Sottilare, R & Proctor, M. (2012). Passively Classifying Student Mood and Performance within Intelligent Tutors. *Educational Technology & Society*, *15* (2), xx–xx.

Passively Classifying Student Mood and Performance within Intelligent Tutors

Robert A. Sottilare[1] and Michael Proctor[2]
[1]U.S. Army Research Laboratory – Human Research and Engineering Directorate
Orlando, Florida 32826, USA // robert.sottilare@us.army.mil
[2]University of Central Florida College, Department of Industrial Engineering and Management Systems (IEMS)
Orlando, Florida, 32816, USA // mproctor@mail.ucf.edu

Abstract

It has been long recognized that successful human tutors are capable of adapting instruction to mitigate barriers (e.g., withdrawal or frustration) to learning during the one-to-one tutoring process. A significant part of the success of human tutors is based on their perception of student affect (e.g., mood or emotions). To at least match the capabilities of human tutors, computer-based intelligent tutoring system (ITS) will need to "perceive" student affect and improve performance by selecting more effective instructional strategies (e.g., feedback). To date, ITS have fallen short in realizing this capability. Much of the existing research models the emotions of virtual characters rather than assessing the affective state of the student. Our goal was to determine the context and importance of student mood in an adaptable ITS model. To enhance our existing model, we evaluated procedural reasoning systems used in virtual characters, and reviewed behavioral and physiological sensing methods and predictive models of affect. Our experiment focused on passive capture of behaviors (e.g., mouse movement) during training to predict the student's mood. The idea of mood as a constant during training and predictors of performance are also discussed.

Keywords

Adaptive tutoring system, Mood, Affect, One-to-one tutoring; Passive measures

Introduction

The goal of this research was to develop an adaptable ITS conceptual model that includes appropriate inputs to determine the affective state of the student being tutored. In support of this goal, we surveyed methods to allow an ITS to classify the affective state (e.g., mood or emotional state) of the student through the passive sensing and interpretation of correlated student behaviors and their physiological responses during training. We investigated methods of classifying affect for virtual characters through procedural reasoning systems and adapted these methods to human students to classify the student's affective state. Understanding the student's affective state, the ITS can use that information along with other student data (e.g., knowledge and progress toward goals) to select instructional strategies (e.g., direction) to optimize learning and performance.

The research discussed below offers perspectives on: one-to-one tutoring; affect and learning; the need for ITS to be capable of "perceiving" student affect; the design limitations of current ITS; models of affect; and enhanced ITS models. The main body of our research contains two primary objectives: the review of methods to sense student behaviors unobtrusively (passively) so as avoid any negative impact on the learning process; and the interpretation of sensed behaviors to build a predictive model of student affect as a basis for make decisions on the delivery of instruction.

A Perspective on One-to-One Tutoring

An ongoing goal in the research and development of ITS has been to increase their adaptability to better serve student needs (Heylen, Nijholt, op den Akker & Vissers, 2003; Hernandez, Noguez, Sucar & Arroyo-Figueroa 2006 and Sottilare, 2009). "The basic tenet of intelligent tutors is that information about the user (e.g., knowledge, skill level, personality traits, mood level or motivational level) can be used to modify the presentation of information so that learning proceeds more efficiently." (Johnson & Taatgen, 2005, p.24).

ITS offer the advantage of one-to-one tutoring where instruction is delivered at a tailored pace based on competency and progress toward instructional goals. The value of expert, one-to-one, human tutoring vice group tutoring (i.e.

traditional classroom teaching) has been documented among students who often score 2.0 standard deviations higher than students in a conventional classroom (Bloom, 1984). Other advantages of one-to-one, human-tutoring over classroom settings is that students ask approximately 26.0 questions per hour versus less than 0.20 questions per hour in the classroom setting (Dillon, 1988; Graesser & Person, 1994). This higher rate of interaction provides additional learning opportunities for weaker students. Stronger students ask fewer questions, but these questions tend to be "deep-reasoning questions" (Person & Graesser, 2003).

Loftin, Mastaglio & Kenney, 2004 assert that while one-to-one human tutoring is still superior to ITS in general, the one-to-one human tutoring approach is neither efficient nor cost-effective for training large, geographically-distributed populations (e.g., military organizations or large multi-national corporations). Given large and potentially diverse student populations, one goal of our research was to identify methods to improve the adaptability of ITS to support one-to-one tutoring. We explored methods to increase ITS "perception" of the student's affective state and thereby theoretically increase the potential of the ITS to effectively adapt to a given student during one-to-one, computer-based instruction.

A Perspective on Affect and Learning

Linnenbrink and Pintrich (2002) identified a connection between affect and learning. They found that many students experience some confusion when confronted with information that does not fit their current knowledge base, but those in a generally positive affective state adapted their known concepts to assimilate the new information. Students in a generally negative state usually reject this new information. This infers the need for tutors (human or otherwise) to be able to perceive and address the affective state of the students when formulating instructional strategies.

Fundamental to our research is an understanding of affect and how it might be modeled within ITS. Affect includes personality traits, mood and emotions which vary in duration, influence and cause (Gebhard, 2005). Affect is evident through student behaviors and physiological changes and may also be measured via self-report surveys. Each method has limitations. Self-report measures may be biased by the student's desire to conform to expectations and may not accurately reflect their true affective state. Physiological measures can vary from person to person and may be misinterpreted. Hybrid approaches that use combinations of these measures to confirm affect are considered more reliable. Being able to accurately predict affect is the first step in using it to select appropriate instruction.

An ITS should choose the content and the method of instruction based on both the student's competence and affective state (e.g., emotions) in the same way "an experienced human tutor manages the emotional state of the student to motivate him and to improve the learning process" (Hernandez, et al, 2006). Rodrigues, Novais & Santos (2005) affirmed that an ITS "must be capable of dynamically adapting and monitoring each student". This highlights the need for computer-based tutors to be able to perceive student state and use this information in formulating instruction.

Design Limitations of ITS

Affect has been modeled extensively within ITS (Core, et al, 2006; Heylen, et al, 2003; and Graesser, et al, 2001), but have generally been used to represent the tutor's affect rather than perceiving affect in students. Incorporating affective perception within ITS is recognized as a key element in the learning process (Bickmore & Picard, 2004; Burleson & Picard, 2004; Johnson, Rickel and Lester, 2000; and Picard, et al, 2004), but Picard (2006) notes the design limitations of ITS including their inability to: recognize student affect; respond appropriately to the student affect; and appropriately express emotion. Picard suggests that the "simplest set" of emotions for an ITS to recognize are the emotions each of us is born with: pleasure, boredom and frustration. The ability to sense and predict even this small set of emotions would greatly expand the capability of ITS to adapt instruction to the needs of the student and optimize learning and performance.

Further expanding the need for ITS to perceive student affect and adapt instruction, Alexander, Sarrafzadeh and Fan, (2003) argued that "an important factor in the success of human one-to-one tutoring is the tutor's ability to identify and respond to affective cues given by the student". We also expected that ITS that model affect would be more effective than equivalent ITS that do not model affect.

Picard (2006) asserted that “no matter how intelligent a tutor is, it will eventually become annoying if it does not have emotional intelligence”. Emotional intelligence is defined as “a set of skills hypothesized to contribute to the accurate appraisal and expression of emotion in oneself and in others, the effective regulation of emotion in self and others, and the use of feelings to motivate, plan, and achieve in one’s life” (Salovey & Mayer, 1990). In other words, ITS that have emotional intelligence are capable of: recognizing the student affect; making the student aware so he can participate in managing his affective state; providing options (e.g., strategies) for the student to control his affective state; using emotion to motivate the student to achieve established objectives.

Today’s ITS provide limited adaptive training capabilities. Sessink, et al (2007) assert that “most traditional learning material [for ITS] targets the ‘average student’, and is suboptimal for students who lack certain prior knowledge, or students who have already attained some of the course objectives”. Csikszentmihalyi (1990) stated that ITS should match “adequate challenge with skill in service of Flow, or optimal experience” within the learning system. Failure to provide adequate challenge could result in boredom or frustration. It is desirable for the ITS to know everything possible about the student’s affect to craft a tailored learning experience.

ITS MODELS

We chose to begin with a documented standard ITS model. Beck, Stern and Haugsjaa’s (1996) Intelligent Tutoring System Model (shown in Figure 1) illustrates the interactions ITS components including: domain knowledge which provides basis for course content, a student model which generally contains information about the progress of the student relative to learning objectives; an expert model which defines the performance of an ideal student; a pedagogical module which makes decisions on instruction and feedback to the student; and finally, a communication interface which presents instruction and feedback to the student. Given the goal to expand the student model to include affect and optimize decisions on instruction, Beck’s ITS model has some significant limitations including the inability to sense the student’s behavior and physiological state.

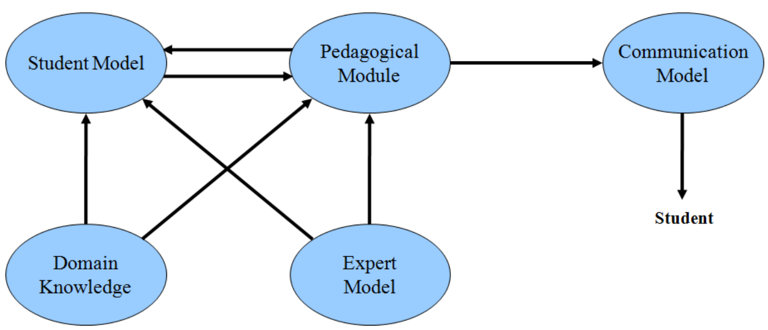

Figure 1: Beck, Stern and Haugsjaa’s (1996) Intelligent Tutoring System Model

Based on the current limitations of ITS and the emphasis in the literature on understanding the affective state of the student in one-to-one tutoring situations, Sottilare (2009) adapted Beck, Stern and Haugsjaa’s (1996) ITS model as shown in Figure 2 to include the student’s affective state and methods to assess affect (e.g., behavioral, physiological measures). This model enhanced Beck, Stern and Haugsjaa’s model by adding capabilities to sense behaviors and physiological data, and transfer this data to the student model to aid in assessing the student’s affective state; and by adding a real-time flow of affective state data from the student model to the pedagogical model so the student’s affective state can play a part in instructional strategy decisions. Sottilare’s (2009) model was used as the foundation for integrating affective classification models based on procedural reasoning systems as discussed below.

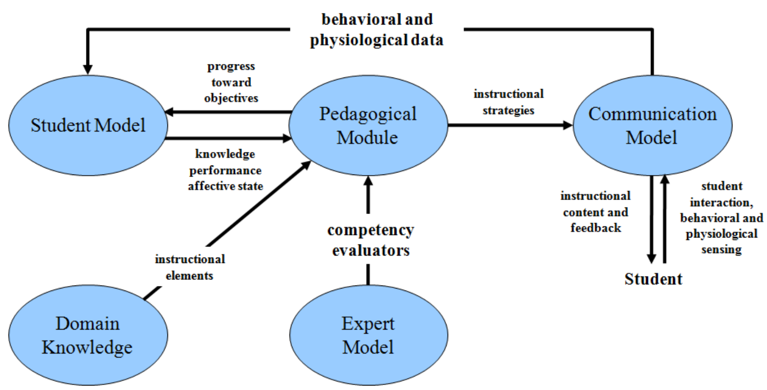

Figure 2: Sottilare's (2009) Intelligent Tutoring/Coaching Model

Procedural Reasoning Systems

Like the logic within ITS, procedural reasoning system (PRS) for virtual humans act as intelligent agents which manage processes (e.g., sensing processes). For this reason, we decided to evaluate the sensing and reasoning processes of PRS to see if there were any transferrable methods for use in ITS. We evaluated several existing models used in embodied characters and their evolution in providing more complex interactions within intelligent agents. We also sought to reuse key elements and adapt these models to support a comprehensive ITS model that includes affect.

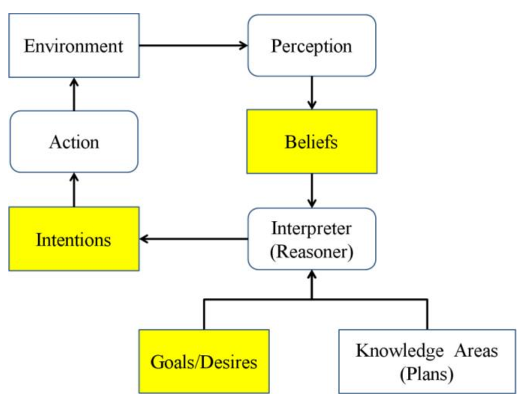

Figure 3: Procedural Reasoning System (Georgeff and Lansky, 1987)

First, we considered the basic procedural reasoning system (PRS) model developed by Georgeff and Lansky (1987). This model is based on a belief-desire-intention framework as illustrated in Figure 3 and includes the following states and processes: beliefs state are facts about the world which are perceived to be true; goals/desires state are a set of needs/wants to be realized; knowledge areas are declarative procedure specifications or sequences; interpreter process is the analysis of relevant goals/desires, beliefs and a set of relevant knowledge areas to determine intentions; intentions state is a set of planned actions; action process include actions taken in order to achieve a goal or satisfy a desire; environment state include any external influencing factors; and finally, the perception process which senses changes in the environment after actions take place.

Parunak, et al, 2006 enhanced Georgeff and Lansky's model with the addition of OCC model of emotions (Ortony, Clore & Collins, 1988). In this enhanced model, beliefs feed an appraisal process as well as the interpreter. The appraisal process determines which of the 22 emotions defined in the OCC model to output to the perception and interpreter processes. Parunak, et al, (2006) expanded the PRS model again with the addition of disposition, "a parameter that distinguishes the varying susceptibility of different agents to various emotions". Parunak considered disposition to be constant over the time horizon of the simulations that they used for experimentation. For our purposes, we considered disposition in terms of a student's varying susceptibility to various emotions. In this light, disposition might be influenced by factors like personality, mood or competency in the learning domain in which the student is engaged.

As noted previously, Linnenbrink and Pintrich (2002) identified a connection between affect and learning. For instance, a personality factors like openness might influence a student's disposition to learning new information. Competency might influence confidence and thereby dispose a student toward specific emotions like frustration if the instruction presented is not in line with their competency level. Therefore, an enhanced student model within an ITS that includes affect along with passive, real-time sensory inputs (e.g., behavioral and physiological measures) would make that ITS better able to adapt to a student's needs. In addition to its effect on perception, note that changes in emotion might also positively or negatively influence desires and plans in this enhanced student model.

Default values for personality, mood and competency are based on instruments like the Big Five Personality Test, the Self-Assessment Manikin (SAM) for mood (Lang, 1980) and a skills/knowledge pre-test to determine initial competency. Emotions are assessed probabilistically by frequent appraisals of behaviors and physiological measures. Competency might also be assessed periodically to determine if the instruction is having the desired effect.

Finally, integrating these concepts with the Intelligent Tutoring/Coaching Model in Figure 2, we see the potential for a more adaptive ITS model in

Figure 4 that accounts for affect, competency, goals/desires, plans and beliefs to tailor instruction to each student's needs, limitations and abilities. This adaptive ITS model uses real-time behavioral and physiological sensing to allow the ITS to "perceive" affective state and project changes in beliefs, desires and intentions.

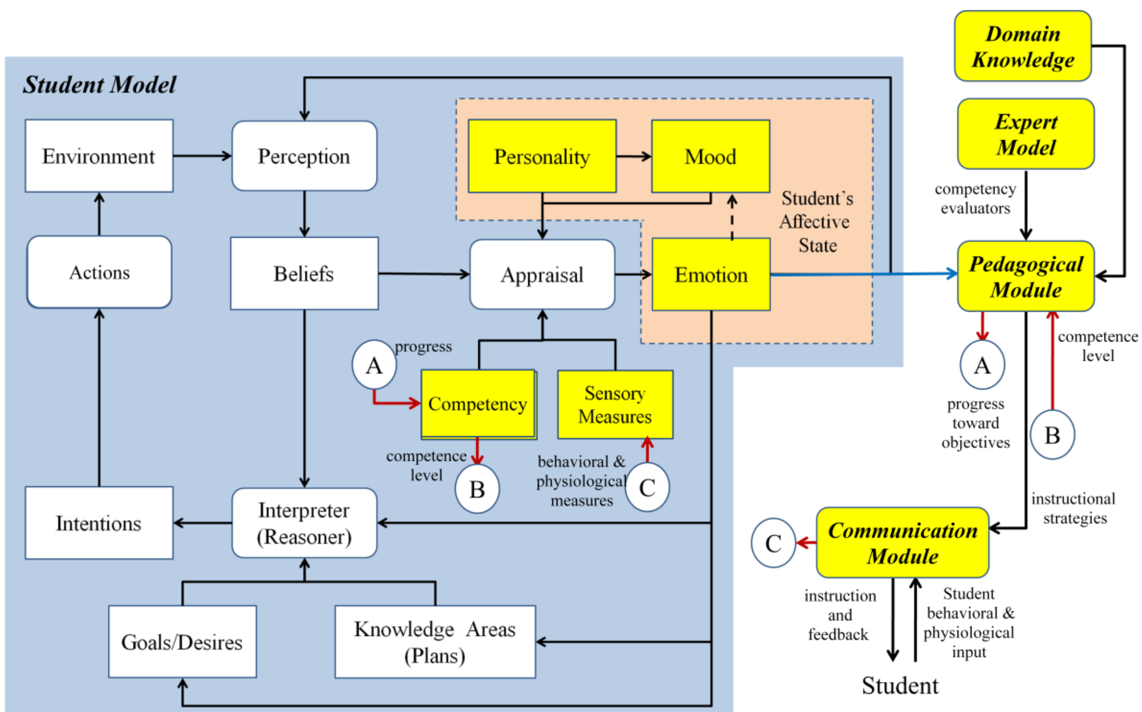

Figure 4: Adaptive ITS with enhanced Student Model

ASSESSING AND MODELING AFFECT

Given the importance we have placed on assessing affect in our adaptive ITS model, the following defines affect and reviews methods to assess and model affect.

Contrasting Personality, Mood and Emotions

Personality has long-term affect and reflects individual differences in cognitive and social processes. The Five Factor Model (FFM) of personality (McCrae & John, 1992) defines personality traits in terms of openness, conscientiousness, extraversion, agreeableness and neuroticism; all of which tend to be relatively stable in adult populations. Personality moderates the speed and degree to which mood and emotions emerge and dissipate.

Moods generally have a moderate duration as opposed to emotions (short duration) and personality (long duration). Mood has a subtle influence on cognition which may go unnoticed even by human tutors and are of unknown cause so they may be realized as the cumulative effect of a series of emotional events. According to Morris (1989), mood has "a great influence on human's cognitive functions" and thereby learning. Mood is represented by Mehrabian's (1996) Pleasure-Arousal-Dominance (PAD) model. Mood moderates emotional change and may be influenced by series of emotional events.

Emotions are a psychological state or process that functions in the management of goals and needs of an individual (Broekens & Degroot, 2004). Davidson (1994) argued that "emotions bias action, whereas moods bias cognition." That infers that emotions may inhibit appropriate actions or cause unintended actions (student behavior), while mood might affect perception and reasoning (input and processing) by the student during the learning process. Emotions are represented by Ortony, Clore & Collins' (1988) twenty-two emotional states (e.g., joy, pride).

Relationship between Mood and Personality

Mehrabian (1996) defined mathematical relationships between the PAD dimensions and the FFM factors of personality. We made the assumption that since personality is stable in healthy adults (McCrae and Costa, 1994) and mood is of moderate duration, that mood may be considered to be stable for training sessions of moderate duration (30 minutes to one hour). We tested mood for stability in our experiment and discuss the validity of this assumption in the results section.

Personality Assessment Methods

Again, since personality traits have been determined to be relatively stable in healthy adults (McCrae and Costa, 1994), we assume that a single evaluation is all that is required to assess the FFM traits. The Big Five Personality Test (John, 2003) is a representative instrument that measures the five FFM dimensions of openness, conscientiousness, extraversion, agreeableness and neuroticism on 0-100 scale.

Mood and Emotion Assessment Methods

Mood and emotion assessments tend to fall into one of three categories (self-report instruments, behavioral indicators and physiological indicators) and may be combined to provide a more accurate assessment of the student's affective state. Lang's (1980) Self Assessment Manikin (SAM) is a self-report instrument composed of a nine-point Likert scale for each mood dimension (pleasure, arousal and dominance). Affective prediction methods include the use of: behavioral measures (e.g., facial landmark sensing, keystroke and mouse movement detection); physiological measures (e.g., heart rate inter-beat interval sensing or galvanic skin response); and other student data (e.g. demographics). Examples of related affective computing research discussed below.

Neji & Ben Ammar (2007) investigated an ITS that included a virtual character and facial sensors. In addition to the two-way conversational input and output, the tutor inferred the emotional state of the student through a machine vision system which sensed changes in distances between facial landmarks and classified their expressions as one of six universal emotional states (joy, sadness, anger, fear, disgust and surprise) or a neutral expression. The student's emotional state was then used in the ITS to determine which tutoring strategy (e.g., sympathizing or non-sympathizing feedback, motivation, explanation, steering). This study took significant steps toward an adaptable tutoring system, but two primary limitations are noteworthy: the cost and portability of the visual system would likely limit the ability of the ITS to support large, distributed populations; and it failed to assess the connection between student affect, the selection of instructional strategies and performance.

D'Mello, Craig , Sullins and Graesser (2006) and D'Mello and Graesser (2007) used frequent conversation patterns to predict affect (i.e. confusion, eureka, frustration) when students emote aloud and their ITS provided feedback, pumps, hints and assertions to influence student's progress. They noted significant relationships between tutor feedback (negative, neutral or positive) and the student's affect. Positive feedback by the ITS was a strong positive predictor of eureka, "a feeling used to express triumph on a discovery" (D'Mello, et al, 2006). The "directness" of the tutor predicted a negative relationship with confusion. Negative feedback from the tutor was the only significant predictor of frustration. The primary drawback to this approach was the requirement for students to "emote aloud" which has some of same drawbacks as other self-report methods and may be incompatible with students with lower FFM openness scores. Another drawback was the low participant throughput for the experiment based on the labor intensive nature of the data collection and analysis. Given the variability among students and the associated time to baseline each student, this method is unsuitable as is for training in large-scale, distributed organizations.

Zimmermann, Guttormsen, Danuser & Gomez (2003) used passive methods to capture computer keystrokes and mouse movement as indirect indicators of affect. Their method falls short in that it does not determine any correlation between computer keystrokes or mouse movement and specific affective states (e.g., happy, sad or frustrated). Their approach also fails to follow up in adapting coaching strategies within tutoring systems to the student's affective state once it is determined.

Yun, Shastri, Pavlidis and Deng (2009) demonstrated passive sensing and interpretation of thermal images to estimate student stress levels. They altered the difficulty levels of game play for users based on singular input from StressCam, which monitors heat dissipation through a thermal imaging-based camera and analysis system. Since stress levels are related with increased blood flow in the forehead and higher blood flow equates to more heat, StressCam passively and continuously senses and interprets thermal images. No specific affective variables were predicted in this research and no evidence was provided regarding the use of this technology for determining instruction. Setup time, system cost and lack of portability may be an issue with this approach.

Whereas many of the machine perception methods reviewed above focus on single data sources (e.g thermal imaging) to predict affect, others pursued multi-source approaches. McIntyre and Göcke (2007) advocated a multimodal approach to reduce the uncertainty associated with physiological measurements.

Conati and Maclaren (2004) inferred six emotions (joy-distress, pride-shame, and admiration-reproach) of the twenty-two emotions defined in the OCC model (Ortony, Clore & Collins, 1988) by assessing their reaction to meeting/not meeting goals during a computer game. They used a probabilistic model, the dynamic decision network (DDN), which used: personality traits; interaction patterns; electromyography (EMG) which measures muscle activity; and skin conductance to indirectly assess student emotions. The drawbacks to this approach are its accuracy (no significant difference from the baseline accuracy of 50%), and inclusion of only 6 of 22 OCC emotions. It also lacks a student performance assessment to gauge the effectiveness of the selected instructional strategies. Finally, the physiological sensing of valence produced false negatives 55% of the time.

Hernandez, Noguez, Sucar & Arroyo-Figueroa (2006) also developed a Bayesian affective student model based on student personality traits, student knowledge state, student goals and the tutoring situation to infer the same six emotions. A weakness of this approach is in its performance as a tutor as compared to human tutors. Pedagogical actions of their model of instruction only agreed with the pedagogical actions of expert human tutors 57% of the time.

Based on approaches cited in the literature and the adaptive tutoring model discussed above, we decided to pursue a passive approach to detecting mood in the context of a self-paced training environment with an embedded tutor. We chose to assess the influence of student state variables (e.g., competency) and student behaviors as predictors of mood state.

HYPOTHESES UNDER TEST

Three hypotheses were evaluated in this research. In reviewing the literature, there was little beyond self-report methods used as indicators of student mood. Given the obtrusiveness of self-reporting methods and concerns that frequent assessments would interfere with the learning process, we determined to assess passive predictors of mood and performance, and to test the assumption of mood as a constant during training.

Hypothesis "A": Mood Prediction

In hypothesis A, we considered predictors of student mood variables. Mood variables were selected as dependent variables over other affective variables (personality or emotion) based on their duration, influence and cause. Mood was also selected because of its influence on cognition. Davidson (1994) argued that "emotions bias action, whereas moods bias cognition" which indicated that mood would likely affect perception and reasoning during learning. Based on this connection, we considered the importance of accurately predicting mood variables and posed the following hypothesis: "student state variables (e.g., amount of sleep, energy level, a priori knowledge level and interest in the topic), student action variables (e.g., mouse movement rates and control-selection rates) and student performance are predictors of student mood variables".

Hypothesis "B": Mood Stability

In hypothesis B, we considered the assumption that mood is relatively constant over moderate periods of time (e.g., a training session lasting thirty to sixty minutes). We considered the minimum number of mood assessments required to accurately reflect the student's mood and still maintain the continuity of the learning process. Would only one or two assessments be required or would mood have to be assessed more frequently or even continuously? To determine stability of mood, we posed the following hypothesis: "student mood variables (pleasure, arousal and dominance) are generally constant during a 30-60 minute training session".

Hypothesis "C": Performance Prediction

In hypothesis C, we considered predictors of student performance and posed the following hypothesis: "student state variables, student action variables and student mood variables are predictors of student performance".

EXPERIMENTAL METHODS

Participants

Cadets from the United States Military Academy (USMA) were the target population for this study. The experimental design focused on the target population with low-moderate competence in tactical combat casualty care (TC3). One hundred-thirty one (131) cadets participated in the study which produced 124 instances of usable data. Of the 124 participants, 108 were males (age M = 18.79, SD = 0.98) and 16 were females (age M = 18.38, SD = 0.62). The 124 participants in this study are approximately 17.7% of the available participants in this population.

Experimental Procedure

Our experimental approach is based on a case study which involved instructing cadets about the elements of hemorrhage control using a multi-media training package called Tactical Combat Casualty Care (TC3). The content of this study was delivered to the participants via a multimedia presentation on a laptop computer. The study content included a biographical survey, a TC3 training course, a pre-course and post-course knowledge assessment, a pre-training and a post-training mood assessment, and a TC3 performance assessment. Interactive controls, coaching strategy logic, timestamp events and automatic recording of each participant's interaction and survey data was accomplished through embedded software. The experimental process is shown in Figure 5 below.

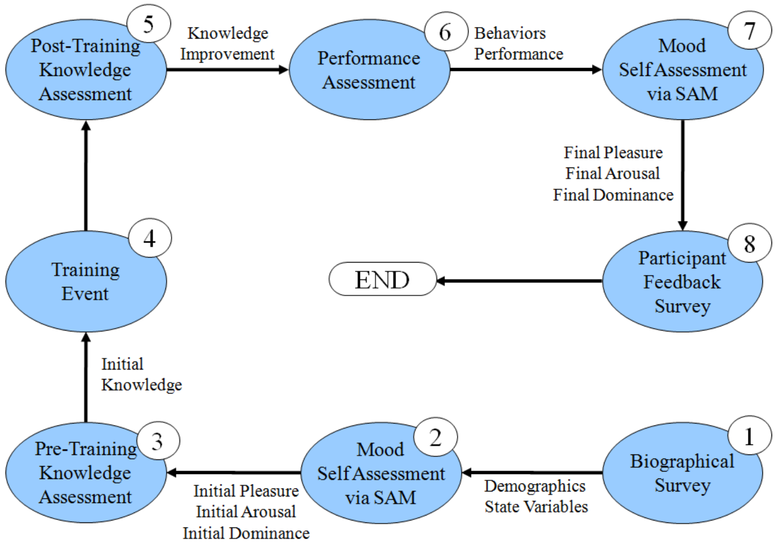

Figure 5: Experimental Procedure

Student state information (e.g., age, gender, energy level, amount of sleep, first aid experience and interest level in TC3) was collected through a biographical survey. The Self-Assessment Manikin (SAM) survey (Lang, 1980) was administered twice to each participant: once prior to the training course and once again after the completion of the TC3 performance assessment. Knowledge assessments were administered prior to the delivery of instruction and after the delivery of instruction.

The training course focused on acquiring knowledge (facts and principles) while the performance assessment focused on applying knowledge and demonstrating skill through decisions and actions. Figure 6 shows a screen capture from the performance assessment which is time sensitive as the student must appropriately treat the wounds within the approximately four minutes for the soldier to bleed out and die.

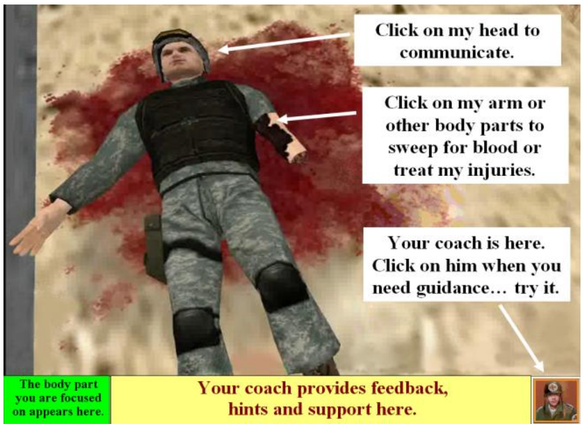

Figure 6: Screen capture from TC3 performance assessment

RESULTS

Hypothesis "A": Mood Prediction

We posed the following hypothesis: "student state variables, student action variables (e.g., behaviors) and student performance are predictors of student mood variables". We analyzed specific student state and behavior variables with respect to the three mood variables: pleasure, arousal and dominance.

Final pleasure measurements were positively correlated with *self-assessed first aid knowledge*, Pearson's r(124) = .177, p = .049; *TC3 interest level*, Pearson's r(124) = .178, p = .048; *initial pleasure*, Pearson's r(124) = .272, p = .002; *initial arousal*, Pearson's r(124) = .194, p = .031; *initial dominance*, Pearson's r(124) = .266, p = .003; *final dominance*, Pearson's r(124) = .541, p < .001; *control selection rate*, Pearson's r(124) = -.182, p = .043; and *performance*, Pearson's r(124) = .177, p = .012.

Final arousal measurements were positively correlated with *initial pleasure*, Pearson's r(124) = .352, p < .001; and *initial arousal*, Pearson's r(124) = .507, p < .001.

Final dominance measurements were positively correlated with *TC3 interest level*, Pearson's r(124) = .233, p = .009; *initial arousal*, Pearson's r(124) = .177, p = .049; *initial dominance*, Pearson's r(124) = .578, p < .001; *final pleasure*, Pearson's r(124) = .541, p < .001; and *performance*, Pearson's r(124) = .393, p < .001.

Examining the significance of the variables and their influence on each mood variable, we developed the models shown in Tables 1 and 2 via linear regression. *Initial pleasure*, *final dominance* and *control selection rate* explained a significant portion of variance in final pleasure scores, R^2 = .36, F(3, 120) = 22.77, p < .001, d = 0.57. *Initial dominance*, *control selection rate*, *performance* and *final pleasure* explained a significant portion of variance in *final dominance scores*, R^2 = .54, F(4, 119) = 35.23, p < .001, f^2 = 1.18. The regression effect sizes (f^2) indicated large effect for both models per Cohen's (1992) guidelines.

Table 1: Regression Model of Pleasure

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	2.096	.854		2.453	.016	.405	3.788
	InitialPleasure	.230	.099	.173	2.323	.022	.034	.426
	FinalDominance	.519	.073	.523	7.086	.000	.374	.664
	ControlSelectionRate	-.090	.037	-.182	-2.472	.015	-.163	-.018

Table 2: Regression Model of Dominance

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	-2.130	.830		-2.566	.012	-3.773	-.486
	InitialDominance	.518	.079	.436	6.531	.000	.361	.675
	ControlSelectionRate	.068	.032	.136	2.101	.038	.004	.132
	Performance	.021	.009	.155	2.304	.023	.003	.039
	FinalPleasure	.417	.067	.415	6.233	.000	.285	.550

Hypothesis "B": Mood Stability

We posed the hypothesis: "student mood variables (*pleasure, arousal and dominance*) are generally constant during a 30-60 minute training session". A non-directional student's t-Test (α = .05) was used to compare the participant's mood measurements at the beginning and end of the experiment. This comparison revealed a statistically reliable difference between initial (M = 6.03, SD = 1.22) and final (M = 5.36, SD = 1.63) pleasure measurements, t = 1.98, p < .01, d = .47; and a statistically reliable difference between initial (M = 4.05, SD = 1.74) and final (M = 5.28, SD = 1.64) arousal measurements, t = 1.98, p < .01, d = .73. For dominance, there was no statistically reliable difference between initial and final measurements.

This indicates that our assumption that mood variables remain constant during a 30-60 minute training session are not supported for pleasure and arousal variables, but may be valid for the dominance variable. The effect sizes (Cohen's d) indicated the research environment had medium effect for pleasure and a large effect for arousal per Cohen's (1992) guidelines.

Hypothesis "C": Performance

We posed the following hypothesis: "student state variables, student action variables and student mood variables are predictors of student performance". We analyzed specific student state, behavior and mood variables with respect to student performance. *Performance* measurements were positively correlated with *initial dominance*, Pearson's r(124) = .289, p = .001; *final pleasure*, Pearson's r(124) = .224, p = .012; *final dominance*, Pearson's r(124) = .393, p < .001; and *mouse movement rate*, Pearson's r(124) = .251, p = .005.

Examining the significance of the variables and their influence on performance, we developed the model shown in Table 3 via linear regression. *Initial dominance*, *mouse movement rate* and *final knowledge* scores explained a significant portion of variance in performance scores, $R^2 = .23$, $F(3, 120) = 11.96$, $p < .001$, $f^2 = 0.30$. The regression effect size (f^2) indicated moderate to large effect per Cohen's (1992) guidelines.

Table 3: Regression Model of Performance

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	32.815	8.317		3.945	.000	16.347	49.283
	InitialDominance	2.917	.716	.330	4.076	.000	1.500	4.333
	MouseMovementRate	.062	.018	.281	3.462	.001	.026	.097
	FinalKnowledge	.247	.081	.245	3.054	.003	.087	.407

DISCUSSION, CONCLUSIONS AND FUTURE RESEARCH

This research examined whether mood could be passively predicted, whether mood could be considered as a constant and the relationship of mood and other student variables were predictors of performance. The objective was to discover relationships that could expand the capability of ITS student models.

Predictive models of pleasure and dominance were developed via regression analysis. Although it was expected that mouse movement and control selection rates would be direct indicators of arousal, there were no reliable predictors of arousal generated during our experimentation.

Mood, which we expected to be stable over the course of a 30-60 minute training session, was significantly different for pleasure and arousal while student's dominance was relatively unaffected by their experience in the training research environment. Pleasure, arousal and dominance changed by an average of -25%, 58% and 18% respectively. The significant changes in pleasure and arousal may be directly connected to the tactical combat casualty care scenario that the students were exposed to during the experiment.

While we did discover reliable predictors of performance, we were surprised that previous experience, training and topic interest were not among those predictors of performance. Experience and training were used as indicators of competence in our adaptive training model. Given the generally low competence level of the sample population, we should have predicted that these variables would not be reliable predictors of performance.

Additional research is needed to fully validate the full adaptive tutoring model discussed herein, but relevant results in predicting future mood states (pleasure and dominance) were realized along with significant predictors of performance. A regression model of future pleasure and dominance states was developed that relied upon few predictors (3 for pleasure and 4 for dominance). A regression model of performance was developed and includes only 3 predictors to account for 23% of the variance in performance. This finding passes the "common sense" test as demonstrated by: high initial dominance which was an indicator of trainee comfort level in working in a new domain and resulted in higher performance scores; high mouse movement rates which indicated a comfort level in working quickly with the user interface; and high final knowledge scores which indicated that the trainee's absorbed the lessons and were ready to apply them in the performance assessment test. The findings in this paper also confirmed findings of Davidson (1994) by linking mood to performance outcomes.

The experimental results reviewed in this article provide much needed production rules to support intelligent tutor instructional strategy decision-making based on student mood attributes and other variables of interest whereas other similar work is focused either on cognitive attributes and/or emotions. Mood was chosen based on its duration which is similar in length (30-60 minutes) as many training sessions. In contrast emotional states may only last a few minutes. Production rules based on mood afford the tutoring system the opportunity to assess and update mood state much less often than production rules based on emotional states.

The results discussed herein are significant in that they were drawn from a significant sample size (power > 0.99). The application of the algorithms developed and discussed in this paper will allow for more accurate classification of student mood and thereby allow computer-based tutors to optimally match instruction (challenge level, flow and feedback) to the student's affective and cognitive needs. This will improve the learning effectiveness of computer-based tutors so they may successfully substitute for more observant human tutors.

While statistically relevant relationships were discovered in an experimental context, the models discussed herein will ultimately be deployed as part of a learning system that integrates and evaluates multiple learning variables. How these results operate as part of a learning system is yet to be determined by empirical evaluations over the long term, but their application in tutors has already begun. Additional unobtrusive methods are being evaluated.

The authors recommend that research be undertaken to develop a more robust model of the "affective sensing-assessment-instructional strategy selection" process. Additional low-cost, passive sensing techniques should be evaluated on a variety of learning and even mobile learning platforms to push the limits of student modeling and instructional strategy selection. We see another research challenge in the sensing of multiple, simultaneous emotions and we also recommend the continued evaluation of techniques to assess performance, beliefs and competency. We also see the need to enhance predictive models through more continuous measures vice limited discrete measures at the beginning and end of key events. Finally, we see the need for a standard set of instructional strategies that are topic independent so they can be implemented across intelligent tutoring systems.

REFERENCES

Alexander, S., Sarrafzadeh, A., & Fan, C. (2003). Pay Attention! The Computer is Watching: Affective Tutoring Systems. In A. Rossett (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003* (pp. 1463-1466). Chesapeake, VA: AACE.

Beck, J., Stern, M., & Haugsjaa, E. (1996). Applications of AI in Education, *ACM Crossroads*, 3(1).

Bickmore, T. & Picard, R.W. (2004). Towards Caring Machines. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Vienna, Austria.

Bloom, B. S. (1984). The 2-Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher,* 13(6), 4-16.

Broekens, J. & DeGroot, D. (2004). Scalable and Flexible Appraisal Models for Virtual Agents. In *Proceedings of the International Conference on Computer Games, Artificial Intelligence,Design and Education (CGAIDE)*, (pp. 208–215). Reading, UK.

Burleson, W. & Picard, R. (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In *Social and Emotional Intelligence in Learning Environments Workshop In conjunction with the 7th International Conference on Intelligent Tutoring Systems*. Maceio - Alagoas, Brasil.

Cohen, J. (1992). Quantitative Methods in Psychology: A Power Primer. *Psychological Bulletin*, 112(1), 155-159.

Conati, C. & Maclaren, H. (2004). Evaluating a Probabilistic Model of Student Affect. In J.C. Lester, R.M. Vicari, & U. F. Paraguac (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science Volume 3220* (pp. 55-66). Berlin: Springer-Verlag.

Core, M. G., Traum, D., Lane, H. C., Swartout, W., Marsella, S., Gratch, J., & van Lent, M. (2006) Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82(11), 685-701.

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper Perennial.

Davidson, R.J. (1994). On emotion, mood, and related affective constructs. In P. Ekman & R.J. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions* (pp. 51-55). New York: Oxford University Press.

D'Mello, S.K., Craig, S.D., Sullins, J., & Graesser, A.C. (2006). Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor's Mixed-Initiative Dialogue. *International Journal of Artificial Intelligence in Education*, 16, 3-28.

D'Mello, S. & Graesser, A. (2007). Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)* (pp 161-168). Amsterdam, The Netherlands: IOS Press.

Dillon, T.J. (1988). *Questioning and Teaching: A Manual of Practice*. New York: Teachers College Press.

Gebhard, P. (2005). ALMA - A Layered Model of Affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS-2005)* (pp 29-36). New York: ACM.

Georgeff, M. & Lansky, A. (1987). Reactive Reasoning and Planning. In *Proceedings of the 6th National Conference on Artificial Intelligence* (pp. 677–682). Seattle, Washington.

Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.

Graesser, A. C., Person, N. K., Harter, D., & Tutoring Research Group. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12(3), 257-279.

Hernandez, Y., Noguez, J., Sucar, E., & Arroyo-Figueroa, G. (2006). Incorporating an Affective Model to an Intelligent Tutor for Mobile Robotics. In *Proceedings of the 36th Annual Frontiers in Education Conference* (pp. 22-27). San Diego, California.

Heylen, D., Nijholt, A., op den Akker R., & Vissers, M. (2003). Socially Intelligent Tutor Agents. In T. Rist, R. Aylett, D. Ballin, & J. Rickel (Eds), *Proceedings of Intelligent Virtual Agents (IVA 2003), Lecture Notes in Computer Science Volume 2792* (pp. 341-347). Berlin: Springer.

John, O.D. & Atof Inc. (2003). The Big Five Personality Test. [Website]. Retrieved October 23, 2006 from http://www.outofservice.com/bigfive/.

Johnson, A. & Taatgen, N. (2005). User modeling. In R. W. Proctor & L.V. Kim-Phuong (Eds.), *The Handbook of Human Factors in Web Design* (pp. 424-438). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, W.L., Rickel, J.W., & Lester, J.C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environment. *International Journal of Artificial Intelligence in Education*, 11, 47-78.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex.

Linnenbrink, E. A. & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, 31, 313-327.

Loftin, B., Mastaglio, T., & Kenney, P. (2004), *Outstanding Research Issues in Intelligent Tutoring Systems*, study commissioned by the Research Development and Engineering Command (RDECOM), Orlando, Florida, USA under contract N61339-03-C-0156, Retrieved October 23, 2006 from http://www.mymicsurveys.com/site/files/pub_4.pdf.

McCrae, R.R., & John, O.P. (1992). An introduction to the five-factor model and its applications. Special Issue: The five-factor model: Issues and applications. *Journal of Personality*, 60, 175-215.

McCrae, R.R. & Costa, P.T. (1994). The stability of personality: observations and evaluations. *Current Directions in Psychological Science*, 3, 173–175.

McIntyre, G. & Göcke, R. (2007). Towards Affective Sensing. In *Proceedings of the 12th International Conference on Human-Computer Interaction (HCII 2007), Lecture Notes in Computer Science Volume 4552*, (pp. 411-420). Berlin: Springer.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14, 261-292.

Morris W. N. (1989). *Mood: The frame of mind*. New York: Springer-Verlag.

Neji, M. & Ben Ammar, M. (2007). Agent-based Collaborative Affective e-Learning Framework. *Electronic Journal of e-Learning,* 5(2), 123 – 134.

Parunak, H., Bisson, R., Brueckner, S., Matthews, R., & Sauter, J. (2006). A Model of Emotions for Situated Agents. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '06)* (pp. 993-995). New York, NY: ACM.

Person, N. K., Graesser, A. C., & The Tutoring Research Group (2003). Fourteen facts about human tutoring: Food for thought for ITS developers. In *Artificial Intelligence in Education 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a View Toward the Classroom* (pp. 335-344). Sydney, Australia.

Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., et al. (2004). Affective learning - a manifesto. *BT Technology Journal,* 22(4), 253-269.

Picard, R. (2006). *Building an Affective Learning Companion*. Keynote address at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan. Retrieved from http://www.its2006.org/ITS_keynote/ITS2006_01.pdf

Rodrigues, M., Novais, P., & Santos, M.F. (2005). Future Challenges in Intelligent Tutoring Systems – A Framework. In A. Méndez-Vilas, B. González-Pereira, J.A. Mesa González, (Eds.), *Recent Research Developments in Learning Technologies, Proceedings from the International Conference on Multimedia, Information and Communications Technologies in Education*, (pp. 929-934). Cáceres, Spain.

Salovey, P. & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, & Personality*, 9, 185-211.

Sessink, O., Beeftink, H., Tramper, J., & Hartog, R. (2007). Proteus: A Lecturer-Friendly Adaptive Tutoring System. *Journal of Interactive Learning Research*, 18(4), 533-554.

Sottilare, R. (2009). Making a case for machine perception of trainee affect to aid learning and performance in embedded virtual simulations. In *Proceedings of the NATO Workshop on the Human Dimensions of Embedded Virtual Simulation*, Orlando, Florida.

Yun, C., Shastri, D., Pavlidis, I. & Deng, Z. (2009). *O'game, can you feel my frustration? Improving user's gaming experience via stress-cam.* In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (ACM CHI)*, (pp. 2195-2204). Boston, Massachusetts.

Zimmermann, P., Guttormsen, S., Danuser, B., & Gomez. P. (2003). Affective Computing - A Rationale for Measuring Mood with Mouse and Keyboard. *International Journal of Occupational Safety and Ergonomics,* 9(4), 539-551.