Understanding Privacy and Trust Issues in a Classroom Affective Computing System Deployment

Shaundra Bryant Daily[1], Dante Meyers[2], Shelby Darnell[1], Tania Roy[1], and Melva T. James[1]

[1] School of Computing, Human-Centered Computing Division, Clemson University, Clemson, South Carolina, USA
[2] Department of Psychology, Old Dominion University, Norfolk, Virginia, USA
{sdaily,sdarnell,taniar,melvaj}@clemson.edu
dmyer010@odu.edu

Abstract. Our research group is in the midst of working with teachers to co-design an affective computing system that uses physiological measures, gathered via wrist worn sensors, to understand how students are engaging with classroom instruction. Optimally, our goal is to find new ways of supporting empathetic practices in the classroom by providing teachers real-time (or reflective) feedback on student engagement. In parallel, with our work with teachers, we are working to pinpoint the privacy and trust issues that might be associated with this type of system. The objective of this paper is to present the results of a series of studies conducted to understand the challenges associated with introducing a pervasive affective computing system into classroom environments. While we focus on physiological sensors, the implications apply to other pervasive technologies as well.

Keywords: Affective Computing, Privacy, Adoption.

Introduction

Affective computing is "computing that relates to, arises from, or deliberately influences emotion." [1] As with any pervasive computing application, issues of privacy and trust for affective computing systems must be a part of the conversation from design to implementation. Reynolds & Picard [2] provide a framework for the evaluation of affective computing systems from a dimensional metaethical position and include privacy and trust, amongst other constructs, as a part of this analysis. Further, they evaluate systems in a small study based on this framework. Lane, et al. [3] raise important considerations as physiological sensing becomes more pervasive, including how to process data for best privacy, how to share data appropriately, unintended leakage of personal information, and who is responsible when collected data causes harm. Similarly, Ameen, et al. [4] caution that serious social unrest might arise if people fear being monitored. They suggest encrypting communications, keeping users anonymous unless completely necessary, and creating public awareness. These guidelines are a part of a larger literature about not only the importance of keeping these

N. Streitz and C. Stephanidis (Eds.): DAPI/HCII 2013, LNCS 8028, pp. 414–423, 2013.

data safe, but also ensuring that the system is able to overcome cultural, psychological, and other social barriers to adaptation and utilization. Although the focus of this paper is not the development of a pervasive affective computing system, we begin by describing our recent efforts here in order to provide the reader with the context.

It is imperative to support a teacher's ability to understand how he/she is connecting with his/her students in the classroom so he/she can adapt his/her pedagogical strategies to meet the needs of his/her diverse learners. Engagement, defined in the literature in a variety of ways, has been associated with student achievement, [5] positive classroom and school climate, [6] and effective instructional practices. [7] Engagement is a multidimensional construct with cognitive, behavioral, and affective dimensions. First, cognitive engagement is related to a student's investment in learning, seeking challenges, going beyond requirements, and self-regulation. [8] Cognitive engagement can be demonstrated by a student's mastery of the full meaning of material, taking the position of an expert rather than a novice. [9] Next, behavioral engagement is related to participation and involvement in activities. This includes observable behaviors such as positive conduct, persistence, effort, and attention. Lastly, affective engagement measures positive and negative reactions to stimuli including teachers, classmates, academics, or school. Positive emotional engagement supports student ties to institutions and is presumed to influence their willingness to work.

To date, much evaluation of each of these dimensions of engagement relies on self-report and other subjective, obtrusive, and inconsistent instruments. [10] The goal of our line of research is to contribute to the creation of a physiology-based, quantifiable, and unobtrusive technique for measuring the affective response associated with the dimensions of engagement with the goal of supporting teachers and students. The key measure in the affective computing system discussed in this paper relies upon wrist-worn sensors that measure electrodermal activity. In short, when sympathetic nervous system activity increases, sympathetic fibers that surround eccrine sweat glands modulate the production of sweat. The skin, in turn, momentarily becomes a better conductor of electricity (i.e., electrodermal activity). This electrodermal activity can be measured as conductance or resistance by different sensors. Here, we focus on skin conductance, for which sensors. [11] These sensors are placed on the fingers, the palm of the hand, or the wrists where there is a large concentration of sweat glands. [12] We have chosen to use the Q sensor to collect skin conductance, temperature, and motion data since this sensor can be worn outside of a laboratory setting (i.e., without being tethered to a computer) and since it is worn on the wrist like a watch, which might increase the possibility of it being unobtrusive to the student.

This sensor provides information about a person's level of arousal provided that other triggers of increased perspiration have been held constant (e.g., temperature). It will not provide any information as to the specific emotion that is being elicited unless other conscious emotion variables are collected. Further, numerous events such as pain, significant thoughts (not related to the current context), lying, exercise, individual changes in biochemistry, and motion artifacts can lead to changes in skin activity. Even with attribution and noise limitations, however, electrodermal activity is a useful measure that has been used in research focusing on stress and anxiety, [13] lie detection, [14] user interface evaluation, empathy, [15, 16] and game assessment. The final

two are most relevant to the system under development in the classroom setting. By measuring skin conductance simultaneously from patients and therapists during a clinical session, Marci et al. [16] found that increased therapist empathy as perceived by the patient correlated with high concordance of skin conductance between the two. In other words, the more empathic the patient felt his/her therapist to be, the stronger the relationship between skin conductance measures. Next, Mandryk et al. [17] found that skin conductance was higher when playing a game against a friend rather than a computer and was correlated with subjective measures of "fun". Further, in a separate study, Mandryk et al. [18] found that a combination of physiological measures, which included skin conductance, were useful in evaluating the emotional response to entertainment technologies.

In order to support teachers in understanding how they are connecting with their students, our approach to this research incorporates two goals: First, we are co-developing, with teachers, principals, and district leaders, a user interface tool that allows a teacher to peruse this engagement data connected with video of his/her classroom activity. Second, we are conducting feasibility studies to understand the social, political, cultural, and psychological barriers to this pervasive affective computing system. The latter is the focus of the rest of this paper.

Methods

2.1 Study Context

In the midst of our participatory research with teachers, a blog post[1] was released about one author's opinion of the research. Unfortunately, the post was based on a mistake on a website connecting our efforts to empower teachers and students to understand engagement with another study trying to understand teacher effectiveness. The interpretation of this work in the blog post was, in effect, that the arousal levels of students would be utilized as a way to evaluate, and possibly fire, teachers. Even though the post, which generated hundreds of responses, was based on inaccurate information, the resulting perceptions of the project are valid and informative of the views people might develop about a pervasive affective computing system. Therefore, the post provided an opportunity, beyond the efforts in the participatory research with teachers, to examine the proposed affective computing system critically.

2.2 Data Sources

We have three main data sources: focus groups, online news articles, and social media sites. Our twenty-four participants in the focus groups have been divided into two cohorts: teachers and district administrators. The first cohort consists of ten sixth- through eighth-grade teachers and a principal from a middle school the Southeast who

[1] Our effort here is not to critique or refute this blog post. In an effort to focus on the privacy and trust topic of the paper and to avoid igniting further discussions of the blog post, we have purposefully chosen not name the author and source of the post.

teach English, science, social studies, and mathematics. All participants are White females with one to fourteen years of experience teaching who responded to a general request from their principal asking if they would be interested in participating in the study. They had no knowledge about the details of the study, so there should not have been a bias towards teachers who were more accepting of affective computing. The second group consists of various administrators from the school district including two principals, a professional development specialist, a responsiveness to instruction specialist, a high school instructor, a professor, and a project executive director.

Our focus groups were divided into three sessions held with each cohort. Consent forms were given to each group and participants were given the opportunity to opt out of participating. During the first session, we first introduced stakeholders to electrodermal activity as a measure, as well as the sensors for measurement. Next, we provided background for the goals of the project and familiarized ourselves with the ways in which they try and understand how their students are engaged in the classroom. We also asked questions related to the gaps they see inherent in their approach and how they could imagine improvement. In addition to field notes collected by a research assistant, we collected drawings from stakeholders to understand how an affective computing system might look. During the second session, we presented a paper-based prototype developed as a result of the previous discussion and asked for feedback and improvements. In an upcoming session, we will bring a tangible prototype to the stakeholders. During three different focus group sessions held thus far, two with the teacher cohort and one with the administrator cohort, field notes were generated from discussions.

The remainder of our sources consist of online news articles and social media sites (n = 522) containing articles written in response to a blog post reporting inaccurate information about the work. These sources were found using the Social Media Listening Center at Clemson, which uses Radian6 technology to filter relevant articles and posts dealing with the topic. Radian6 is a social media monitoring platform that gathers data, in part, from Facebook, Twitter, blogs, blog comments, message boards and online forums, news groups, podcasts, reviews on e-commerce sites, experience sharing sites, and mainstream news sites. In order to accomplish this task, a filter was created for terms in the initial blog post for a one-month time period that included the initial blog post.

2.3 Analysis

An initial analysis of the focus group and social media data consisted of computing percentages for positive, negative, and neutral sentiments. Data were coded by two raters as positive if the poster expressed clear agreement with the technology, discussed beneficial implications, or provided positive suggestions; coded as negative if the poster used profanity, totalitarian references, expressed anger, disgust, fear, or resentment; and neutral if remarks were indefinite (i.e., forwarded online article or no valenced opinion presented). Finally, an interrater reliability analysis was performed to determine agreement between raters.

Next, thematic analysis was utilized to generate themes from the focus group notes and online data. Thematic analysis is "an accessible and theoretically-flexible approach to analyzing qualitative data". [19] This method, widely used in psychology, calls for the demarcation of a qualitative data corpus into themes. Thematic analysis is valuable when attempting to understand a data corpus whose information is based on notes from study groups and blog posts with comments, in other words, data that were not collected under experimental conditions. All data were also analyzed using thematic analysis procedures which include building familiarity, generating codes, identifying features, finding, confirming, and defining themes for reporting. [19] The focus group and social media data presented separate themes after analysis and are explained in detail in the next sections.

Results

3.1 General Sentiments

Table 1 below depicts the positive, negative, and neutral sentiments found in the online news articles and other social media. The interrater reliability for the coders was found to be 78.7% with Cohen's Kappa = 0.55, indicating moderate agreement. [20]

Table 1. Results from analysis of positive, negative, and neutral sentiments

Groups	% Negative	% Positive	% Neutral
Focus Groups (n = 24)	2%	98%	0%
Articles & Social Media Coder A (n = 522)	31.4%	6.0%	62.6%
Articles & Social Media Coder B (n = 522)	25.5%	6.3%	68.1%

3.2 Social Media Themes

Theme One. Many of the authors and commenters believe that this project is just another effort to control aspects of citizen's private lives. The name "Big Brother" (a totalitarian dictator from George Orwell's novel entitled *Nineteen Eighty-Four*) [21] was commonly used. One author posted, "it is not too far of a stretch to assume that tracking bracelets could one day be used to weed out students or teachers that do not buy into…agenda[s]… By monitoring what is being taught and how students respond to it, Big Brother could theoretically read the human mind in real time, which has some fairly disturbing implications."

Theme Two. The second theme was that the technology was being used to *evaluate teacher performance*. Authors thought the technology would be used to evaluate teachers rather than help and empower the classrooms. One author posted, "Using

students' emotional responses to various learning material as a metric of how well a teacher is performing is a flawed approach that could send many quality, veteran educators packing their bags." Another author posted, "A student's physical reaction to a classroom lesson soon could be used to judge how successful—or unsuccessful—an educator is in keeping students engaged." Lastly, an author suggested, "The student reactions recorded on the bracelets' sensors could be added to a host of more traditional teacher evaluation methods such as test grades, administrator observations, and student surveys."

Theme Three. Many believed that the technology would not be able to distinguish what is actually engaging a student. In other words, the technology would not work. One poster wrote, "In any case, even if a child is giving off highly engaged skin signals, how would the machines know whether he or she is deeply engaged in a beautiful daydream rather than 14th-century English literature?" Another author asked a similar question, "How would the bracelet tell if a student is responding to a teacher and not to something his friend whispers in his ear?" Authors and posters also cautioned at the fact that people can be deceptive. One poster maintained, "It's a fair point, but in terms of the GSR's actual effectiveness, there's one thing researchers should bear in mind: Children are very, very good at cheating."

3.3 Focus Group Themes

Theme One. The first of the focus group themes was *informative feedback*. Generally, the teachers and administrators expressed desires for the engagement pedometer to provide information that would help them adjust their lesson plans. One teacher insisted the technology should "Alert teachers of low levels of engagement, so teachers can monitor or re-engage students." Another said, "For lessons [the technology should tell] if the teachers consistently have low engagement or high engagement at specific points in the lessons." Some teachers wanted the information from the tool as a reflective feature while others wanted real-time information, or as one teacher stated, "Instruction intervention as class proceeds."

Theme Two. The second theme was the teachers having the option of using the pedometer to *view individuals or groups* within the class. One teacher asked to be able to "Target particular students upon request." Some teachers insisted that since class is not always individual work, this would be helpful. One teacher said, "Can each student's dots [points representing students] also have a number so we can track, if we move from individual to pair to group work?" Some teachers also expressed the desire to observe patterns and behaviors of students over time. One teacher suggested "A way to follow a particular student through lessons, class schedules, and four-week periods as a way of tracking progress/engagement."

Theme Three. The third theme was the desire to have *access to data immediately* for interpretation. Some teachers expressed that they wanted the data to come straight to different devices they have such as iPads or other mobile devices. One teacher maintained, "I would want data immediately to my PC or other device so I could access it to be able to adjust instruction and monitor students/groups." In relation to Theme

One, rather than having an alert for feedback, teachers want to interpret the data themselves. One teacher explained that it would be useful if "Each student [was] able to be monitored and the teacher [was] able to see live data to use immediately to be able to intervene and instruct in a different way to benefit all students."

Discussion

In this research, there is an obvious discrepancy between sentiments expressed between the two data sets. Most focus group participants found the idea amenable to use in classes and provided suggestions on the design of the tool. All the themes from the focus groups were about aspects they wanted in a tool to augment their own abilities to support students in their classrooms. Only once or twice were ideas expressed around who would be in control of the data and what the possible negative uses of this information were. This general positive sentiment, of course, can be attributed to how the information was presented to these groups. In the focus group sessions, we presented the technology, described its limitations (e.g., noise and attribution errors), discussed opportunities to opt into the research, and asked for teacher input. Although we asked for both feedback and concerns (i.e., should this be developed at all), this approach did not espouse the same number of negative reactions as the blog post.

The results confirm previous suggestions presented in the Introduction about privacy and trust of pervasive computing systems. First, public awareness of accurate information is crucial. Developers of pervasive affective computing systems must make sure that people understand the ins, outs, and limitations of what is being developed. The challenges associated with the interpretation of electrodermal activity are an important limitation that must be discussed when presenting the current system. Second, who has access to data and how they will be utilized is important. In this case, steps must be taken to prevent the technology from being used to evaluate the teacher. Rather, this system can be used by the teacher to help him/her understand better how students are responding to his/her pedagogical approach. The student, of course, cannot be forgotten in this picture. Although the technology has been framed for empowering the teacher, the system should not be used to evaluate the student either. Instead, it can be used to facilitate teachers' understanding of their impacts on students so they can be better supported. In addition, it can be used by the students to understand their own engagement. Finally, Ackerman [22] defines privacy as "the ability of an individual to control the terms under which their (sic) personal information is acquired and used". In order for systems to be respectful of teachers' and students' choices, opportunities must be provided to opt in, or out, at any moment. For our technology, teachers and students should have the choice about providing their engagement data to the system.

4.1 Limitations

The main limitations are the scope of analysis and biases. The scope of analysis limitation deals with the number of responses from the focus group compared to those

from social media. Given the uniqueness of our research there exists no readily available information, in similitude to the notes taken during the focus group, other than what has been collected, from which to draw themes. The number of data points for the focus group is twenty-four, while the number of data points for social media is five hundred twenty-two. Another limitation is how each group was introduced to the topic. The focus group was introduced to the topic by researchers whose intentions are to build a tool to enable teachers to improve classroom experiences, while the social media group was introduced to the topic by a blog post that did not accurately represent the work. These disparate introductions impacted how the discussions proceeded. Finally, biases exist for interpretation of the sentiments. After performing the coding for the social media group, the two coders had moderate to substantial, instead of strong agreement. Even with definitions in place for the coding scheme, what we as researchers see as positivity and negativity can still differ.

Conclusions

From this research, we have public perception data from a real situation confirming that those interested in introducing pervasive technologies must fully describe a system's goals and possible limitations, they must emphasize respectful (e.g., opt in) opportunities to use the technology, and they have to provide intended users with choices about how, when, and where their data can and cannot be utilized. Kotter and Cohen [23] posit that "[p]eople change what they do less because they are given analysis that shifts their thinking than because they are shown a truth that influences their feelings." By providing a visualization that is informative, not evaluative, for teachers to understand how their classroom instruction is impacting their students (possibly, in unforeseen ways), we imagine that this type of influence is possible. Our continued understanding of privacy and trust issues generated by the research presented in this paper lays the groundwork for the realization of this vision.

References

1. Picard, R.W.: Affective Computing, 1st edn. The MIT Press, Boston (2000)
2. Reynolds, C., Picard, R.W.: Affective Sensors, Privacy and Ethical Contracts. In: Proceedings of the Conference on Human Factors in Computing Systems, Vienna, Austria, April 24-29, pp. 1103–1106. ACM, New York (2004)
3. Lane, N.D., Miluzzo, E., Lu, H.L.H., Peebles, D., Choudhury, T., Campbell, A.T.: A Survey of Mobile Phone Sensing. IEEE Communications Magazine (2010), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5560598 (retrieved)
4. Ameen, M.A., Liu, J., Kwak, K.: Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications. Journal of Medical Systems 36(1), 93–101 (2012), http://link.springer.com/article/10.1007%2Fs10916-010-9449-4?LI=true# (retrieved)

5. Appleton, J.J., Christenson, S.L., Furlong, M.J.: Student Engagement with School: Critical Conceptual and Methodological Issues of the Construct. Psychology in the Schools 45(5), 369–386 (2008)
6. Appleton, J.J., Christenson, S.L., Kim, D., Reschly, A.L.: Measuring Cognitive and Psychological Engagement: Validation of the Student Engagement Instrument. Journal of School Psychology 44(5), 427–445 (2006), doi:10.1016/j.jsp.2006.04.002.
7. Mant, J., Wilson, H., Coates, D.: The Effect of Increasing Conceptual Challenge in Primary Science Lessons on Pupils' Achievement and Engagement. International Journal of Science Education 29(14), 1707–1719 (2007), doi:10.1080/09500690701537973.
8. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research 74(1), 59–109 (2004), doi:10.3102/00346543074001059.
9. Yazzie-Mintz, E., McCormick, K.: Finding the Humanity in the Data: Understanding, Measuring, and Strengthening Student Engagement. In: Christenson, S.L., Reschly, A.L., Wylie, C. (eds.) Handbook of Research on Student Engagement, pp. 743–761. Springer (2012), http://www.springerlink.com/content/v53410867366517m/abstract/ (retrieved)
10. Fredricks, J.A., McColskey, W., Meli, J., Mordica, J., Montrosse, B., Mooney, K.: Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments (Issues & Answers Report, REL 2011–No. 098). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast, Washington, DC (2011)
11. Picard, R.W., Daily, S.B.: Evaluating Affective Interactions: Alternatives to Asking What Users Feel. In: CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches, Portland, OR (April 2005)
12. Boucsein, W.: Electrodermal Activity. Plenum Press, New York (1992)
13. Fenz, W., Epstein, S.: Gradients of Physiological Arousal in Parachutists as a Function of an Approaching Jump. Psychosom. Med. 29(1), 33–51 (1967)
14. Podlesny, J.A., Raskin, D.C.: Physiological measures and the detection of deception. Psychological Bulletin 84(4), 782–799 (1977), doi:10.1037/0033-2909.84.4.782.
15. Marci, C., Orr, S.: The Effect of Emotional Distance on Psychophysiologic Concordance and Perceived Empathy Between Patient and Interviewer. Applied Psychophysiology and Biofeedback 31(2), 115–128 (2006), doi:10.1007/s10484-006-9008-4
16. Marci, C.D., Ham, J., Moran, E., Orr, S.P.: Physiologic Correlates of Perceived Therapist Empathy and Social-Emotional Process During Psychotherapy. Journal of Nervous and Mental Disease 195(2), 103–111 (2007)
17. Mandryk, R.L., Inkpen, K.M., Calvert, T.W.: Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. Behaviour & Information Technology 25(2), 141–158 (2006), doi:10.1080/01449290500331156.
18. Mandryk, R.L., Atkins, M.S., Inkpen, K.M.: A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 1027–1036. ACM, New York (2006b), doi:10.1145/1124772.1124926
19. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. Qualitative Research in Psychology 3(2), 77–101 (2006), doi:10.1191/1478088706qp063oa.

20. Rietveld, T., van Hout, R.: Statistical Techniques for the Study of Language and Language Behaviour. Mouton de Gruyter, Berlin (1993)
21. Orwell, G.: Penguin, London (1977,1984)
22. Culnan, M.J.: Protecting privacy online: is self-regulation working? J. Public Policy Market 19(1), 20–26 (2000)
23. Kotter, J.P., Cohen, D.S.: The Heart of Change: Real-Life Stories of How People Change Their Organizations. Harvard Business Press, Cambridge (2002)

.