkably consistent across tutors. Specifically, the delta scores for three items (1, 8, and 9) were significantly *less* than 0, while delta scores for item 7 was significantly *greater* than zero. These results indicate that students' reported less engaging (item 8), less enjoyable (item 9) and less fruitful interactions (item 1) in Session 2 compared to Session 1. However, students considered the tutors to more closely resemble a human tutor (item 9) after Session 2 compared to Session 1.

Assessing tutor effects on delta scores. We investigated whether there were tutor differences in delta scores with four independent-samples t-tests. There were no significant differences across tutors for delta scores associated with engagement, enjoyment, and learning.

However, delta scores pertaining to resemblance to a human tutor was marginally significantly greater for the Supportive AutoTutor (M = .972, SD = 1.32) compared to the Regular AutoTutor (M = .462, SD = 1.37), $t(77) = 2.03$, $p = .11$, $d = .38$.

Correlations of delta scores with learning. We correlated the delta scores that significantly differed from zero (perceptions of learning, engagement, enjoyment, and resemblance to humans) with proportional learning gains for Session 2. Partial correlations that corrected for prior knowledge (proportion of correct responses on pretest) were separately computed for each tutor. The results indicated that the extent to which the students considered the tutors to resemble a human tutor was correlated with proportional learning gains. The correlation was substantially stronger for the Supportive tutor ($r = .516$, $p = .001$) compared to the Regular tutor ($r = .290$, $p = .082$).

The delta score for engagement was also significantly correlated with learning gains for the Supportive tutor ($r = .372$, $p = .023$) but not for the Regular tutor ($r = -.161$, $p = .342$).

Predicting delta scores. We attempted to identify features of the tutors that could predict changes in perceived resemblance to human tutors. This was accomplished by conducting two multiple linear regressions for the Regular and Supportive tutors. The dependent variable in each regression was the delta score for *resemblance to human*. The independent variables were students' changes in perception of the extent to which the (1) tutor showed emotions, (2) tutor's emotions were natural, (3) tutor's feedback was appropriate, and (4) tutor was difficult to use. A tolerance analysis indicated that there were no multicollinearity problems, so all four variables were considered in the model with a stepwise regression procedure.

Significant models were discovered for both the Regular, $F(1, 36) = 5.74$, $p = .022$, $R^2adj. = .114$, and the Supportive tutor, $F(2, 33) = 6.07$, $p = .006$, $R^2adj. = .225$. Changes in students' perceptions of the quality of tutor feedback were significant positive predictors for both the Regular ($\beta = .371$, $p = .022$) and the Supportive tutor ($\beta = .426$, $p = .009$). Students' change in the perception of the extent to which tutor's showed emotions was a negative predictor for the Supportive tutor ($\beta = -.403$, $p = .013$) but not for the Regular tutor.

General Discussion

We analyzed a novel affect-sensitive ITS and a traditional non-affect-sensitive ITS with respect to (a) expectations of learning, motivation to learn, and effort exerted, (b) perceived difficulty, engagement, and enjoyment, and (c) the quality of the tutor emotions, feedback, and naturalness. Students' impressions of the Supportive AutoTutor were compared to the Regular AutoTutor after each learning session. We also investigated changes in students' perceptions across sessions in an attempt to identify malleable factors that were predictive of learning.

The results were illuminating in a number of respects. With two exceptions, there were no major differences in students' perceptions of the tutors after either session. The interesting patterns emerged when changes in students' perceptions were quantified and analyzed. One finding was that students perceived their second tutorial session with both tutors to be less engaging and less enjoyable. This was an expected finding due to the difficulty of the task and tedium of the learning sessions. Indeed, longer learning sessions have been associated with increased boredom (D'Mello & Graesser, 2010).

Another finding was that students' expectations of learning with both tutors decreased in the second session. This was a surprising finding because proportional learning gains were significantly greater in Session 2 (M = .288, SD = .382) compared to Session 1 (M = .453, SD = .298) for the Supportive tutor, $t(38) = 2.36$, $p = .023$, $d = .48$, but not for the Regular tutor ($p = .946$). Expectations of learning were not correlated with actual learning gains for either tutor, which is consistent with previous research which indicates that students cannot accurately gauge their learning (Glass, Kim, Evens, Michael, & Rovick, 1999; Person, Graesser, Magliano, & Kreuz, 1994).

Perhaps the most interesting finding was that training time had a positive impact on the Supportive tutor's perceived resemblance to human tutors and this resemblance was highly correlated with learning gains. Similar effects were discovered for the Regular tutor, but these effects was more muted. One interpretation of this finding is that learning is positively impacted when students form a social bond with the animated pedagogical agents that embodies a computer tutor (Reeves & Nass, 2003). However, it takes some time for students to bond with the tutor to the extent to which they consider the tutors to resemble humans. Furthermore, this social bond is strengthened when the agents make an effort to mirror some of the pragmatics of human-human communication, such as the display of empathetic and supportive emotional expressions.

The results of the present study have a number of important implications for the design of computer tutors, particularly for ITSs that aspire to model human tutors. The key finding was that learning was positively associated with students' perceptions of how closely the ITSs resembled human tutors. Perceived resemblance to humans was larger for the Supportive AutoTutor that attempted to model the motivational moves of human tutors (Lepper & Chabay, 1988). Providing appropriate feedback with respect to progress is one important factor that positively contributes to students' perceptions of the likeness of ITSs to human tutors. However, it takes some time for the student and the Supportive tutor to interact before students consider the tutor to be more humanlike. Further research with additional measures, tutors, and longer training sessions is needed to fully understand the malleability of students' perceptions of computer tutors and its impact on learning and engagement.

Acknowledgments

We thank our research colleagues in the Emotive Computing Group at the University of Memphis the Affective Computing group at MIT.

This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1108845). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Calvo, R., & D'Mello, S. (Eds.). (2011). *New perspectives on affect and learning technologies*. New York: Springer.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267-303.

Corbett, A. (2001). Cognitive computer tutors: Solving the two-sigma problem. In M. Bauer, P. Gmytrasiewicz & J. Vassileva (Eds.), *Proceedings of 8th International Conference on User Modeling* (pp. 137-147). Berlin / Heidelberg: Springer.

D'Mello, S., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20*(2), 147-187.

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

D'Mello, S., Craig, S., Fike, K., & Graesser, A. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In J. Jacko (Ed.), *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction* (pp. 595-604). Berlin/Heidelberg: Springer.

D'Mello, S., & Graesser, A. (in press). Emotions during Learning with AutoTutor. In P. Durlach & A. Lesgold (Eds.), *Adaptive Technologies for Training and Education*. Cambridge, U.K: Cambridge University Press.

Daniels, L. M., Pekrun, R., Stupnisky, R. H., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology, 101*(4), 948-963.

Forbes-Riley, K., & Litman, D. (2009). Adapting to student uncertainty improves tutoring dialogues. In V. Dimitrova, R. Mizoguchi & B. Du Boulay (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 33-40). Amsterdam: IOS Press.

Glass, M., Kim, J., Evens, M., Michael, J., & Rovick, A. (1999). *Novice vs. expert tutors: A comparison of style.* Paper presented at the Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, IN.

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

Graesser, A., & 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*(3), 524-536.

Issroff, K., & del Soldato, T. (1996). *Incorporating motivation into computer-supported collaborative learning*. Paper presented at the Proceedings of European conference on artificial intelligence in education, Ficha Tecnica, Lisbon.

Lepper, M., & Chabay, R. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl & A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems* (pp. 242-257). Hillsdale, NJ: Erlbaum.

Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377-390.

Pekrun, R. (in press). Academic emotions. In T. Urdan (Ed.), *APA educational psychology handbook* (Vol. 2). Washington, DC: American Psychological Association.

Person, N., Graesser, A., Magliano, J., & Kreuz, R. (1994). Inferring What the Student Knows in One-to-One Tutoring - the Role of Student Questions and Answers. *Learning and Individual Differences, 6*(2), 205-229.

Reeves, B., & Nass, C. (2003). *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK: Cambridge University Press.

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling learner affect with theoretically grounded dynamic bayesian networks In S. D'Mello, A. Graesser, B. Schuller & J. Martin (Eds.), *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction* (pp. 286-295). Berlin Heidelberg: Springer-Verlag.

Schutz, P., & Pekrun, R. (Eds.). (2007). *Emotion in education*. San Diego, CA: Academic Press.

VanLehn, K., Graesser, A., Jackson, G., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*(1), 3-62.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology, 4*(3/4), 129-163.