Emotional Rollercoasters: Day Differences in Affect Incidence during Learning

Sidney K. D'Mello
University of Notre Dame
384 Fitzpatrick, Notre Dame, IN 46556, USA
sdmello@nd.edu

Abstract

This study investigated between-session variability in affective state incidence during learning with an intelligent tutoring system (ITS). We tracked 12 affective states (plus neutral) while three learners completed their statistics coursework in eight 45-minute weekly sessions with the ALEKS ITS. The results indicated that engagement/flow was the most frequent affective state that demonstrated considerable stability across sessions. Nine other affective states (anger, anxiety, boredom, contempt, curiosity, disgust, eureka, happiness, surprise) were rarely observed both within and across sessions. Confusion and frustration occurred with some frequency but there was considerable between-session variability in the incidence of these states. Followup analyses indicated that some of the between-session variability could be explained by initial (before session) positive and negative mood and initial physiological arousal (electrodermal activity). Implications of the findings for basic research on affect and learning and for affect-aware ITSs are discussed.

Introduction

Affect-aware or affect-sensitive intelligent tutoring systems (ITSs) aim to incorporate affect in their pedagogical decision making (Conati & Maclaren, 2009; D'Mello et al., 2010; Forbes-Riley & Litman, 2011; Sabourin, Mott, & Lester, 2011; Woolf et al., 2009). This strategy is motivated by a preponderance of research on the inextricable coupling between affect and cognition during learning (e.g., Clore & Huntsinger, 2007; Isen, 2008). Affect-aware ITSs come in many flavors (see Calvo & D'Mello, 2011 for a review). For example, Affective AutoTutor is an ITS that automatically detects learner boredom, confusion, and frustration from facial features, body movements, and discourse/context cues and responds with motivational dialog moves (D'Mello et al., 2010). UNC-ITSPOKE is an ITS that automatically detects (from acoustic-prosodic features and lexical cues) and responds to the certainty and correctness of learner spoken responses (Forbes-Riley & Litman, 2011). These two affect-aware ITSs can be considered to be *reactive* in that they first detect and then respond (react) to certain affective states *after* they occur.

As one might expect, there are a number of fundamental challenges that need to be addressed over the course of building an effective reactive affect-aware ITS. One initial step involves identifying the affective states that learners' experience during interactions with the ITS. A second step consists of developing automated methods to detect those states. The third step typically involves augmenting the decision making engine of the ITS to incorporate the sensed states. A fourth step might also involve emotion synthesis strategies, especially for ITSs endowed with embodied conversational agents.

The present paper focuses on one of the most fundamental aspects in building a reactive affect-aware ITS – identifying the affective states that occur during learning with the ITS. This is not a novel area of research as numerous studies have tracked affective incidence during learning with ITSs and other learning environments (see Related Work section). However, the present study takes a different perspective on affective incidence. Instead of focusing on affective incidence *across* learners *within* a single session with the ITS, as is typically done, our emphasis is on affective incidence *within* learners *across* multiple sessions with the ITS. In other words, the emphasis is on assessing between-session variability in affect incidence across multiple learning sessions. This information is critical for both the scientific goal of understanding affect during learning as well as for the engineering goal of building affect-aware ITSs (see Motivation and Overview section).

Related Work

Over the last decade, researchers have been actively investigating the incidence of affective states that occur during interactions with learning technologies. This is a rich area of research as can be evidenced by considerable diversity in learning technologies, content areas, student populations, and methodologies used to track affect (Afzal & Robinson, 2011; Alhothali, 2011; Craig, Graesser, Sullins, & Gholson, 2004; Lester, McQuiggan, & Sabourin, 2011; Mills & D'Mello, 2012; Rodrigo & Baker, 2011a, 2011b). Some of the learning technologies studied include ITSs, such as AutoTutor, Aplusix, Scatterplot Tutor, educational games like Crystal Island and The Incredible Machine, and computer interfaces for problem solving, reading comprehension, and essay writing. Some of the topics covered included physics, chemistry, biology, algebra, geometry, logic puzzles, analytical reasoning, critical thinking, computer literacy, social studies, and argumentative writing. The studies feature diverse populations of learners including middle-school, high-school, and college students across multiple countries. Finally, learner affect has been tracked with assorted methodologies such as online observations, emote-aloud protocols, cued-recall, video coding, and delayed self-report (see Porayska-Pomsta, Mavrikis, D'Mello, Conati, & Baker, 2013 for a review).

D'Mello (in press) recently reported a meta-analysis on affect incidence as reported in 24 such studies. The main findings were that: (a) engagement/flow was the most frequent affective state that consistently occurred across studies; (b) boredom confusion, curiosity, happiness, and frustration, occurred with some frequency but varied substantially across studies; while (c) contempt, anger, disgust, sadness, anxiety, delight, fear, and surprise were consistently infrequent across studies. This meta-analysis provides some preliminary evidence to suggest that affect-aware ITSs might want to initially consider focusing on engagement/flow, boredom, confusion, curiosity, happiness, and frustration.

Motivation and Overview of Current Research

The studies noted above have made impressive progress identifying the affective states that occur during learning with technology. However, one commonality in these studies is that they mainly focus on tracking affect of a sample of learners in one (most cases) or two or three (occasionally) sessions. This is by no means a criticism, because the goal of these studies is to identify the most prominent affective states *across* learners. However, an equally important question is whether learner affect is stable or exhibits variability across sessions. This is the focus of the present study.

Affective states can be *dispositional* in that learners experience particular affective states consistently across sessions (e.g., boredom due to dislike of a subject like biology) or *situational,* where affect incidence varies across sessions (e.g., more boredom when learning about cells in one session compared to ecological succession in another session). It is also possible that the likelihood of an affective state being situational, dispositional, or a combination of both varies for different affective states. Identifying which affective states are situational vs. dispositional can inform both basic research on the incidence of affect during learning with ITSs and on applied work focused on developing affect-sensitive ITSs. More specifically, when an affective state is more or less dispositional, then tracking it over one or two sessions might be sufficient to obtain a reliable indicator of its incidence. On the other hand, one or two session estimates of affect incidence will likely be unreliable for affective states that are more situational.

The present research investigated the extent to which 12 affective states (and neutral) demonstrated dispositional or situational properties in a data set involving three learners who completed eight 45-minute sessions with an online statistics tutoring system called ALEKS or **A**ssessment and **LE**arning in **K**nowledge **S**paces (Canfield, 2001; Falmagne, Cosyn, Doignon, & Thiéry, 2006). Using a case study methodology, we monitored affect of a small set of learners across multiple sessions, because our primary goal was to assess affect incidence within individual learners but across sessions (as opposed to across learners but within a session). Therefore, although the N of 3 learners is small, the data set encompasses more than 1,000 minutes of tutoring ($45 \times 8 = 360$ minutes per learner) and is sufficient for an initial investigation into which affective states are more dispositional vs. situational during learning.

Method

Participants (Learners)

Participants (called learners) were three undergraduate students enrolled in an introductory Psychological Statistics course at a large public University in the mid-south U.S. The learners signed up for a special section of the course consisting of independent study with ALEKS rather than the traditional in-class section. Learners were recruited during the first (and only) in-class meeting. Learners A and B were Caucasian females while Learner C was an African-American male. Learners were compensated $175 for participating in the study and an additional $50 bonus for completing all eight sessions.

Measures

The three measures of interest include learners' self-reported affective states, their initial moods, and their physiological arousal.

Discrete Affective States

Learners self-reported their current affective state by selecting one out of 12 affective states plus neutral at multiple points in the session (see Procedure). The affective states included basic emotions (anger, disgust, contempt, happiness, and surprise), learning-centered affective states (anxiety, confusion, boredom, curiosity, eureka, engagement/flow, and frustration), and neutral. Learners had access to a sheet with definitions of the affective states throughout the sessions.

Initial Mood

Learners initial (before the learning session) mood was measured with the Positive and Negative Affect Schedule-Expanded (PANAS-X) (Watson & Clark, 1994; Watson, Clark, & Tellegen, 1988), which is a validated and one of the most widely used measure of mood.

Physiological Arousal

Learners physiological arousal (electrodermal activity or EDR) was measured throughout the learning session with the Thought Technology SA9309M skin conductance sensor. Learners placed the two sensor cuffs on the index and middle fingers of their non-dominant hands. Due to equipment failure, EDR data was missing for two sessions for Learner B and for three sessions for Learner C.

ALEKS

ALEKS is a web-based ITS for math, statistics, science, and other domains. It relies on a combination of periodic knowledge assessment, reassessment, and tutoring. It is based on knowledge space theory which maintains that knowledge in well-defined domains can be represented as a collection of interrelated states (e.g., prerequisite relationships). ALEKS determines the current knowledge state of the learner by carefully selecting problems/questions based on their prior knowledge levels and then develops an individualized plan to help the learner advance deeper into the knowledge space (Falmagne et al., 2006). The version of ALEKS used in the present study covered basic math and statistics concepts, such as descriptive statistics, probability, distributions, hypothesis testing, and introductory inferential statistics.

Procedure

Learners were tested individually in a laboratory environment across eight sessions (per learner). Learners were informed that they would complete their normal statistics coursework with ALEKS, but would do it in our lab once a week. Accordingly, each learner scheduled a weekly session prior to beginning the study.

Activities in each session involved: (a) putting on the cuffs of the EDR sensor, (b) completing the PANAS-X, (c) completing a 45 minute tutorial session with ALEKS, and (d) completing some additional questionnaires (not discussed here).

Interactions with ALEKS were self-paced and topics were selected based on a combination of ALEKSs assessment of suitable topics (based on its knowledge space model) and on learner choice (subset of suitable topics). Affective states were self-reported every three minutes via a popup window consisting of a list of the 12 states (and neutral). Learners were simply asked select one state from the list.

Results and Discussion

The results are organized as follows. First, we identified the frequent affective states that learners experienced during interactions with ALEKS. Second, we investigated between-session variability in affective states. Third, we assessed if affective states could be predicted from initial (before learning session) mood and physiological arousal.

Affective State Incidence

The 14 affect reports collected per session were first proportionalized so that the sum of affect proportions for a learner across a session would be 1. Means and standard deviations of affect proportions for each learner across sessions are shown in Table 1. We performed one sample t-tests comparing each mean to 0 to assess which affective states significantly showed non-zero incidence across sessions. A one-tailed test was used because we know the direction of the mean since a proportion cannot be negative. Means that were significantly (p < .05) greater than zero are annotated in bold in Table 1.

The results yielded a number of interesting patterns. First, despite the small sample, there was notable consistency across learners. Engagement/flow appeared to be the most frequent affective state; comprising a mean (across learners) of 56.7% of all affect reports. Mean proportions of confusion and frustration were significantly greater than zero on five out of the six t-tests, but were lower than engagement/flow. Together, these two states comprised 20.7% of the affect reports. While engagement/flow, confusion, frustration, and neutral (mean of 9.7%) comprised 86.9% of the data, the remaining eight states were extremely rate and collectively comprised a mere 13.1% of the affect reports. Hence, the subsequent analyses focus on engagement/flow, confusion, and frustration.

Table 1. Means and standard deviation (in parentheses) of affect proportion across sessions

Affect	Learner A	Learner B	Learner C
Eng./Flow	**.56 (.25)**	**.63 (.16)**	**.51 (.36)**
Confusion	**.15 (.22)**	**.06 (.07)**	.07 (.17)
Frustration	**.24 (.26)**	**.03 (.04)**	**.07 (.10)**
Anger	.00 (.00)	.01 (.03)	.01 (.03)
Anxiety	.00 (.00)	.04 (.09)	.01 (.03)
Boredom	.00 (.00)	.00 (.00)	.01 (.03)
Contempt	.01 (.02)	.01 (.03)	.00 (.00)
Curiosity	.00 (.00)	**.11 (.09)**	.01 (.03)
Disgust	.00 (.00)	.00 (.00)	.01 (.03)
Eureka	.00 (.00)	.02 (.05)	.02 (.05)
Happiness	.01 (.02)	**.04 (.05)**	.01 (.03)
Surprise	.03 (.10)	.01 (.03)	.03 (.05)
Neutral	.00 (.00)	**.05 (.07)**	**.24 (.19)**

Notes. Eng./flow = engagement/flow; Bolded cells indicate means that are significantly greater than zero with one-tailed one-sample t-tests.

Between-session Variability in Affect Incidence

We divided the mean (across sessions) proportional occurrence of each affective state with the standard deviation as a rough measure of the signal (mean) to noise ratio (standard deviation). A signal to noise ratio (SNR) greater that 1 would imply more signal than noise and vice versa for values less than 1. These results are shown in Table 2, where we note SNR's under 1 for confusion and frustration for all three learners. On the other hand, engagement/flow demonstrated strong and consistent SNRs across learners.

Table 2. Signal to noise ratio (mean divided by standard deviation) affect proportions across sessions

Affect	Learner A	Learner B	Learner C
Eng/Flow	2.22	3.89	1.40
Confusion	0.70	0.75	0.41
Frustration	0.93	0.72	0.73

To further investigate the low SNRs for confusion and frustration, proportional occurrence of these states for the individual sessions and learners are depicted in Figure 1 and Figure 2, respectively. We note that confusion levels were highly variable across sessions. Learner A reported non-zero confusion for only four out of the eight sessions and confusion levels in Sessions 1 and 6 were almost double than those of Sessions 7 and 8. Learner B also reported non-zero confusion for only four of the eight sessions, while Learner C reported some confusion for only two sessions.

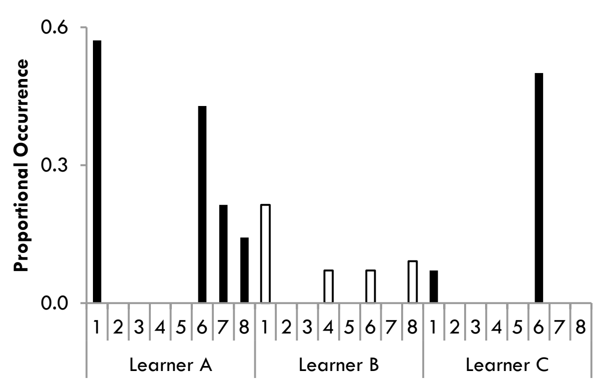

Figure 1. Confusion levels across sessions

A somewhat similar pattern was observed for frustration. Although Learner A reported some frustration for six out of the eight sessions, Learners B and C reported non-zero frustration for at most half of the sessions. Taken together, these results suggest that the mean confusion and frustration levels across sessions (reported in Table 1) might not be a very reliable due to the considerable between-session variability in the incidence of these states.

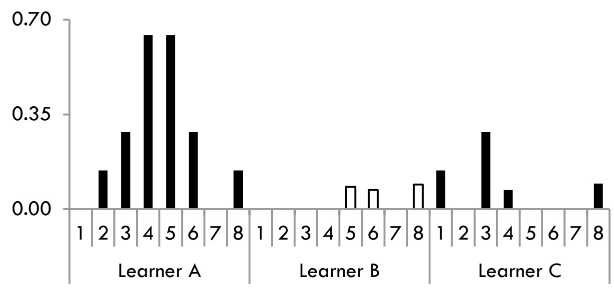

Figure 2. Frustration levels across sessions

Initial Mood and Arousal as Predictors of Affect

We analyzed if affective states could be predicted from initial mood states and initial physiological arousal. We focused on general positive and general negative mood from the PANAS-X and these variables could range from 1 to 5. The mean EDR values over the first 10 seconds of each session were considered as a proxy for initial physiological arousal.

The analyses proceeded by regressing proportional occurrence of confusion, frustration, and engagement/flow on initial positive and negative affect (Model 1) and on physiological arousal (Model 2). Due to the repeated nature of the design, where sessions are nested within learners, a mixed effects modeling approach was adopted (Pinheiro & Bates, 2000). The pre-learning (initial) variables were the fixed effects (independent variables) and learner and session number were the random effects. The lme4 library in R was used for the requisite computation.

Initial negative mood was a marginally significant predictor of confusion, $F(1, 21) = 3.67$, $p = .061$, $B = -.128$, but initial positive mood was not a significant predictor, $F(1, 21) = .033$, $p = .857$, $B = -.005$. Surprisingly, negative mood was a negative instead of a positive predictor, which is contrary to expectations. The model for frustration was more in line with expectations, with initial negative mood as a significant positive predictor, $F(1, 21) = 8.52$, $p = .008$, $B = .203$, and initial positive mood as a marginally significant negative predictor, $F(1, 21) = 2.96$, $p = .100$, $B = -.073$. Neither initial negative mood, $F(1, 21) = 1.86$, $p = .187$, $B = -.185$, nor initial positive mood, $F(1, 21) = .166$, $p = .688$, $B = .019$, were significant predictors of engagement/flow, however, the coefficient for initial negative mood was in the expected direction.

Initial physiological arousal was a significant positive predictor of confusion, $F(1, 17) = 9.33$, $p = .007$, $B = .109$, but not for frustration, $F(1, 17) = .148$, $p = .705$, $B = .029$, or engagement/flow, $F(1, 17) = .883$, $p = .361$, $B = -.073$. In summary, these results suggest that some of the between-session variability in affective state incidence for confusion and frustration can be explained by initial mood states and physiological arousal.

General Discussion

The present study tracked 12 affective states (plus neutral) while three learners completed their statistics coursework in eight 45-minute sessions with the ALEKS ITS. In this section we provide a summary of our major findings along with implications and discuss limitations and future work.

Main Findings and Implications

The results were illuminating in a number of respects. First, our findings were mostly consistent with the previous meta-analysis on the affective states that occur during learning with technology (D'Mello, in press). We discovered that engagement/flow was the most frequent affective state, confusion and frustration were somewhat frequent, while the remaining nine states rarely occurred. However, while the previous meta-analysis also indicated that boredom was a prominent affective state, this was not observed in the data. The low boredom rates in the present study might be attributed to the material being directly relevant to learners' course grade and academic progress. In general, the present study supports the claim that confusion and frustration are two important negative states for affect-aware technologies to sense and respond to.

Our second important finding was that there was considerable variability in levels of confusion and frustration across sessions. A modicum of between-session variability in affect state incidence is to be expected due to changes in the content and other factors. However, the fact that the same learners reported confusion/frustration levels of zero in some sessions, and more than 50% in others, is notable and of some concern. This suggests that measurement of these states in one session, as is typically done in this area of research, might not provide very reliable estimates of general affect incidence across sessions, at least for confusion and frustration. The considerable between-session variability in the incidence of these states might also pose problems for automated affect detectors trained on data collected in a single session, because these systems will have difficulties adjusting to changing baselines (prior probabilities of incidence).

Our third finding pertains to sources of between-session variability in affect incidence. One source consists of what the learners bring to the session and this was tracked via initial mood states and initial levels of physiological arousal. Another source of variability consists of more domain-specific factors, such as the learning content and the dynamics of the student-tutor interaction. We focused on the first source and found that some of the variability in incidence of confusion and frustration could be explained by initial mood states (particularly negative moods) and initial levels of physiological arousal. Automated affect detectors can presumably use this information to model between-session variability in affect incidence.

Limitations and Future Work

It is important to acknowledge some limitations with the present study that should be rectified in future work. One limitation pertains to the sample size of three learners. Despite the small sample size, it should be noted that the overall interaction time (1080 minutes) was roughly equivalent to the median interaction time (1377 minutes) of The 24 reviewed in the aforementioned meta-analysis (D'Mello, in press). Nevertheless, a larger sample would allow us to investigate between-learner variability in addition to the between-session variability studied here. The participant recruitment procedure is also a limitation because of the potential for self-selection biases. Therefore, it would be important to replicate the findings with randomly selected learners, but this is more difficult to do for authentic tutorial sessions. Third, the present study did not include any learning measures because it was difficult to develop uniform knowledge assessments since each learner progressed at his or her own pace (i.e., tutorial content varied across learners). This limitation can be addressed by replicating the study with a more fixed content sequence, so that learning measures can be obtained and correlated with affect incidence.

Concluding Remarks

The present study provides an initial investigation into between-session variability in affect incidence during learning with an ITS. Nine out of the 12 states exhibited dispositional properties by being consistently present (engagement/flow) or consistently absent (remaining eight states) across sessions. Interestingly, confusion and frustration, which are two of the most interesting and impactful affective states in learning contexts, exhibited situational variability. Although we identified initial mood and initial physiological arousal as potential factors to explain this variability, the next step is to ascertain how these pre-learning variables interact with tutorial events (content, problems, performance, feedback, etc.) to predict the incidence of confusion and frustration.

Acknowledgments

The author gratefully thanks Xiangen Hu for helping with participant recruitment. This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1235958). 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

Afzal, S., & Robinson, P. (2011). Natural affect data: Collection and annotation. In R. Calvo & S. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 44-70). New York, NY: Springer.

Alhothali, A. (2011). *Modeling user affect using interaction events.* (M.S.), University of Waterloo, Waterloo, Canada.

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

Canfield, Ward. (2001). ALEKS: A Web-based intelligent tutoring system. *Mathematics and Computer Education, 35*(2), 152-158.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393-399. doi: 10.1016/j.tics.2007.08.005

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.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29*, 241-250. doi: 10.1080/1358165042000283101

D'Mello, S. (in press). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., . . . Graesser, A. (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 the 10th International Conference on Intelligent Tutoring Systems* (pp. 245-254). Berlin / Heidelberg: Springer.

Falmagne, Jean-Claude, Cosyn, Eric, Doignon, Jean-Paul, & Thiéry, Nicolas. (2006). The assessment of knowledge, in theory and in practice *Formal concept analysis* (pp. 61-79): Springer.

Forbes-Riley, K., & Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication, 53*(9-10), 1115-1136. doi: 10.1016/j.specom.2011.02.006

Isen, A. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. Haviland-Jones & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548-573). New York, NY: Guilford.

Lester, J.C., McQuiggan, S.W., & Sabourin, J.L. (2011). Affect recognition and expression in narrative-centered learning environments. In R. Calvo & S. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 85-96). New York: Springer.

Mills, C., & D'Mello, S. K. . (2012). Emotions during writing on topics that align or misalign with personal beliefs. In S. Cerri, W. Clancey, G. Papadourakis & K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 638-639). Berlin Heidelberg: Springer-Verlag.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.

Porayska-Pomsta, K., Mavrikis, M., D'Mello, S. K., Conati, C., & Baker, R. . (2013). Knowledge Elicitation Methods for Affect Modelling in Education. *International Journal of Artificial Intelligence in Education, 22*, 107-140.

Rodrigo, M., & Baker, R. (2011a). Comparing learners' affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning, 6*(1), 43-66.

Rodrigo, M., & Baker, R. (2011b). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In R. Calvo & S. D'Mello (Eds.), *New Perspective on Affect and Learning Technologies* (pp. 183-202). New York: Springer.

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.

Watson, D., & Clark, L. A. . (1994). The PANAS-X: Manual for the positive and negative affect schedule-Expanded Form. Iowa City: University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect - the Panas Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

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.