Physiological Responses to Events during Training: Use of Skin Conductance to Inform Future Adaptive Learning Systems

Megan Hardy[1], Eric N.Wiebe[2], Joseph F. Grafsgaard[3], Kristy Elizabeth Boyer[3], & James C. Lester[3]
[1]Department of Psychology, North Carolina State University, Raleigh, NC, USA
[2]Department of Science Technology, Engineering, and Mathematics Education, North Carolina State University, Raleigh, NC, USA
[3]Department of Computer Science, North Carolina State University, Raleigh, NC, USA

Understanding the role of physiological responses within the behavioral, cognitive, and affective domains of a training intervention are an important step towards designing augmented adaptive systems that respond to the learner's cognitive and affective states. Multiple studies have shown that specific affective states are related to learning (Craig, Graesser, Sullins, & Gholson, 2004; Graesser & D' Mello, 2011; Kort, Reilly, & Picard, 2001). This paper explores trainees' skin conductance responses to specific behavioral events and theorized cognitive and affective events, and their relationship to learning during a training session within the programming domain. A series of independent samples t-tests revealed that students who exhibited a skin conductance response (SCR) to the behavioral event of compile begins, as well as to affective events of displays of uncertainty, negative feedback, and minimizations of failure had significantly higher learning gain and post-test scores than students who did not exhibit a SCR to these events. These findings provide a step towards understanding the relationship between the physiological measure of skin conductance and affective experiences of the learner in the course of events during a training session, and inform the design of adaptive training and learning systems.

INTRODUCTION

Augmented systems aspire to maximize human cognition through a union of humans and computational systems. Augmented training systems can serve to improve human performance through appropriately adapting and responding to a learner's state. This union of human and machine in the form of adaptive systems have tremendous potential for the training field (Stanney et al., 2009). In order to develop such systems, it is important (1) to know what aspects of a learner's affective state influence performance during a training intervention, (2) have valid and reliable measures of these states, and (3) be able to integrate this knowledge into a computational model that appropriately responds to the learner's state to maximize cognition and improve performance.

Recent research indicates that both cognition and affect play an important role in training and learning interventions (Craig et al., 2004; Graesser & D'Mello, 2011; Kort et al., 2001). This work builds off foundational cognition research on how affective components engagement and motivation are involved with cognitive processes like decision-making and memory (e.g., Ashbury, & Isen, 1999). Stein & Levine (1991) propose a theoretical model that predicts that when there is an emotional episode (or affective response) that learning almost always occurs. The physiological basis of this model lies in the activation of the autonomic nervous system (ANS) in response to a mismatch between existing schemas and incoming information. When this ANS response occurs in parallel to a cognitive appraisal or meta-level analysis of the situation, the learner experiences an affective state that may be beneficial to learning (Stein & Levine, 1991).

Research has shown that affective states of engagement (i.e. flow) and motivation have had significant influences on learning within a training session (Craig et al., 2004; Woolf et al., 2009). An optimal state for learning, the *flow state,* is described as an ideal learning state where learners are engaged and absorbed in material (Csikszentmihalyi, 1990). To maintain or enhance the *flow state*, learners must be presented with tasks that are both achievable and challenging (Csikszentmihalyi, 1990). Appropriate amounts of challenge can allow the learner to confront points of cognitive disequilibrium (contradictions, obstacles, contrasts, or surprises), often resulting in a deep comprehension of the material during learning (Piaget, 1970). Cognitive disequilibrium may occur when the learner experiences confusion. Therefore, states of confusion have been pointed to as an affective state that is useful to model in adaptive learning systems (Craig et al., 2004; Rozin & Cohen, 2003). Appropriately assessing the level of engagement (or other state based measures) during a task can supplement understanding of the learner's cognitive and affective processes.

Typically, measures of a learner's cognitive and affective states (i.e., engagement, interest, motivation) are gathered post hoc. Previous research questions the ability of these measures to accurately portray a learner's state throughout the experience (e.g., Podsakoff, & Organ 1986, Sharek, 2012). Besides being potentially disruptive, a major limitation of such measurement tools is that they were developed to access the psychological state of the learner after the exposure to training. This makes it difficult to detect temporal changes in learner engagement or other cognitive and affective states – a necessary capability for adaptive system development and use. While still a useful tool as part of research and development of adaptive systems, there is a need to expand the data sources used in design and deployment.

The problem presented to adaptive system developers is determining which secondary measures are reliable, valid, and useful measures of detecting a learners changing affective state in real time. Adaptive systems would benefit from a real time, non-disruptive measure of a user's state.

Prior research points to the use of facial, postural and physiological measures to investigate affective phenomena in the context of learning environments (Grafsgaard, Boyer, Wiebe, & Lester, 2012; Grafsgaard, Fulton, Boyer, Wiebe, & Lester, 2012; Mello, Taylor, & Graesser, 2004). Studies of physiological activity during an interaction have linked skin conductance to affective responses during learning (Rebelo, Noriega, Duarte, & Soares, 2012). Skin conductance reflects emotional, or affective, responses (Boucsein, 2012). Additionally, skin conductance responses have also been used as indicators of emotional learning (Lanzetta & Orr, 1986; Ohman & Dimberg, 1978). Skin conductance may provide researchers and developers with additional insight into the cognitive and affective state of the learner. Yet research is just beginning to evaluate skin conductance in training and adaptive learning contexts.

In an adaptive system, real time objective measures of psychological processes that monitor learner state may supplement and complement post hoc measures of engagement, flow or other state based measures. Rebelo et al. (2012) point out that in measuring a user's experience during a task, real time measurements have several advantages. They are a "cleaner" measure of user state when compared to subjective post hoc self-reports. They are less invasive and less likely to interrupt the user (less likely to disrupt flow). Additionally, skin conductance analyses can be triangulated with other data sources (e.g., facial expressions, postural position, specified events) to determine valence of the arousal event.

Skin Conductance

Electodermal Activity (EDA) is a term that describes any electrical phenomena in the skin (Boucsein, 2012). Skin conductance, a specific measure of EDA, can be measured non-invasively by applying a low constant voltage to the skin (Benedek & Kaernbach, 2010). Skin conductance has two major components; tonic and phasic. The tonic component, also referred to as skin conductance level (SCL), is activity that slowly varies over time (Benedek & Kaernbach, 2010). The phasic component consists of abrupt increases in the conductance of the skin forming a peak that is followed by a slower decline back to baseline (Benedek & Kaernbach, 2010). When an abrupt increase in phasic activity passes a threshold (often set between .01 and .05 μS), it is referred to as a significant skin response (SCR). SCRs are arousal responses that occur in response to a stimulus or event.

The sympathetic nervous system (SNS) controls arousal responses within the body. When an arousal response is triggered, the sudomotor neurons of the SNS send a signal to the sweat (eccrine) glands that the sudomotor neurons innervate (Dawson, Schell, & Filion, 2000). This process can be measured as a change in skin conductance and is a linear correlate with arousal (Lang, 1995). Although, only the SNS directly innervate the sweat glands, some adrenergic fibers (within the autonomic nervous system) exists in close proximity to the sudomotor fibers of the SNS (Dawson et al., 2000). This might provide a physiological linkage to Stein & Levine's (1991) theory on how affective responses originated by the autonomic nervous system aid in the learning process.

Physiological measures like skin conductance, may move the theoretical understanding of the role of affect in learning forward, and may prove to be robust enough to be utilized in a real time adaptive system. The focus of this paper is centered around learning-related behavioral events during a training session on introductory Java programming and the skin conductance data collected during these sessions. Our goal is to examine the efficacy of skin conductance measures to inform models of learner states within an adaptive system. We integrate observable behavioral events, events that point to certain cognitive and affective states, as well as the skin conductance responses to these events into a model to predict learning within a training intervention.

In the context of this training session, our hypotheses are:

(1) Student engagement is positively related to learning.
(2) Specific behavioral events initiated by a student or tutor may evoke a physiological response.
(3) Students' display of specific affective states may evoke a physiological response that relate to learning gain.

METHOD

Participants

Students (N = 38) were recruited to take part in a computer mediated training sessions designed to teach introductory Java programming concepts. Students were not allowed to participate if they reported prior experience that would indicate that they knew the material covered in the curriculum.

Materials

Students interacted with human tutors via a web-based interface that provided a coding interface, a chat pane as well as task information to be viewed on one screen. Events occurring within the tutoring session were logged and stored in a database with a timestamp for syncing the events. Five trained coders tagged student dialogue for specific occurrences. Logged behavioral events and tagged dialogue events with their descriptions can be found in Table 1. Skin conductance was measured through a skin conductance bracelet (see Figure 1). The sensor collected data at 32 Hz for each student per lesson.

Procedure

Each student was randomly assigned one tutor for a series of six lessons across the semester. Lessons consisted of programming tasks that mapped to fundamental computational concepts. Tutors guided the task progression, and by design, previous tasks could not be revisited. Each tutoring session covered one lesson and was limited to a maximum of forty minutes.

Table 1.

Logged Behavioral and Tagged Dialogue Events

Behavioral Events	
Compile Begin	Logged when student clicked the compile button
Compile Error	Logged when student received compile error
Compile Success	Logged when student received compile success message (no errors were in code)
Code	Logged when students were editing their program within code window
Dialogue Events	
Uncertainty	Instances of student displays of uncertainty or confusion
Minimization of Failure	Tutor reassurance after a mistake, setback, or other poor performance
Negative Feedback	The tutor gives negative assessment of performance
Positive Emotion	Instances of student displays happiness, joy, excitement
Positive Feedback	The tutor gives positive assessment of performance
Maximization of Success	Tutor praise for good performance (beyond positive feedback)

Prior to beginning a session, students completed a lesson content-based pre-test. After the pre-test, tutors and students interacted through the web-based interface that provided the task content, basic programming functionality with the capability to compile and run programs interactively, and a chat pane. After each session, students completed a post-test consisting of the same lesson based content questions as in the pre-test. Students also answered a post survey composed of a modified User Engagement Survey (UES), (O'Brien and Toms, 2010), and the NASA- TLX, (Hart and Staveland, 1988). The modified UES included the Focused Attention, Endurability, and Involvement subscales. Perceived Usability, Aesthetics, and Novelty subscales of the UES were omitted as they primarily related to experience with the interface rather than the learning task. The NASA-TLX measuring cognitive load, entailed response items for Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level.

Cognitive Affective Task Analysis

In order to aid the selection of events occurring within the training session that correspond to behavioral events, theorized cognitive events, and affective events, a modified cognitive task analysis was performed. Cognitive task analysis is a technique used to document thought processes underlying behavioral events (see Schraagen, Chipman, & Shalin, 2000). A trained dialogue coder combined knowledge of existing transcripts and a top down approach to implement an adapted method of cognitive task analysis to include behavioral events, theorized cognitive events and theorized affective events of students during interactions of a tutoring task. A training task was divided into five subtasks, summarized below.

Subtask 1. Processing Phase. Cognitively, students process the requirements of the task and evaluate the current or existing state of their knowledge in comparison to what is required to complete the task. Students reach a decision point on whether

Figure 1. Study Setup. Skin conductance data measured through a Skin Conductance Bracelet.

or not they possess the knowledge and skills to complete the objective of the task and proceed with planning.

Subtask 2. Planning Phase. If students do not possess the knowledge required to complete the task, the planning phase might include segments of questions and answers between the student and the tutor, or statements made by the tutor with acknowledgements from the student. Affectively, students might display uncertainty during this phase. Dialogue with the tutor may be to gain specific information to implement the task or it may be meta-level information as to how to pose questions that allow them to go about acquiring the information they need. When students understand the plan, they may make a statement to the tutor, affectively indicating a positive emotion or excitement, or they might skip straight to the implementation phase.

Subtask 3. Implementation Phase. The implementation phase begins with typing code within the task window. Here an iterative process of code creation (typing) and code evaluation takes place. This process could be supplemented with feedback provided by the tutor and/or by segments of questions and answers between the student and tutor. Affectively, students could have some affective response to the type of feedback given by the tutor, or display uncertainty within the question and answer events. This iterative process continues until the student feels the code is ready to test.

Subtask 4. Testing Phase. Testing the code begins with a compile begin event. Depending on the level of involvement with the task, students may internally exhibit an affective display of anticipation or excitement at the compile begin event. In instances of compile errors students may have an affective response. When there is a compile success, students likely receive feedback from the tutor aimed to affectively support their accomplishment.

Subtask 5. Evaluation & Reflection Phase. Students are solidifying the programming elements and concepts used within the task. This cognitive process is supported through a combination of statements made by the student, tutor-led tests of knowledge, and/or tutor-led demonstrations.

Within each of these subtasks, the events of display of uncertainty, negative feedback, and minimizations of failure may be indicators that cognitive disequilibrium has occurred. Students would only display uncertainty if they were unsure if what they had coded or a statement they had made was correct. The tutor would only give negative feedback or minimize failure when a student had made a coding error or stated something incorrectly. Each of these events suggests a gap between existing knowledge of the student and correct knowledge required to complete the task. These events, as well as behavioral events occurring during the testing phase are points of interest in investigating skin conductance phenomena within the training session.

Data Analysis

Skin Conductance Analysis. For the purposes of this paper, data from a single lesson (Lesson 1) were analyzed. Log events as well as dialogue events were time synced with each students EDA data. Within this data, events occurred in close proximity to each other. Skin conductance responses to events that overlap can be difficult to analyze because the SCRs can be superposed on one another. To mitigate this problem, we chose a method of analysis recommended by Benedek & Kaernbach (2010). This method, Continuous Decomposition Analysis (CDA), decomposes skin conductance data into its tonic and phasic components. This decomposition allows for the extraction of overlapping SCRs. CDA was performed using Matlab software, Ledalab. The software allows for event related data analysis using SCRs. The threshold for detecting a SCR was set to a minimum change in amplitude of .02µS occurring within 1 – 5 seconds of the onset of the specified event.

RESULTS

To address hypothesis 1, Pearson product-moment correlations were run to assess the relationship between engagement subscales and normalized learning gains. Preliminary analyses showed the relationships to be linear, normally distributed as assessed by Shapiro-Wilk ($p > .05$), and there were no outliers. There were significant positive correlations between each engagement subscale, as well as the average of the three subscales, to learning gains (see Table 3). As engagement ratings increased on the focused attention, felt involvement and endurability subscales, overall learning gains increased.

Table 3.

Correlations between Engagement and Learning Gains

Engagement Variables	*R*	R^2
Focused Attention	.323*	.104
Felt Involvement	.349*	.122
Endurabiliy	.383*	.147
Overall Engagement	.415**	.172

Note: *p<.05, ** p<.01

To investigate hypotheses 2 and 3, a series of independent-samples t-tests were conducted. First, analyses looked at SCR's for the behavioral events of *compile begin*, *compile error* and *compile success*.

For compile begin, an independent-samples *t*-test determined that there were differences in learning gains between students who exhibited a SCR to *compile begin* events and students who did not exhibit a SCR to *compile begin* events. Learning gains for each group were normally distributed, (Shapiro-Wilks, $p > .05$), and homogeneity of variances was not violated (Levene's test $p > .05$). Results showed that learning gains were significantly greater for students who exhibited a SCR to *compile begin* events ($M = .59$, $SD = .34$) than students who did not exhibit a SCR ($M =.32$, $SD = .30$), $t(36) = 2.58$, $p = .014$. Students who displayed a SCR to *compile begin* events answered, on average, 59% of the questions missed on their pre-test correctly on their post-test, whereas students who did not display a SCR only answered 32% of the questions missed on their pre-test correctly on their post-test. This finding is also mirrored in post-test scores. Post-test scores were significantly greater for students who exhibited a SCR to *compile begin* events ($M = .80$, $SD = .15$) than those who did not ($M = .68$, $SD = .16$), $t(36) = 2.82$, $p = .029$. Students displaying a SCR scored, on average, 80%, while students who did not display a SCR scored, on average, 68%.

No significant differences in learning gains were found between students who exhibited a SCR to *compile success* events and those who did not exhibit a SCR. Similarly, no significant differences in learning gains were found between students who did or did not exhibit a SCR to *compile error* events. However, an independent samples t-test revealed that students who displayed a SCR ($M = 21.75$ $SD = 19.95$) to compile error events reported higher levels of frustration on the NASA-TLX than students who did not display a SCR to compile error events ($M = 9.15$ $SD = 14.92$), $t(36)=$, $p = .037$). Although frustration was not normally distributed, (Shapiro-Wilks $p > .05$), there were no outliers, and the assumption of homogeneity of variance was not violated.

Independent samples t-tests were run to examine SCRs after affective events pointing to the existence of a cognitive disequilibrium (*uncertainty*, *negative feedback*, and *minimization of failure*). Results showed that students who exhibited a SCR ($M = .78$, $SD = .15$) after any of these affective events had significantly higher learning gains ($t(25) = 4.34$, $p < .001$, equal variances not assumed) than students who did not exhibit a SCR ($M = .35$, $SD = .36$). Additionally, students who exhibited a SCR ($M = . .88$, $SD = .09$) had significantly higher post-test scores ($t(36) = 2.58$, $p = .014$), than students who did not exhibit a SCR ($M = .70$, $SD = .17$) to these affective events.

DISCUSSION

The purpose of our analyses were to investigate the role of both post hoc measures of engagement and real time measures of skin conductance within a training session. This was done with the goal of better understanding how these measures might inform the research and development of adaptive systems. We found that the post hoc measure of engagement was positively correlated with learning gains. This finding is in line with theory and prior empirical findings with similar training systems. A modified version of cognitive task analysis successfully identified both behavioral events and affective events which preceded physiological responses (i.e., SCR's) related to learning. Results from event analyses show that students who respond to the *compile begin* event have higher learning gains and score higher on the post-test. The amount of arousal could point to the level of investment that students have in the training session. That is, occurrence of SCR's may point to an emotional commitment to the result of the code compilation that emerges from an increased level of engagement towards learning. For affective events analyses, we found that students that displayed a physiological response to a display of uncertainty, negative feedback, or minimizations of failure given by the tutor, have higher learning gains and score higher on the post-test than those who did not exhibit a response. Again, emotional response to uncertainty or mistakes may be linked to high engagement towards learning. Similar patterns in SCR's did not occur with the compilation results (*compile success* or *compile error*). However, appropriately, a self-report measure of frustration did seem to be related to high SCR's occurring with compile errors.

CONCLUSION

Physiological responses to behavioral, cognitive and affective states are important to investigate in order develop effective implementations of augmented training systems. Cognitive task analysis proved to be a useful tool in identifying events that might be of interest in a training intervention. Examining events identified by this analysis during a lesson within a tutor mediated training intervention, resulted in finding significant differences in learning between students who exhibited skin responses and those who didn't, supporting prior theory building around the role of the autonomic nervous system in the learning process. These event types can be readily identified in similar training systems and used in conjunction with real-time physiological data to provide adaptive support for trainees. Although there were limitations, our findings supports the research literature on learner affective states in the context of training interventions, as well as provide a step forward in assessing the usefulness of skin conductance measures in future design of adaptive systems.

A limitation of our analyses was the small data set. The next phase of work will focus on the compilation of more data points to allow for regression analyses and in-depth consideration of other variables computed by CDA, such as response latency and amplitude of response. Future analyses could involve a time series analysis of events to inform a bottom up version of a cognitive task analyses

REFERENCES

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*(3), 529-550. doi: 10.1037/0033-295X.106.3.529

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80-91. doi: 10.1016/j.jneumeth.2010.04.028

Boucsein, W. (2012). *Electrodermal activity*. New York: Springer.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of psychophysiology*. New York: Cambridge University Press.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*(3), 241-250. doi: 10.1080/1358165042000283101

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The Electrodermal system. In Cacioppo, J.T., Tassinary, L.G., & Berntson, G. (Eds.). *Handbook of psychophysiology*, (pp. 200-223). Cambridge University Press.

D'Mello, S., Taylor, R. S., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In *Proceedings of the 29th annual meeting of the cognitive science society*, 203-208.

Graesser, A., & D'Mello, S. (2011). Theoretical Perspectives on Affect and Deep Learning. In R. A. Calvo & S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (255–271). doi:10.1007/978-1-4419-9625-1

Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012). Analyzing Posture and Affect in Task-Oriented Tutoring. Artificial Intelligence. *Proceedings of the 25th Florida Artificial Intelligence Research Society Conference*, 438-443.

Grafsgaard, J. F., Fulton, R. M., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012). Multimodal analysis of the implicit affective channel in computer-mediated textual communication. *Proceedings of the 14th ACM international conference on Multimodal interaction*, 145-152.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human and Mental Workload*. *1*(3), 139-183.

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. *Proceedings IEEE International Conference on Advanced Learning Technologies,* 43–46. doi: 10.1109/ICALT.2001.943850

Lanzetta, J. T., & Orr, S. P. (1986). Excitatory strength of expressive faces: Effects of happy and fear expressions and context on the extinction of a conditioned fear response. *Journal of Personality and Social Psychology, 50*(1), 190-194. doi: 10.1037/0022-3514.50.1.190

Ohman, A., & Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses: A case of "preparedness"? *Journal of Personality and Social Psychology, 36*(11), 1251-1258.

Piaget, J. (1970). *Genetic epistemology*. New York: Columbia University Press.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544. doi: 10.1177/014920638601200408

Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using virtual reality to assess user experience. *Human Factors: The Journal of Human Factors and Ergonomics Society, 54*(6), 964-982. doi:10.1177/0018720812465006

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of americans. *Emotion, 3*(1), 68-75. doi: 10.1037/1528-3542.3.1.68

Schraagen, J. M., Chipman, S. F., & Shalin, V. (2000). *Cognitive task analysis*. Mahwah, N.J: L. Erlbaum Associates.

Sharek, D. J. "Investigating Real-time Predictors of Engagement: Implications For Adaptive Video Games and Online Training." Unpublished dissertation. North Carolina State University, 2012.

Stanney, K. M., Schmorrow, D. D., Johnston, M., Fuchs, S., Jones, D., Hale, K. S., Ahmad, A., & Young, P. (2009). Chapter 7 augmented cognition: An overview. *Reviews of Human Factors and Ergonomics, 5*(1), 195-195. doi: 10.1518/155723409X448062

Stein, N. L., & Levine, L. J. (1990). Making sense out of emotion: The representation and use of goal-structured knowledge. *Psychological and Biological Approaches to Emotion*, 45-73.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3), 129–164.