UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Posture as a Predictor of Learner's Affective Engagement

Permalink
https://escholarship.org/uc/item/7hs9v2hr

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
D'Mello, Sidney S.
Chipman, Patrick
Graesser, Art

Publication Date
2007

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Posture as a Predictor of Learner's Affective Engagement

Sidney D'Mello[1] (sdmello@memphis.edu)
Patrick Chipman[2] (pchipman@memphis.edu)
Art Graesser[2] (a-graesser@memphis.edu)

[1]Department of Computer Science, [2]Department of Psychology
University of Memphis, Memphis TN, 38152, USA.

Abstract

This research demonstrates the utility of automatically monitoring a student's posture to track the affective states of boredom (low engagement) and flow (high engagement), which have been shown to influence learning. After a tutoring session with AutoTutor, the affective states of the student were rated by the learner, a peer, and two trained judges. Our results indicated that the affective state of flow was manifested through heightened pressure exerted on the seat of a pressure sensitive chair. Boredom, in turn, was associated with an increase in the pressure exerted on the back coupled with a rapid change in pressure on the seat, perhaps indicative of a state of restlessness. We also investigated the diagnosticity of each of the posture features and the reliability of a computer automatically discriminating episodes of boredom versus flow, which is a major discrimination in any affect-sensitive tutoring system.

Keywords: Posture patterns; affective states; emotions; learning; AutoTutor; classifying affect; Intelligent Tutoring Systems

Introduction

The task of maintaining a student's engagement in educational activities is extremely challenging. Establishing and maintaining the engagement of learners is especially critical in situations with high degrees of learner control, such as in distance education, computer-based tutoring, and informal learning environments. For instance, with web-based instruction, individuals are one-mouse-click-away from ending the session. Several traditional approaches have directly addressed this problem, such as collaborative learning (Palincsar & Brown, 1984), apprenticeship learning (Rogoff, 1990), educational games and simulations (Ferrari, Taylor, & VanLehn, 1999), and inquiry learning (Chinn & Malhotra, 2002). All of these approaches both promote active learning and offer scaffolding to sustain motivation and engagement. They also structure the learning environment so that it matches a student's zone of proximal development (Brown, Ellery, & Campione, 1998) and learning rate (Metcalfe & Kornell, 2005). Determining the appropriate level of difficulty is a non-trivial task, however. It is not obvious what metrics accurately scale the relative difficulty of complex topics such as those in undergraduate science courses. Each individual has his/her own specific zone of proximal development which changes over time in response to instruction. The goal of providing training that hits this 'moving target' is exceptionally difficult.

An exciting new alternative involves the use of *emotion-sensitive* intelligent tutoring systems (ITSs). These tutoring systems attempt to incorporate the learner's emotions (or affective states) into their pedagogical strategies in order to enhance engagement, motivation, and learning (D'Mello, et al., 2005; Forbes-Riley & Litman, 2004; Kort, Reilly, & Picard, 2001). A fundamental challenge in the development of an emotion-sensitive ITS involves reliably measuring the affective states of the learner. This requires the development of a user model that captures the manner in which learners intentionally or implicitly exhibit affect in a naturalistic learning environment. With the user model in hand, the next step is to develop a computational system to automatically diagnose and incorporate the emotions of learner into the pedagogical strategies of the ITS.

This paper addresses these challenges by an in-depth analysis of body posture, a dimension that has been rarely investigated by cognitive scientists. We identified the manner in which learners express particular emotions by modulating their gross body language. We used automated posture tracking hardware and software instead of human coders. This allowed us to develop automated algorithms to identify learner emotions on the basis of the detected patterns between posture and affect.

The use of posture to infer affect is interesting because it is rarely the case that posture is intentionally monitored by humans. The significance of nonverbal behaviors in expressing affect is widely acknowledged, but the vast majority of the scientific literature is restricted to the monitoring of facial features (Ekman & Friesen, 1978), speech contours (Ang et al., 2002), and physiological signals such as electromyography, heart rate, and skin conductance. Our rationale for expecting posture features to be diagnostic of affect in learning environments is motivated by embodied theories of cognition (Clark, 1997; Glenberg, Havas, Becker & Rinck, 2005; de Vega, 2002). Theories of embodied cognition postulate that cognitive processes are constrained substantially by the environment and by the coupling of perception and action. If the embodied theories are correct, then the cognitive and emotional states of a learner are manifested in their body language. An added advantage of monitoring posture patterns is that these motions are ordinarily unconscious, unintentional, and thereby not susceptible to social editing, at least compared with facial expressions and gestures. Ekman and Friesen (1969), in their studies of deception, have coined

the term *nonverbal leakage* to refer to the increased difficulty faced by liars, who attempt to disguise deceit, through less controlled channels such as the body when compared to facial expressions.

A few studies have documented the importance of posture in expressing affect (e.g. Coulson, 2004; Schouwstra & Hoogstraten, 1995; Wallbott, 1998). However, the impetus of these investigations has been directed towards "basic" emotions (i.e. anger, fear, sadness, enjoyment, disgust, and surprise, Ekman & Friesen, 1978). While these basic emotions are ubiquitous in everyday experience, there is a growing body of evidence that suggests that they rarely play a significant role in learning (D'Mello, et al. 2006; Graesser et al., 2006; Kort, Reilly, & Picard, 2001). In particular, the prominent affective states during tutoring are boredom, flow, frustration, and confusion (Graesser et al., 2006).

Some of these emotions (i.e., affect states) might be viewed as being cognitive states rather than emotions by some colleagues, whereas other researchers would classify them as either emotions or affect states (Barrett, 2006; Meyer & Turner, in press). 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.

Within the context of engagement, the two most relevant of these affect states are boredom and flow. Craig et al. (2004) reported that increased levels of boredom were negatively correlated with learning (r = -.39) while students learned about computer literacy topics with an intelligent tutoring system. In contrast, the flow experience (i.e., high engagement, Csikszentmihalyi, 1990) was positively correlated with learning (r = .29).

Therefore, our first step in exploring body posture and emotions during learning examined the affective states of boredom (low engagement) and flow (high engagement), viewing them as approximate endpoints on a continuum of engagement. Perhaps the most relevant research involving the monitoring of posture patterns to infer engagement was conducted by Mota and Picard (2003). They analyzed temporal transitions of posture patterns to classify the interest level of children while they performed a learning task on a computer. Children are much more active than the college students we investigated, so the algorithms to detect affective states might be more subtle than their temporal transitions.

Methods

Participants

The participants were 28 undergraduate students from a mid-south university who participated for extra course credit.

Materials

AutoTutor. AutoTutor is a fully automated computer tutor that simulates human tutors and holds conversations with students in natural language (Graesser et al., 2005; Graesser, et al, 1999). AutoTutor helps students learn Newtonian physics and computer literacy by presenting challenging problems (or questions) from a curriculum script and engaging in a mixed-initiative dialog while the learner constructs an answer.

Body Pressure Measurement System (BPMS). The BPMS system, developed by Tekscan™ (1997), consisted of a thin-film pressure pad (or mat) that could be mounted on a variety of surfaces. The pad was paper thin with a rectangular grid of sensing elements. Each sensing element provided a pressure output in mmHg. Our setup had one sensing pad placed on the seat of a Steelcase™ Leap Chair and another placed on the back of the chair (see Figure 1).

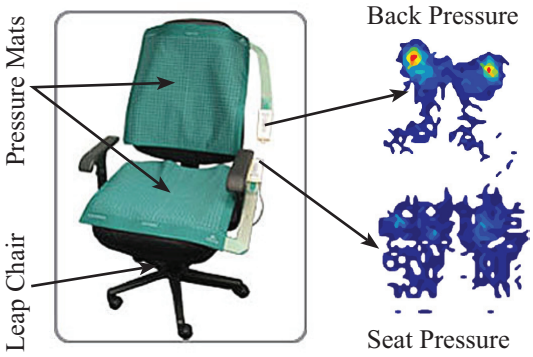

Figure 1: Body Pressure Measurement System

The output of the BPMS system consisted of 38x41 matrix (rows x columns) with each cell in the matrix monitoring the amount of pressure as reported by the corresponding sensing element. Therefore, in accordance with our setup, at each sampling instance (1/4 second), matrices corresponding to the pressure in the back and the seat of the chair were recorded for future, offline analyses.

Procedure

The study was divided into two phases. The first phase was a standard pretest–intervention–posttest design. The participants completed a pretest with multiple-choice questions, then interacted with the AutoTutor system for 32 minutes on one of three randomly assigned topics in computer literacy (Hardware, Internet, Operating Systems), and then completed a posttest. During the tutoring session, the system recorded a video of the participants' face, their pressure patterns, and a video of their computer screen.

The second phase involved *affect judgments* by the learner, a peer, and two trained judges. A list of the affective states and definitions was provided for all judges. The states were boredom, confusion, flow, frustration, delight, neutral and surprise. The selection of emotions was based on previous studies of AutoTutor (Craig et al., 2004; D'Mello et al., 2006; Graesser et al., 2006) that collected observational data and *emote aloud* protocols while college students learned with AutoTutor.

The affect judging session proceeded by displaying video streams of both the learner's screen and face, which

were captured during the AutoTutor session. The raters were instructed to make judgments on what affective states were present in each 20-second interval; at these points the video automatically paused for their affect judgments. Four sets of emotion judgments were made for the observed affective states of each participant's AutoTutor session. First, for the *self* judgments, the participant watched his or her own AutoTutor session immediately after having interacted with it. Second, for the *peer* judgments, participants came back a week later to watch and rate another participant's session on the same topic in computer literacy. Finally, two *trained* judges independently rated all of the sessions. These judges had been trained on how to detect facial action units according to Ekman's Facial Action Coding System (Ekman & Friesen, 1978) and on characteristics of dialogue.

Data Treatment

Analysis of Agreement among Judges. Interjudge reliability was computed using Cohen's kappa for all possible pairs of judges: self, peer, trained judge1, and trained judge2. There were 6 possible pairs altogether. The kappas were reported in Graesser et al. (2006): self-peer (.08), self-judge1 (.14), self-judge-2 (.16), peer-judge1 (.14), peer-judge2 (.18), and judge1-judge2 (.36). These kappa scores revealed that the trained judges had the highest agreement, the self and peer pair had lowest agreement, and the other pairs of judges were in between. It should be noted, however, that the kappa scores increase substantially [self-peer (.12), self-judge1 (.31), self-judge2 (.24), peer-judge1 (.36), peer-judge2 (.37), and judge1-judge2 (.71)] when we focused on observations in which the learner declared they had an emotion, as opposed to many random points when they were essentially neutral. The kappa scores are on par with data reported by other researchers who have assessed identification of emotions by humans (e.g. Ang et al., 2002; Forbes-Riley & Litman, 2004).

Extraction of Posture Features. At each sampling point (1/4 second) the BPMS system provides a spatial map of the pressure exerted on the seat and the back of the chair. By averaging across each of the 1558 sensing elements on the back and seat pads one obtains the *back net pressure* and the *seat net pressure*. However, since we are primarily interested in posture patterns during an emotional experience (as indicated by the self, peer, or 2 trained judges) these features were only computed when an emotional episode was recorded.

We also considered two additional features that attempted to measure the rate of change in pressure exerted by the learner on the back and seat of the chair during an emotional episode. That is, by computing the rate of change in pressure 2 seconds before and 2 seconds after the learner was judged to have experienced an emotion, we were able to operationally define a measurement of the amount of body activity of the learner when he or she experienced an emotion. These features are termed the *back pressure change* and the *seat pressure change*.

Data Selection. Three data sets were constructed by temporally integrating the 4 posture features with the emotion judgments of the raters. Specifically, the four posture features (independent variables) were assessed in predicting the emotion of the learner (dependent variable). The first two models consisted of posture features aligned with judgments of the affective states of boredom and flow provided by the self ($N_{BOREDOM}$ = 483; N_{FLOW} = 593) and the peer ($N_{BOREDOM}$ = 582; N_{FLOW} = 605). The third model was constructed by considering affect ratings where both trained judges agreed on whether the learner was experiencing boredom or flow ($N_{BOREDOM}$ = 268; N_{FLOW} = 224).

Results

Relating Posture with Boredom and Flow

Figure 2 presents descriptive statistics (mean + 95% confidence interval, CI) of an item-level analysis for each of the posture features, segregated by boredom and flow. The results indicate that boredom is accompanied by an increase in the pressure exerted on the back of the chair (see Figure 2a). This pattern was statistically significant for the data sets in which the affect judgments were provided by the peer ($F(1,1185) = 8.51$, p[1] $< .01$) and the trained judges ($F(1,490) = 9.53$) but not the self judgments ($F < 1.6$). In contrast, the affective state of flow appears to be manifested by a heightened pressure exerted on the seat of the chair (Figure 2b). This relationship was statistically significant across all three data sets, $F_{SELF}(1,1076) = 13.78$; $F_{PEER}(1,1185) = 5.44$, $p < .05$; $F_{JUDGES}(1, 490) = 52.47$.

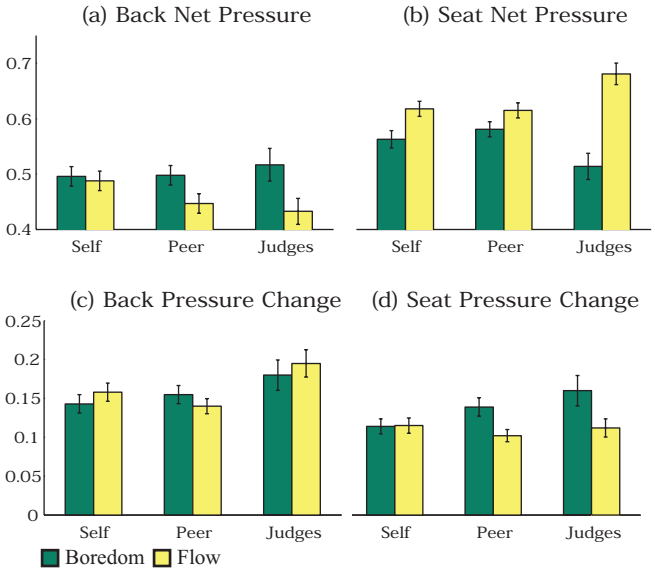

Figure 2: Descriptive Statistics for Posture Features Segregated by Boredom and Flow

It appears that the change in pressure exerted on the back of the chair is quantitatively similar for boredom and flow.

[1] $p < .01$ in all analyses unless specified otherwise.

Therefore, this feature does not appear to be very useful in discriminating these emotions (Figure 2c). However, Figure 2d indicates that boredom is typically accompanied by heightened level of activity on the seat of the chair (i.e., fidgeting). This trend was observed for the data sets where the affect judgments were provided by the peer ($F(1,1185) = 13.41$) and the trained judges ($F(1,490) = 7.43$) but not the self judgments ($F < 1$).

The above mentioned relationships between the posture features and the affective states of boredom and flow can be aligned within a *proclivity-arousal* framework. One can think of heightened pressure in the seat as resonating with a tendency to position one's body towards the source of stimulation (i.e., high proclivity for inclining toward the AutoTutor interface, or a short distance between the nose and the screen). On the other hand, an increase in pressure on the back of the chair suggests that the learner is leaning back and detaching himself or herself from the stimulus (low proclivity). On the basis of these operational definitions of proclivity to a stimuli, our results are in the expected directions. Specifically, the affective state of flow is being manifested by an increased pressure on the seat of the chair, which would be indicative of high proclivity to the tutoring stimulus during periods of engagement. However, during episodes of boredom the learners seem to lean back, presumably disengaging from the learning environment.

Our results also indicate that boredom is accompanied by an increase in the rate of change of pressure exerted on the seat. Heightened arousal is associated with the boredom experience, as learners mentally disengage from the tutor and divert their cognitive capabilities to fidget around and alleviate their ennui. This pattern of increased arousal accompanying disengagement (or boredom) replicates the general findings by Mota and Picard (2003), where they monitored activity related posture features and discovered that children fidget when they were bored while performing a learning task on a computer.

Diagnosticity of Posture Features

Our discovery, in summary, is that the level of pressure in the seat and the back of the chair, and the level of activity in the seat are the main posture features that discriminate boredom from flow, two ends of the engagement continuum. This raises the question as to which of these features is the most diagnostic of engagement. We conducted a small computer simulation in which the C4.5 algorithm (Quinlan, 1993) was used to construct a decision tree capable of discriminating between boredom and flow. For ease of interpretation, each posture feature was dichotomized (i.e. low vs. high) and we only present the decision tree obtained from the data set in which the two trained judges agreed on the learner's emotion (Figure 3).

The C4.5 algorithm operates by first computing the entropy (noise or impurity) associated with the data. Then, the information gain (the reduction in the entropy) that each feature provides is estimated. Hence, features that yield a higher information gain achieve greater diagnostic power and are subsequently used as branching nodes (seat net pressure) before features with lower information gain (back net pressure and seat pressure change).

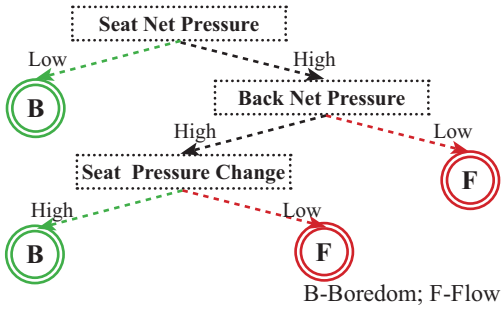

Figure 3: Decision Tree to Discriminate Boredom and Flow

Discriminating between Boredom and Flow

The next important step is to investigate how well a computer can automatically discriminate between boredom and flow on the basis of the learner's posture patterns. The analyses proceeded by first expanding the set of posture features to include small changes in the pressure exerted on the back and the seat before and after the learner's emotional experiences. Specifically, we computed the difference between the net pressure exerted three seconds before and after an emotional episode. We also measured the difference between the net pressure exerted on the back and the seat during emotion E_t and E_{t-1} (i.e., the net pressure for the previous emotional episode). Finally, two features that examined the net pressure coverage on the back and the seat were included. These variables measure the proportion of non-negative sensing units on each pad.

We investigated the reliability by which three, theoretically distinct, machine learning techniques could discriminate between engagement (flow) and lack thereof (boredom). These included a Bayesian model, a neural network, and a simple nearest neighbor classifier.

Table 1. Classification Accuracy (Kappa Scores)

Affect	**Overall**			**Boredom**			**Flow**		
Judge	K	B	N	K	B	N	K	B	N
Self	.37	.27	.28	.36	.01	.19	.38	.53	.37
Peer	.55	.27	.35	.52	.00	.20	.58	.56	.49
Judges	.48	.35	.40	.39	.16	.37	.57	.54	.43
Mean	.47	.29	.34	.42	.05	.25	.51	.54	.43

Notes. K: Nearest Neighbor, B: Bayes, N: Neutral Network

The reliability (kappa scores) associated with each of the aforementioned classifiers was computed for each of the 3 data sets. K-fold cross-validation (k = 10) was run in tests with training and testing components. In k-fold cross-validation the data set (N) is divided into k subsets of

approximately equal size (N/k). The classifier is trained on (k-1) of the subsets and evaluated on the remaining subset. Accuracy statistics are measured after the process is repeated k times. The overall accuracy is the average of the k training iterations. Goutte (1977) has shown k-fold cross-validation to be superior to other techniques for small data sets. The Waikato Environment for Knowledge Analysis (WEKA, http://www.cs.waikato.ac.nz/ml/weka/) was used to perform the requisite computation.

Table 1 presents overall kappa scores as well as individual accuracy metrics in assessing the reliability by which the three machine learning schemes discriminated between boredom and flow when the affective states of a learner were categorized by three different judges. The results indicate that the classifiers were successful in discriminating between boredom and flow at rates significantly higher than chance. The simplest strategy of assigning emotion categories by consulting the neighbors of a test instance yielded performance scores that were quantitatively higher than algorithms that attempted to construct an explicit model from the data (i.e., Bayesian Classifiers and Neural Networks). Classifiers that were trained and evaluated with the data sets of peers and trained judges were on par and higher than those obtained by the self judges. This might be because the differences associated with back pressure (Figure 2a) and the changes in seat pressure (Figure 2d) were not significant when emotion judgments were provided by the self judges.

Discussion

This exploratory research provides new findings on the relationship between a learner's posture and the affective states related to engagement. Our results indicated that boredom is manifested in two distinct forms. The first is consistent with the preconceived notion of boredom in which a learner stretches out, lays back, and simply disengages. However, a counter-intuitive finding is that *boredom* was associated with a form of restlessness manifested by rapid changes in pressure on the seat of the chair. It is not clear as to whether these two bodily expressions of boredom are isolated, combine, or interact during experience. Finer grained analyses would be required to tease apart these alternatives.

The affective state of *flow* was associated with a heightened pressure in the seat of the chair with minimal movement. This may imply that the learner is mentally engaged in absorbing the material and thereby devotes a smaller amount of cognitive processing towards trivial bodily motion as explained by the proclivity-arousal framework.

It is interesting to note that posture predicts affect even though posture information was not on the radar of the judges during their ratings of the learners' affect states of the learners. Perhaps some basic bodily movement could conceivably be inferred from the video of the participants' faces. It is also unclear why the posture patterns associated with the peer and trained judges' affect ratings were more synchronous than the self judgments.

Uncertainty remains as to what exactly should be the gold standard for deciding what emotions a learner is truly having. Should it be the learner, a peer, or the trained judges? 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 peers or trained judges. One possibility is "social desirability;" self judgments are less accurate when posture indicates boredom because learners do not want to admit feeling bored. An alternative position is that the self judges were utilizing internal cues (such as recollection from episodic memory) in judging their emotions. These were obviously unavailable to the other judges. Therefore, the peer and trained judges were forced to rely on bodily measures in inferring the learner's affect. Perhaps a composite score from all viewpoints is most defensible.

This research has highlighted the efficacy of monitoring bodily measures of a learner as a viable channel to infer complex mental states. Some researchers have challenged the role of non-verbal behaviors in communicating affect (Trimboli & Walker, 1987). They argue that experiments that extoll the virtues of non-verbal communication of affect are typically plagued by problems related to experimenter bias (i.e., the nature of the stimuli or the intention of the experimenter are not camouflaged). However, our research on body movements and those of others (Mota & Picard, 2003) were conducted in ecologically valid settings in which no actors are used and no attempts were made to intentionally invoke affect.

Many questions remain unanswered in this exploration of body movements, emotions, and learning. Are the significant relationships between cognition and bodily movements predicted by the various theories of embodied cognition (Clark, 1997; Glenberg, 2005; de Vega, 2002)? To what extent can affect and cognition individually predict bodily activity? Does a combination of these channels increase their predictive power? Do relationships between cognition, affect, and bodily movement generalize above and beyond individual differences in experiencing and manifesting affect? Answers to these questions will help us explore theories of embodied cognition in addition to the synchronization of emotions with complex learning.

Acknowledgments

We thank our research colleagues at the University of Memphis and MIT, as well as Steelcase Inc. for providing us with the Tekscan Body Pressure Measurement System at no cost. This research was supported by the National Science Foundation (REC 0106965 , ITR 0325428, REESE 0633918). 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.

References

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., & Stolcke, A. (2002). Prosody-based automatic detection of annoyance

and frustration in human-computer dialog. In J. H. L. Hansenet al (Eds.), Proceedings of the International Conference on Spoken Language Processing (ICSLP'02) (pp. 2037-2039). Denver.

Barrett, L. F. (2006). Emotions as natural kinds? *Perspectives on Psychological Science, 1*, 28-58

Brown, A. L., Ellery, S., & Campione, J. C. (1998). Creating zones of proximal development electronically. In J. G. Greeno and S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 341-367). Mahway, NJ: Lawrence Erlbaum.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*:2-175-218.

Clark, A. (1997). *Being there: Putting brain body and world together again*. Cambridge, MIT Press.

Coulson, M. (2004). Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28, 117-139.

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.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper-Row: NY.

de Vega, M. (2002). Del significado simbólico al significado corpóreo. *Estudios de Psicología*, 23, 153-174. [From symbolic meaning to embodied meaning].

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,* 7-13. New York: AMC Press.

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.

Ekman, P. & Friesen, W.V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88-105

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

Ekman, P., & Friesen, W. (1972). Constants across culture in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Ferrari, M., Taylor, R., & VanLehn, K. (1999). Adapting Work Simulations for Schools. *Journal of Educational Computing Research, 21* (1), 25-53.

Forbes-Riley, K. and D. Litman. 2004. Predicting Emotion in Spoken Dialogue from Multiple Knowledge Sources. In Proceedings of the *Human Language Technology Conference* of the North American Chapter of the Association for Computational Linguistics, 201-208. Boston, MA.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding Language in Bodily States: The Case for Emotion. In R. Zwaan and D. Pecher (Eds.) The grounding of cognition: The role of perception and action in memory, language, and thinking. Cambridge: Cambridge University Press (2005)

Goutte, C. (1997). Note on free lunches and cross-validation. *Neural Computation*, *9*, 1211-1215.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the TRG (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research, 1*, 35-51.

Graesser, A. C., P. Chipman, B. C. Haynes, and A. Olney. 2005. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education* 48: 612-618.

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.

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). IEEE Computer Society.

Metcalfe, J., & Kornell, N. (2005) A Region of Proximal Learning Model of Study Time Allocation. Journal of Memory and Language, 52, 463-477.

Meyer, D. K., & Turner, J. C. (in press). Reconceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*.

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.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension –monitoring activities. *Cognition and Instruction 1*(2), 117-175

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann, Los Altos.

Rogoff, B. (1990). Apprenticeship in thinking: Cognitive development in social context. New York: Oxford University Press.

Schouwstra, S., & Hoogstraten, J. (1995). Head position and spinal position as determinants of perceived emotional state. *Perceptual and Motor Skills*, 81, 673-674.

Tekscan (1997). *Tekscan Body Pressure Measurement System User's Manual*. Tekscan Inc., South Boston, MA, USA.

Wallbott, N. (1998). Bodily expression of emotion European Journal of Social Psychology, 28, 879-896.