Toward Real-time Brain Sensing for Learning Assessment: Building a Rich Dataset

Shelby Keating
Drexel University
Philadelphia, PA 19104, USA
sak347@drexel.edu

Erin Walker
Arizona State University
Tempe, AZ 85287, USA
eawalke1@asu.edu

Anil Motupali
Arizona State University
Tempe, AZ 85287, USA
amotupal@asu.edu

Erin Solovey
Drexel University
Philadelphia, PA 19104, USA
ets36@drexel.edu

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA ACM 978-1-4503-4082-3/16/05. http://dx.doi.org/10.1145/2851581.2892496

Abstract
By integrating real-time brain input into personalized learning environments, it would be possible to capture a learner's changing cognitive state and adapt the learning experience appropriately. Working toward this goal, we aim to develop a robust system that can classify a user's cognitive state during a learning activity, using brain data collected with functional near-infrared spectroscopy, an emerging non-invasive neuroimaging tool. This paper describes preliminary steps we have taken toward this objective as well as the underlying vision and research goals. This work has implications for online education as well as the growing fields of brain-computer interfaces and physiological computing.

Author Keywords
educational data mining; functional near-infrared spectroscopy; machine learning; brain-computer interfaces.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

Introduction
With the growth of online and computer-based learning environments, there now exists an unprecedented amount of data on how students solve problems and

build knowledge. Many educational technologies collect fine-grained clickstream data on student activities within a learning environment, including the steps students take to solve problems, correctness of their actions, hints requested, and additional scaffolding received. Large data mining efforts using this data have provided substantial knowledge on how students learn, including optimal practice schedules for learning materials [2], skills students require to solve a particular problem [14], and the range of off-task behaviors that students may engage in [5]. However, processes related to deep learning, such as reflecting on errors and confronting misconceptions, often occur at times when students are thinking and thus, during a *pause* in the log data. Of course, pauses during use of educational systems can also be indicative of many other cognitive or motivational states such as conversing with a teacher, engaging in off-task behavior, daydreaming, or simply focusing on the wrong problem features [7,18,28]. The predictive power of pauses varies across datasets based on when they occur, what happens after, and its length. If pauses during learning could be better understood and characterized, it might be possible to build predictive models of when students are learning from online environments.

It may be possible to leverage brain data and other physiological sensing to disambiguate what is occurring during pauses in log data. There are indications that fMRI can be used to detect when students are thinking deeply about a problem [1], and EEG can be used to predict learners' scores on an assessment [16], their correctness and confidence in their answers [22], and aspects of their mental state (e.g., workload, engagement, and distraction) [40]. In addition, physiological sensors have been explored for forming better predictions of learning from educational data. Eyetracking has been used along with log data to differentiate between where successful and unsuccessful students attend to on the interface [35,32], predict self-explanation [17], and indicate boredom and curiosity [27]. Other researchers have attempted to predict states such as distraction, confidence, frustration, and excitement based on collections by physiological sensors [13,3,42].

This project combines two complementary techniques -- educational data mining and functional near-infrared spectroscopy (fNIRS) brain imaging – with the overall goal of understanding the cognitive processes that occur during pauses while using learning environments. fNIRS is an emerging non-invasive neuroimaging tool [15] that has been used to measure cognitive state in real-time while participants complete computer-based tasks [39]. It is now a realistic tool for researchers, being less expensive, more portable and more comfortable to wear [38]. The continuous nature of physiological measurements allow us to fill in pauses in log data with insights from brain data, with the goal of differentiating different types of pauses to better understand the learner's cognitive state. The detailed educational log data provides contextual information about what events occur before and after the pauses that will enable better interpretations of the brain activity. By combining brain and log data, we can explore critical moments and better understand what is occurring during individual use of a learning environment.

In this late-breaking work paper, we first present the theoretical foundations for our approach, describing the types of cognitive states we wish to identify and how brain data might help us do so. We then discuss the steps we have taken so far as well as future work.

Figure 1. fNIRS brain sensing setup

Background

Educational Data Mining

Our project aims to increase the potential for robust learning: learning that is retained over time, transfers to other situations, and enables future learning in new contexts [29]. There are several obstacles for the use of a learning environment to result in robust learning. For example, some students "game the system" attempting to get correct answers by using system features in unintended ways rather than by building up knowledge in the problem domain [6] often times by repeated help requests or systematically guessing. During mind wandering, another form of disengagement, students engage in internal non-task thoughts [37,33], limiting the knowledge they develop. Even motivated students may actively attend to their work in the learning environment but be unable to master the required skill, continuing to make repeated errors after many attempts (called wheel-spinning) [19].

A large focus in educational data mining research has been to predict student learning outcomes based on data on student problem-solving processes. Predictions can be made about robust learning outcomes [10,8,12, 20] and non-productive behaviors such as gaming [21, 6,34], mind-wandering [31], and wheel-spinning [19]. In these predictive models, timing information is heavily used. For example, Gowda et. al [20] used 25 features, 12 which involved timing and 4 which involved pauses, to predict whether student learning was shallow. Pauses in system logs can represent either positive learning events (i.e. the student thinking deeply about the topic) or undesirable cognitive states (i.e. mind-wandering or wheel-spinning). Some predictive models interpret fast response times as indicators of low effort and slow times as indicators of engagement in deeper processing [4,11,25]; others identify long response times as indicators of disengagement or distraction [20,10] and short times as indicators of fluency [8,9].

Understanding what occurs during pauses and using them for predictions is further complicated by the fact that a "typical" pause length varies across problems, sessions, and students [26]. Educational data mining researchers often use additional contextual features surrounding the pauses to get more accurate reads on what cognitive states occur within the pauses, spanning the instructional events triggering the response times [4], student knowledge of the relevant skill [7], or problem difficulty [11]. Despite the use of these contextual features, there are still inconsistencies across different analyses, and often features that are predictive in one analysis will not be predictive in a second. In addition, it is likely that students cycle through different cognitive states within a given pause [36].

Real-time Brain Sensing in Natural Settings with fNIRS

The variance described above motivates our ongoing brain sensing work, as we want to better understand what occurs during pauses that account for differences between problems, students, and natural variation within a student's learning behaviors. Real-time brain sensing has the potential to fill some gaps and is becoming a realistic tool for HCI researcher. In fNIRS neuroimaging, sensors are placed and secured on the head with a headband (Fig. 1). The sensors use near-infrared light to detect hemodynamic changes associated with neural activity in the brain during tasks [15].

Because fNIRS primarily detects light that travels 1-3 cm into the cortex, most fNIRS research focuses on the easily accessible frontalpolar cortex (FPC) which lies

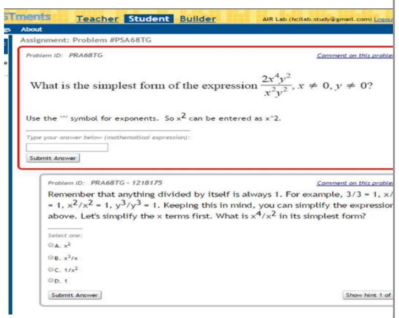

Figure 2. ASSISTments system with Scaffolding

behind the forehead. The brain processes that activate the FPC are mainly high level executive functioning, and have been shown to play a part in memory, problem solving, judgment, planning, coordinating, controlling and executing behavior. Several researchers have found that the FPC is highly activated during the development of expertise, and less activated once reached [40,29,23]. From this work, there is strong evidence that the FPC will have activation patterns that are detectable during various aspects of learning.

Most work on FPC activation was done using fMRI or positron emission tomography (PET) imaging which is not realistic for real-time, ecologically valid measurement. In contrast, fNIRS avoids many restrictions of other techniques, most notably it has significantly lower costs and far fewer mobility restrictions. Additionally, PET requires hazardous material ingestion and fMRI is not practical for use with computers due to the strong magnetic fields. Unlike these more intrusive measures, EEG has been used successfully in brain-computer interface research. However, it typically requires a longer set up time, is more susceptible to motion artifacts, and has lower spatial resolution than fNIRS.

Research Questions

We see potential in combining brain data with educational data mining for learning assessment and are currently building a large, rich dataset to explore this and build a foundation for future research. There are many research questions that can be explored with our dataset:

1. Can we detect cognitive states in learning contexts that have been detected in non-learning contexts?
2. Can we replicate findings from log data analysis with our population and learning environment?
3. What is occurring within the brain data during moments of interest in intelligent tutoring log data, represented by pause-related features typically associated with cognitive states? Does the activity appear to be congruent with what we expect to be going on? Across what percent of similar pauses?
4. How can the brain data collected augment predictions made using log data?
5. How does brain activity within pauses vary between students, and within the same students across multiple sessions?
6. How do our models interact with self-reported cognition and motivation?

Preliminary Work

We have taken several steps toward our research goals. First, we conducted a proof-of-concept study in which we collected preliminary fNIRS data from 11 participants as students used a basic e-learning platform for fraction addition. The proof-of-concept study demonstrated that we could successfully integrate brain data features with log data features. Next, we transitioned to using the ASSISTments platform (Figure 2) to better suit our needs [24]. We built up a repository of developmental math problems, those that prepare student for more advanced topics, then ran a pilot study with 3 Drexel University student participants for the purpose of testing various developmental math topics within ASSISTments. We aimed to discover which topics would be appropriate for our target population of non-STEM college students such that the students would be able to learn the topic during a tutored Intelligent Tutoring System (ITS) session. We included problems on geometry, probability, functions, algebraic simplification, factoring, square roots, inequalities, and linear equations. Of these topics, all subjects got prob-

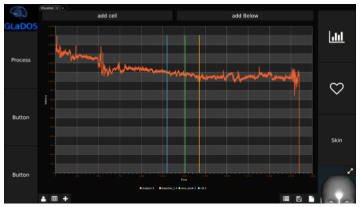

Figure 3. fNIRS Dashboard for visualization and analysis of fNIRS data.

lems on geometry, probability, algebraic simplification, and square roots incorrect on their first attempt; however, during the interview period after the ITS session, the probability question was the only one that the participants rated as difficult. The students also stated that once they remembered how to do the geometry and simplification problems they were easy. Based on these results, it was decided that probability and algebraic simplification would be the topics covered in our study. The probability topic will ideally provide insight into periods when a student does now know how to do a problem and when the student should be learning. The simplification problems will give insight into when a student needs to remember how to complete a problem or already knows how to complete it. By using these two topics, there are various cognitive and motivational states that can be accessed and predicted based on the collected log and brain data.

In addition to running the pilot study and preparing for upcoming mass data collection, we have also been developing a tool to extract learning data and visualize brain data (Figure 3). The team has been developing programs to extract and analyze important information from ASSISTments log data and developing a toolbox to analyze and visualize brain data from fNIRS. Together, this toolbox and these programs will provide a more in-depth view into a learner's state and will eventually be available for use by other researchers.

Building a Rich Dataset

With these preliminary steps complete, we are now in our data collection phase. We are iteratively collecting data from groups of 20 participants at a time, recruited from students in developmental mathematics classes at Drexel University. The goals of this collection are to 1) collect controlled data with known characteristics in well-studied tasks and 2) collect data during realistic, naturalistic learning experiences to uncover predictive features in the data and improve our understanding of learning outcomes.

Procedure

The core data collection procedure is as follows. After obtaining informed consent, participants are introduced to ASSISTments [24] and undergo a short training. We then place the fNIRS sensors on their forehead along with peripheral physiological sensors, including cardiovascular, respiratory, Galvanic skin response and eye tracking. Once the sensors are arranged, the user session proceeds through calibration tasks, a 15 minute pre-test consisting of conceptual checks, problems of mixed difficulty levels, challenge problems, and transfer problems. The participant then completes a 40-minute intervention problem set consisting of probability and simplification problems of varying difficulty in random order. The intervention problem set is "tutored", in that students will be able to request help and get feedback on their problem-solving. The problems are organized so that 10 minutes are spent on each topic at a time. The student then completes 10 minutes of tutored "challenge" problems, which are of higher difficulty than the intervention problems. The challenge problems are designed to learn how well students can use their developing knowledge to master a new topic. Finally, students take an untutored posttest isomorphic to the pretest. At the end of the session, the fNIRS as well as all physiological sensors are removed and the student undergoes a debriefing and completes a questionnaire.

Planned Analysis

Our data collection phase is ongoing. Below we describe some of our plans for analysis to answer our main research questions.

Exploratory Analyses

The large dataset we create will allow us to build robust machine learning classification models of cognitive states such as elevated cognitive workload, which has been demonstrated to be detectable by fNIRS [38,39]. A second necessary component of our research is determining the degree to which we can replicate previous educational data mining findings within our research context. We will extract features and identify feature combinations typically associated with cognitive states of interest such as *wheel-spinning* and *mind-wandering*, and use our toolkit to tag those pauses with that additional cognitive state information. Then, following the methodologies presented in previous papers, we will test a series of statistical and machine learning models for conducting feature selection and predicting learning based on log data. Our dependent variable will be the deep and shallow learning outcomes surveyed.

Core Analyses

We will identify time ranges in the log data associated with the predictive features, as well as the other theoretically significant features we have surveyed. In an exploratory fashion we will examine brain activity during those time ranges and determine interesting activity features. We will evaluate how consistent brain activity is across like regions, where it is different, and potentially why. As part of this process, we will explicitly look for variations in brain activity between students. This analysis should help us isolate the most detectable cognitive processes using the intersection of brain and log data within our particular context. We will include the brain features extracted in our learning predictive models by combining them with intelligent tutoring log features both absent of and in conjunction with pause information. We will focus on brain features that appear indicative of the detectable cognitive and motivational states and see if the brain data improves our model accuracy.

Secondary Analyses

As we begin to understand the cognitive and motivational states that are possible to detect within the brain data, we will add additional measures to try to extract further ground truth assessments of those states (i.e. we may ask students to report how distracted they were during a particular problem, or include a measure of cognitive load at the end of each session). In addition, we will include self-reported measures of the cognitive states we are examining [18]. During a learning activity, students are periodically asked to respond to (Y/N) questions about whether their mind is wandering. This self-report validation will provide an additional source of validation for the conclusions we draw regarding individual students' learning process.

Conclusion

The combination of educational data mining and brain sensing techniques has the potential to facilitate the detection of critical cognitive and motivational states during use of an online learning environment. We have conducted several pilot studies in order to explore this potential integration, and are currently engaged in a large-scale data collection. This data collection will lead to the construction of a data set that will allow us to answer important research questions relating to the nature of cognitive states during learning experiences

Acknowledgments

The authors would like to thank the following people: Stevie Mari Hawkins, Gloria Houseman, Calan Farley, Paritosh Gupta, Ruixue Liu, the Drexel Advanced Interaction Research Lab students, Robert Jacob and the Tufts HCI Lab, Missy Cummings and the MIT Humans and Autonomy Lab, Neil Heffernan, Cristina Heffernan. We would also like to thank the NSF, CRA-W and CDC for supporting this work through the CREU award.

and the design of highly adaptive learning environments that can respond to those cognitive states. We look forward to engaging with the CHI community to discuss our approach and ongoing results.

References

1. Anderson, J. R., Betts, S., Ferris, J. L., & Fincham, J. M. (2011). Cognitive and metacognitive activity in mathematical problem solving: prefrontal and parietal patterns. *Cognitive, Affective, & Behavioral Neuroscience, 11*(1), 52-67.
2. Anderson, J. R. & Pavlik, P.I. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101.
3. Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009, July). Emotion Sensors Go To School. In AIED (Vol. 200, pp. 17-24).
4. Arroyo, I., Mehranian, H., & Woolf, B. P. (2010). Effort-based tutoring: An empirical approach to intelligent tutoring. In *Educational Data Mining 2010*.
5. Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1059-1068). ACM.
6. Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *Intelligent tutoring systems* (pp. 531-540). Springer Berlin Heidelberg.
7. Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223-241.
8. Baker, R. S., Gowda, S., & Corbett, A. (2010). Automatically detecting a student's preparation for future learning: Help use is key. In *Educational Data Mining 2011.*
9. Baker, R. S., Gowda, S. M., & Corbett, A. T. (2011, January). Towards predicting future transfer of learning. In *Artificial Intelligence in Education* (pp. 23-30). Springer Berlin Heidelberg.
10. Baker, R. S., Gowda, S. M., Corbett, A. T., & Ocumpaugh, J. (2012, January). Towards automatically detecting whether student learning is shallow. In *Intelligent Tutoring Systems* (pp. 444-453). Springer Berlin Heidelberg.
11. Beck, J. E. (2004). Using response times to model student disengagement. In *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments* (pp. 13-20).
12. Beck, J. & Want, Y. (2012). Incorporating Factors Influencing Knowledge Retention into a Student Model. In *Educational Data Mining 2012*.
13. Blanchard, N., Bixler, R., Joyce, T., & D'Mello, S. (2014, January). Automated Physiological-Based Detection of Mind Wandering during Learning. In *Intelligent Tutoring Systems* (pp. 55-60). Springer International Publishing.
14. Cen, H., Koedinger, K., & Junker, B. (2006, January). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent tutoring systems* (pp. 164-175). Springer Berlin Heidelberg.
15. Cen, H., Koedinger, K., & Junker, B. (2006, January). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent tutoring systems* (pp. 164-175). Springer Berlin Heidelberg.
16. Chaouachi, M., Heraz, A., Jraidi, I., & Frasson, C. (2009). Influence of Dominant Electrical Brainwaves on Learning Performance. In *World Conference on E-Learning in Corporate, G*overnment, Healthcare, and Higher Education (Vol. 2009, No. 1, pp. 2448-2454).
17. Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. Knowledge-Based Systems,20(6), 557-574.

18. D'Mello, S. & Bixler, R. (2014). Toward Fully Automated Person-Independent Detection of Mind Wandering. In *User Modeling, Adaptation, and Personalization* (pp. 37-48). Springer International Publishing.

19. Gong, Y. & Beck, J. E. (2013). Wheel-spinning: Students who fail to master a skill. In *Artificial Intelligence in Education* (pp. 431-440). Springer Berlin Heidelberg.

20. Gowda, S. M., Baker, R. S., Corbett, A. T., & Rossi, L. M. (2013). Towards automatically detecting whether student learning is shallow. *International journal of artificial intelligence in education, 23*(1-4), 50-70.

21. Heffernan, N. T. & Walonoski, J. A. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *Intelligent Tutoring Systems* (pp. 382-391). Springer Berlin Heidelberg.

22. Heraz, A., & Frasson, C. (2009). Predicting Learner Answers Correctness through Brainwaves Assesment and Emotional Dimensions. In *AIED* (pp. 49-56).

23. Jenkins, I. H., Brooks, D. J., Nixon, P.D., Frackowiak, R. S. J., Passingham, R. E. (1994). Motor Sequence Learning: A study with positron emission tomography. *Journal of Neuroscience 14*(6): 3775-3790.

24. Heffernan, N. & Heffernan, C. (2014). The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. IJAIED. 24(4), 470-497.

25. Hershkovitz, A., & Nachmias, R. (2008). Developing a log-based motivation measuring tool. In *Educational Data Mining 2008*.

26. Hershkovitz, A., & Nachmias, R. (2009). Consistency of students' pace in online learning. In *Educational Data Mining 2009*.

27. Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In*Intelligent Tutoring Systems* (pp. 29-38). Springer International Publishing.

28. Kalyuga, S. (2007). Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective. *Educational Psychology Review*,*19*(3), 387-399.

29. Koedinger, K.R., Pavlik, P., McLaren, B.M., & Aleven, V. (2008). Is it Better to Give than to Receive? The Assistance Dilemma as a Fundamental Unsolved Problem in the Cognitive Science of Learning and Instruction. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the *30th Annual Conference of the Cognitive Science Society* (pp. 2155-2160). Austin, TX: Cognitive Science Society.

30. Leff, D. R., Elwell, C. E., Orihuela-Espina, F., Atallah, L., Delpy, D. T., Darzi, A. W., Yang, G.Z. (2008). Changes in prefrontal cortical behavior depend upon familiarity on a bimanual co-ordination task: An fNIRS study. *NeuroImage 39*. 805-813.

31. Litman, D. & Drummond, J. (2010). In the zone: Towards detecting student zoning out using supervised machine learning. In *Intelligent Tutoring Systems* (pp. 306-308). Springer Berlin Heidelberg.

32. Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and instruction*, *20*(2), 167-171.

33. Mills, C., D'Mello, S., Lehman, B., Bosch, N., Strain, A., & Graesser, A. (2013, January). What Makes Learning Fun? Exploring the Influence of Choice and Difficulty on Mind Wandering and Engagement during Learning. In *Artificial Intelligence in Education* (pp. 71-80). Springer Berlin Heidelberg.

34. Muldner, K., Burleson, W., Van de Sande, B., & VanLehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User modeling and user-adapted interaction*, *21*(1-2), 99-135.

35. Najar, A. S., Mitrovic, A., & Neshatian, K. (2014). Utilizing eye tracking to improve learning from examples. In *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge* (pp. 410-418). Springer International Publishing.

36. Shih, B., Koedinger, K. R., & Scheines, R. (2011). A response time model for bottom-out hints as worked examples. *Handbook of educational data mining*, 201-212.

37. Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review, 14*(2), 230-236.

38. Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R. J. K. (2009). Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines. *Proc User Interface Software and Technology (UIST).*

39. Solovey, E. T., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., & Jacob, R. J. K. (2012). Brainput: Enhancing Interactive Systems with Streaming fNIRS Brain Input. *Proc. ACM CHI'12 Conf.*

40. Stevens, R. H., Galloway, T., & Berka, C. (2007). EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In User Modeling 2007 (pp. 187-196). Springer Berlin Heidelberg.

41. Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, *11*(11), 1040-1046.

42. 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*(3-4), 129-164.