Abstract Title Page

Title:

EEG estimates of cognitive workload and engagement predict math problem solving outcomes

Authors and Affiliations:

Carole R. Beal & Federico Cirett Galan, School of Information: Science, Technology and Arts, The University of Arizona

Abstract Body

Background / Context:

The growth of technology-based instruction over the last two decades has led to an increased interest in estimating the extent to which students are trying to learn while at the computer. In the traditional classroom, the instructor has many sources of information about students' attention and apparent effort, but when the student is working with a keyboard and monitor and may even be at another location, it can be difficult to judge if he or she is concentrating, distracted or simply "fooling around." Recent research has focused on using behavioral data such as latencies and specific patterns of keystrokes to estimate students' engagement (Baker, Walonoski, Heffernan, Roll, Corbett & Koedinger, 2008; Beal, Mitra & Cohen, 2007; Johns & Woolf, 2006). In addition, some studies have involved the use of physical sensors such as eye tracking, skin conductance and even fMRI to track students' behavior while at the computer (Arroyo, Cooper, Burleson, Woolf, Muldner & Christopherson, 2009; Fincham & Anderson, 2010; Kappor, Burleson & Picard, 2007). Although results have been intriguing, one limitation of this line of research has been the primary focus on student motivation rather than the cognitive processes associated with learning and performance.

An alternative that is attracting attention is the use of electroencephalography (EEG), referring to patterns of electrical activity from the brain that can be detected on the scalp. Technology for capturing EEG signals has progressed considerably, to the point where the user can wear a lightweight recording unit that transmits data wirelessly for analysis. Such units have been used in a variety of tasks that require sustained attention and cognitive effort, including long-haul truck driving, missile tracking and submarine systems control, with impressive results (Berka, Levendowski et al., 2007). Education researchers have now begun to use this type of device to track students' cognitive activity during problem solving. Stevens, Galloway, Berka, Johnson and Sprang (2008) compared novices and experts as they solved a series of chemistry problems, and reported that the two groups showed distinct patterns of visual attention that correlated with problem solving time and accuracy. Mostow, Chang and Nelson (2011) used a single-channel EEG recorder with both adults and children as they read difficult and easy text passages. The EEG data were used to train a machine classifier, and results were significantly better than chance at discriminating the reading of adults and children, as well as predicting the difficulty of the text. Chaouachi, Jraidi and Frasson (2011) recorded EEG activity with a six-channel unit as users solved cognitive tasks varying in difficulty, and verified that their estimates of the users' mental workload were correlated with the difficulty of the tasks. Thus, initial results appear to suggest that EEG might be a valuable technology for directly assessing a student's level of cognitive effort.

Purpose / Objective / Research Question / Focus of Study:

In the present study, we focused on the use of EEG data about cognitive workload and sustained attention to predict math problem solving outcomes. EEG data were recorded as students solved a series of easy and difficult math problems. Sequences of attention and cognitive workload estimates derived from the EEG signals were used to train a Support Vector Machine ([SVM], a machine learning classifier) to predict the outcome of the problem: correct or incorrect answer. We were also interested in learning if the EEG estimates would be different for easy and hard problems, as suggested by the results of Chaouachi et al. (2011), and if the estimates would be related to the students' self-report of how difficult the problem was.

Setting:

Data for this study were collected in a laboratory setting.

Population / Participants / Subjects:

There were 16 participants in the study (8 males, 8 females). Participants were college students who were at least 18 years old and gave active written consent for participation. They received either a small stipend or course credit for participation.

Intervention / Program / Practice:

Each person participated in a 90 minute session, which included informed consent procedures, fitting the EEG headset, completing a 30 minute baseline calibration task, and solving math problems presented at the computer while wearing the EEG headset.

Research Design:

Participants solved eight multiple-choice math problems presented at the computer while wearing the EEG headset. Math problems were taken from a set of released SAT items; there were four easy problems and four hard problems, with difficulty level determined by information from the College Board. Each problem had four answer options. The items were presented to students within an online tutoring system that recorded the time on the problem (initial presentation on the screen to first answer selection) as well as the outcome (correct, incorrect answer chosen). Problems were presented in one of two sequences (easy, easy, hard, hard, easy, easy, hard, hard or hard, hard, easy, easy, hard, hard, easy, easy) across subjects. Participants were provided with an eight-page paper booklet to use as they solved the math problems. Each page included an area for working out the problem on paper (if needed). After completing each problem, participants were also asked to rate how difficult the problem had been, how confident they felt about their performance, and their level of frustration. Ratings were made on a five-point Likert-type scale.

Data Collection and Analysis:

Collection. The electroencephalogram (EEG) data were recorded from nine sensors integrated into a mesh cap covering the upper half of the head, along with two reference signals attached to the mastoid bones (behind the ears) and two sensors attached to the right clavicle and to the lowest left rib to record the heart rate (although the heart rate data were not used in the study). The location of each sensor was determined by the International 10-20 System [10] to ensure standardized reproduction of tests. This cap was equipped with a small wireless transmission unit. A small USB dongle received the wireless transmissions to a PC computer with Windows (XP/Vista/7) 32 bit operating system.

Processing. Each second, 256 EEG signals were transmitted from the unit and converted to Theta, Alpha, Beta and Sigma wave signals (ranging from 3 Hz to 40 Hz). These signals were processed by Advanced Brain Monitoring (2011) proprietary software to produce classifications of mental states, meaning the probability that the participant was in a particular state in epochs of one second. States included Engagement, Distraction, Drowsiness and Cognitive Workload. Engagement includes estimates of cognitive activities such as information gathering, visual scanning and sustained attention, and Workload is a measure of effortful cognitive activity (Berka et al., 2007). We focused on Engagement and Workload because levels of Drowsiness were very low in the study, and Distraction is essentially the inverse of Engagement.

SVM training. The Engagement signals (one per second) were processed by converting each raw signal into one of three equal-sized bins, with limits set at 0.333 and 0.666 of the Cumulative Distribution Function. The count of signals below 0.333 were considered as the Low State, values between 0.333 and 0.666 as the Medium state, and those between 0.666 and 1.0 as the High state. Then the Engagement signal sequence was scanned and tagged to produce a transition probability table (TPT). The TPT consisted of nine cells representing the transition probabilities between the three states. The TPT was then tagged with the problem outcome (+1 for a correct answer and -1 for an incorrect one). A file was constructed with 16 records (one for each participant) including the outcome tag and nine features to serve as input to a Support Vector Machine (SVM). Because the sample size was small (each problem had only 16 records) libsvm (Fan, Chen & Lin, 2005) was used with leave-one-out cross-validation enabled. A similar method was used to process the Workload signals.

Findings / Results:

Problem solving results. Of the 16 participants, 15 completed all eight problems and one completed seven of the eight items. The outcome data set thus consisted of 127 completed math problems, with 49 answered incorrectly (38.6%).

Predictions from EEG. Prediction accuracies from the SVM classifier are shown in Table 1. As may be seen in Table 1, the SVM predictions were consistently higher than chance (25%, or one answer out of four possible) and also higher than the base rate of performance for each problem. That is, using the EEG data to predict whether the student would solve the problem correctly is more accurate than simply knowing that, for example, there is a 56% chance on average of a correct solution to the Towns problem, based on the actual performance of the problem solvers.

(please insert Table 1 about here)

As shown in Table 1, there were some differences for the easy versus the hard problems: On average, the SVM classifier performed better on the easy problems, with the Engagement and Workload signals being equally good predictors (83% both). Combining the Engagement & Workload signals (six features in total) as input for the SVM actually reduced the classification accuracy slightly (80%). On hard problems, the Engagement signal was a better predictor than the Workload signal (78% vs. 67%). Using the maximum value between the result of the SVM classification of the Engagement and Workload signals incremented the cross validation accuracy for the easy problems up to 87% and for the hard problems to 78%.

Relation with self-report. We were interested in whether students' self-reported experiences while problem solving might be related to their cognitive workload as indicated by the EEG estimates. We started by attempting to verify that the hard problems actually seemed more challenging. Participants took more time to answer the hard problems than the easy ones. A matched-pairs t-test indicated that this difference was significant, ($t(15) = 3.746$, $p < .01$). Subjects rated themselves as more frustrated on hard problems than on easy problems, $t(14) = 2.84$, $p < .05$, and rated hard problems as more difficult than easy problems, $t(12) = 6.273$, $p < .01$. Also, as shown in Table 1, students' problem solving performance was worse for the hard problems on average relative to the easy problems.

In an exploratory analysis, we considered the extent to which the reports of problem difficulty, frustration and confidence were correlated. Results are shown in Table 2. The strong inter-correlations suggested a principal components analysis, which indicated that there was one underlying factor that accounted for 72% of the variance in the self-report responses. Therefore, we extracted the estimate for this latent factor for each participant, and then used the mean workload scores to predict these estimates. Results indicated that for individuals with higher average workload scores, the self-reported experience was one of reduced confidence, greater frustration and the perception that the problems were more challenging.

(please insert Table 2 about here)

Conclusions:

These preliminary results suggest that using the estimates of Engagement and Workload from EEG data can predict problem solving outcomes better than the base rate of performance on the problems. In general, predictions that utilized both signal sources tended to be better, although this was not consistently observed. Predictions were somewhat better for easy problems, perhaps because students solved those items more quickly; prediction accuracy tends to decline with the amount of time the student works on the problem, perhaps due to increased noise in the sensor data streams. Alternatively, we are investigating internal structure in the more extended problem solving sequences, for example, to try to identify points where the student may realize that he or she needs to reexamine the problem and try another approach. If students' problem solving has such structure, for example, re-reading the problem and starting over two minutes into the problem, it may be more challenging for the machine classifier to make good predictions from the complete sequence. Applying methods for finding structure in multivariate time series data would be a logical next step to address this possibility.

One major limitation of the machine learning approach used in this research is that it is typically not possible to identify what specific components or characteristics of the source data are most important in making the classification of a right or wrong outcome. On the other hand, it may not be necessary if the goal is to use the data to understand what the student is doing and whether intervention would be useful. An advantage of the machine learning approach is that, once trained with a subset of data, the SVM classifier can run in real-time, raising the possibility that the EEG data could be used to intervene proactively, that is, well before the student is about to answer incorrectly. In recent work, we are examining the quality of the predictions made during the course of the problem solving sequence, and initial results suggest that good accuracies are obtained within the first 20 seconds after the student is presented with the problem. Overall, the results were consistent with other recent studies suggesting that the application of neuroscience methods may be a valuable addition to education research.

Appendices

Appendix A. References

Advanced Brain Monitoring. (2011). B-Alert X-10, 2011, http://www.b-alert.com/x10.html.

Arroyo, I., Cooper, D., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In Proceedings of the 2009 Conference on Artificial Intelligence in Education, pp. 17-24.

Baker, R. S., Walonoski, J., Heffernan, N., Roll, I., Corbett, A. & Koedinger, K. (2008). Why students engage in gaming the system behavior in interactive learning environments. Journal of Interactive Learning Research, 19, 185-244.

Beal, C. R., Mitra, S., & Cohen, P. R. (2007). Modeling learning patterns of students with a tutoring system using Hidden Markov Models. Proceedings of the 1007 Conference on Artificial Intelligence in Education, pp. 238-245.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T.,Olmstead, R. E., Tremoulet, P. D. & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, Space and Environmental Medicine, 78(5 Suppl), B231-B244.

Chaouachi, M., Jraidi, I., & Frasson, C. (2011). Modeling mental workload using EEG features for intelligent systems. In J. A. Konstan et al. (Eds.), Proceedings of UMAP, Lecture Notes in Computer Science 6787, pp. 51-61. Berlin: Springer-Verlag.

Fan, R. E., Chen, P. H. & Lin, C. J. (2005). Working set selection using second order information for training Support Vector Machines. Journal of Machine Learning Research, 6, 1889-1918.

Fincham J. M., & Anderson J. R. (2010). Using neural imaging and cognitive modeling to infer mental states while using an Intelligent Tutoring System. In Baker, R.S.J.d., (Ed.), Proceedings of the 3rd International Conference on Educational Data Mining, Vol. 3, pp. 51-60.

Johns, J., & Woolf, B. P. (2006). A dynamic mixture model to predict student motivation and proficiency. Proceedings of the American Association of Artificial Intelligence, IOS Press, Boston MA.

Kapoor, A., Burleson, W. & Picard, R.W. (2007). Automatic prediction of frustration, International Journal of Human-Computer Studies, 2007, 65(8), 724-736.

Mostow, J., Chang, K., & Nelson, J. (2011). Towards exploiting EEG input in a reading tutor. Proceedings of the 15th International Conference on Artificial Intelligence in Education, pp. 230-237. Berlin: Springer Verlag.

Stevens, R. H., Galloway, T., Berka, C., Johnson, R., & Sprang, M. (2008). Assessing students' mental representations of complex problem spaces with EEG technologies. Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomic Society, New York, NY.

Appendix B. Tables and Figures
Not included in page count.

	Cross validation accuracy				
Easy problems	Base rate Accuracy	Engagement	Workload	Eng. & Wkld	Max(Eng, Wkld)
Bows	75%	75%	88%	88%	88%
Class	50%	88%	69%	75%	88%
Summation	93%	93%	93%	93%	93%
Village	44%	75%	81%	63%	81%
Average	**65%**	**83%**	**83%**	**80%**	**87%**
Hard problems	Base rate Accuracy	Engagement	Workload	Eng. & Wkld	Max(Eng, Wkld)
Bus	62%	75%	75%	69%	75%
Fence	62%	81%	69%	63%	81%
Towns	56%	81%	63%	69%	81%
Triangle	50%	75%	63%	63%	75%
Average	**57%**	**78%**	**67%**	**66%**	**78%**
Both	**62%**	**80%**	**75%**	**73%**	**83%**

Table 1. Accuracy of SVM predictions for easy and hard math problems, relative to base rate accuracy for each problem.

	Confidence	Frustration	Difficulty
Confidence	1	-0.65*	-0.81**
Frustration	-0.65*	1	0.64*
Difficulty	-0.81**	0.64*	1

Table 2. Correlations for self-reports of confidence, frustration and problem difficulty. * p < .05, ** p < .01.