



Utilizing Log-Based and Neurophysiological Measures to Understand Engagement and Learning with Intelligent Tutoring Systems

Yushuang Liu¹ , Ido Davidesco¹ , Bruce McLaren² , J. Elizabeth Richey³ , Xiaorui Xue¹, Leah Teffera², Hayden Stec², Hyosun Lee⁴, Jiayi Zhang⁵, Suyi Liu¹, and Elana Zion-Golumbic⁶

¹ Boston College, Chestnut Hill, MA 02467, USA
yushuang.liu@bc.edu

² Carnegie Mellon University, Pittsburgh, PA 15213, USA

³ University of Pittsburgh, Pittsburgh, PA 15260, USA

⁴ University of Connecticut, Storrs, CT 06269, USA

⁵ University of Pennsylvania, Philadelphia, PA 19104, USA

⁶ Bar-Ilan University, 5290002 Ramat Gan, Israel

Abstract. Computer-based intelligent tutoring systems (ITSs) are a key to learning in many educational scenarios, but not all students engage effectively with them. Student engagement with ITSs is typically assessed via log data (e.g., problem-solving time), which only partially captures its multidimensional nature. In the current project, we studied how students engage with an intelligent tutor using a combination of log-based, electroencephalography (EEG), and eye-tracking measures. A total of 56 high school students participated in a school-based study with a pretest-intervention-posttest design. During the intervention, students watched short videos and solved chemistry problems using an ITS. Preliminary analysis shows that log-based measures substantially improved model fit, explaining 26% more variance in posttest scores than a model using only pretest scores. Within the model, error rate and hint request rate were significant predictors, while self-reported effort and time spent on problems were not. Hint requests were negatively associated with learning outcomes, emphasizing the need to assess how students use hints and other assistive features of ITSs. Adding EEG and eye-tracking measures did not significantly improve the overall performance of the model. However, a more fine-grained, problem-by-problem analysis revealed that frontal EEG alpha power significantly predicted the number of errors on individual intervention problems. In the future, this finding may support the development of EEG-informed ITSs that can provide more personalized assistance. More generally, these results highlight the value of multimodal measures of engagement in educational technology research.

Keywords: Intelligent tutoring systems · engagement · EEG · eye-tracking

1 Introduction

Computer-based intelligent tutoring systems (ITSs) are a critical component in many educational settings, particularly in K-12, including in-person and online learning. ITSs are designed to complement traditional classroom learning by providing students with personalized guidance and support that are tailored to their learning progress and individual needs [1–3]. ITSs are designed to be engaging, but not all students learn effectively with them, possibly due to differences in engagement [4, 5].

The current study builds on the Interactive-Constructive-Active-Passive (ICAP) theoretical framework, which proposes a direct link between the level of engagement in an instructional activity and its corresponding learning outcomes [6]. According to ICAP, passive activities, such as listening to a lecture, lead only to superficial learning (e.g., recall). Conversely, constructive and interactive activities that require high levels of engagement, such as solving science problems, ultimately lead to robust learning [7]. This framework suggests that examining student engagement with ITSs may help explain differences in learning outcomes.

There is a growing consensus that engagement should be viewed as a multidimensional construct [8, 9]. Fredricks and colleagues [10] proposed three components of engagement: (1) Behavioral engagement: learners’ participation in the learning process, including effort and persistence (e.g., note-taking); (2) Cognitive engagement: learners’ mental investment in the learning task, such as allocation of attention; and (3) Affective engagement: feelings and attitudes about the learning task or learning context (e.g., interest, boredom). Typically, students’ engagement with ITSs is measured via log-based student interaction data (e.g., time to complete a problem), but these data do not fully capture all three dimensions of engagement. For example, it is challenging to determine how students cognitively engage with assistive features of ITS - such as hints - based on log-data alone (do students read hints?).

In the current study, we incorporated a set of neurophysiological measures (Electroencephalography and eye-tracking) to capture a more comprehensive picture of how students engage with ITSs. Specifically, we focus on EEG activity in the alpha band (8–12 Hz), which has been associated with lapses of attention [11, 12] and mind-wandering [13, 14]. We also focus on gaze position as a proxy for the locus of attention [15–17]. These two neurophysiological measures could provide information on the cognitive engagement of learners [9].

The current study addresses the following research questions: RQ1: What are the relations between tutor log-based and neurophysiological measures? RQ2: How much of the variability in learning outcomes with ITS can be attributed to log-based and neurophysiological measures?

2 Methods

2.1 Participants

56 students participated voluntarily ($Mean_{age} = 15.89$ years, 52% female, 66% white), currently taking a chemistry course, from two high schools in the USA. All received permission from their caregivers, and signed an assent form, as required by our Institutional Review Board. We excluded 3 students from EEG and 12 from eye-tracking

due to poor data quality (EEG: >50% artifacts; Eye-tracking: > 30% missing or >1° calibration error).

2.2 ITS

The specific ITS platform that was used in this project is a web-based chemistry tutor that has been validated extensively in previous studies (Fig. 1) [18–20]. During tutored problems, students could request context-sensitive and on-demand hints, as well as access an on-screen calculator. As students worked through each problem, the tutor also provided automatic error messages, adaptive support, and correctness feedback.

2.3 Procedure

Students participated in a self-paced study (~3 h) in a quiet school office. They first completed a pretest (~ 50 min), consisting of four stoichiometry problems (isomorphic to the tutored problems) and three conceptual knowledge questions. Within one week, students participated in the EEG and eye-tracking session (preparation: ~ 30 min), including watching content-related videos and solving six tutored problems, each followed by a self-reported effort rating (~ 1 h). A posttest concluded the session (~ 40 min). Pre/posttests had two isomorphic forms (A and B), counterbalanced across students.

2.4 Log-Based Measures

The current paper focuses on three measures that were extracted from the log file for each tutored problem – duration on problem, number of hint requests, number of errors—as well as self-reported efforts.

2.5 Eye-Tracking

Eye-tracking data were collected using a Tobii Pro Fusion screen-based eye tracker. The ITS display was divided into eight Areas of Interest (AOIs): problem statement, hint window, calculator, four problem steps, and the final result (see Fig. 1). The current paper focuses on the proportion of time looking at each AOI.

2.6 EEG

EEG was recorded using a 32-channel mbt Smarting Pro system. Data were band-pass filtered (1–30 Hz), cleaned of noisy channels/periods, and re-referenced. Ocular/muscle artifacts were removed via ICA. Preprocessed data were epoched into 1-s intervals, tapered with a Hanning window, and transformed (4–30 Hz) via FFT. Alpha power (8–12 Hz) was normalized by total 4–30 Hz power [21], then averaged across frontal, central, and posterior ROIs. Analyses was carried using Matlab R2024b, with EEGLAB 2024.0 [22] and FieldTrip [23].



Fig. 1. A screenshot of the Chemistry Tutor with areas of interest for eye gaze analysis.

3 Results

Students' averaged 3.54 (SD = 1.22) on the pretest and averaged 4.37 (SD = 1.08) on the posttest, out of a total possible score of seven. A paired-sample t-test indicated that students' performance significantly improved ($t(55) = 5.55, p < .001$).

To address RQ1 ("What are the relations between tutor log-based and neurophysiological measures?"), we first assessed the correlations among log-based measures for each tutored problem. We also averaged the measures across problems and computed the correlations between the aggregate measures. Overall, self-reported effort, problem duration, number of errors, and number of hint requests were positively, significantly correlated with each other (Table 1).

Next, a series of linear regressions were used to test whether the neurophysiological measures significantly predicted each one of the log-based measures. This preliminary analysis indicated that in four out of the six tutored problems, frontal EEG activity in the alpha band significantly predicted the number of errors on problems ($\beta_s = -.91$ to $-.52$ and $p_s = .001$ to $.04$). The regression coefficients were negative, indicating that higher alpha power was associated with fewer errors. Alpha power did not significantly predict self-reported effort, duration, or number of hint requests. Among the eye-tracking measures, the only significant association across problems was between the proportion of time looking at the hint window and number of hint requests.

To address RQ2 ("How much of the variability in learning outcomes with ITS can be attributed to log-based and neurophysiological measures?"), a regression model was used to predict total posttest performance based on pretest performance. Adding the log-based measures to the model explained significantly more variance than pretest alone ($R^2_{\text{Change}} = .27, p < .001$). Specifically, the model predicting posttest performance with pretest explained 27.6% of the variance ($F(1, 54) = 20.58, p < .001, R^2 = .28, R^2_{\text{Adjusted}} = .26$), while the model with the log-based measures predicted 53.6% of the variance ($F(5, 50) = 11.56, p < .001, R^2 = .54, R^2_{\text{Adjusted}} = .49$). Within the extended model, pretest score ($\beta = .28, p = .017$) was a significant, positive predictor and total number of

errors ($\beta = -.45, p = .018$) and total number of hint requests ($\beta = -.29, p = .034$) were significant, negative predictors of posttest score. Duration on problem ($\beta = .15, p = .41$) and self-reported effort ($\beta = -.04, p = .66$) were not significant predictors of posttest performance. A preliminary analysis that included the neurophysiological measures did not improve the model ($R^2_{\text{Change}} = .11, p = .47$).

Table 1. Correlations among test scores, log-based measures, and neurophysiological measures

	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Pretest	.53*	-.51*	-.31*	-.50*	-.03	-.11	-.05	-.04	.02	.11	.03	-.25	.02	-.12	.08
2. Posttest		-.67*	-.58*	-.55*	-.04	-.09	-.15	-.08	.01	-.24	.22	-.04	.004	-.14	.09
3. Errors			.67*	.82*	.03	.27	.27*	.20	-.08	-.03	-.34*	.13	.15	.18	-.02
4. Hints				.63*	-.05	.31*	.37*	.29*	-.08	-.01	-.43*	.05	.03	.25	.06
5. Duration					.03	.34*	.28*	.26	.04	-.11	-.31*	.19	.16	.07	-.07
6. Effort						-.15	-.03	-.06	.45*	-.14	-.35*	-.32*	.11	.22	-.17
7. Alpha frontal							.85*	.84*	-.11	.07	.21	.04	-.12	-.34*	.03
8. Alpha central								.85*	-.04	.14	.13	-.05	-.12	-.37*	.04
9. Alpha posterior									-.002	.16	.14	-.04	.05	-.38*	-.02
10. Gaze hint										-.11	-.23	-.22	.06	-.07	-.61*
11. Gaze area 1											.14	-.33*	-.08	-.25	-.07
12. Gaze area 2												.04	.04	-.39*	-.25
13. Gaze area 3													.02	-.09	-.04
14. Gaze area 4														.14	-.32*
15. Gaze solution															-.15
16. Gaze stat.															

Note. Gaze = proportion of time spent looking at each AOI. Stat. = Problem Statement. (*) $p < .05$.

4 Discussion

In this study, we examined how log-based ITS measures relate to neurophysiological measures of engagement (RQ1) and how each set of measures predict learning outcomes (RQ2). Consistent with prior ITS research, our analysis indicates that log-based measures are strongly correlated among themselves [24]. Among the EEG measures, our preliminary analysis indicates that frontal EEG activity in the alpha band was significantly associated with numbers of errors. Previous studies reported that increases in alpha power are associated with lapses of attention and disengagement [21, 25]. However, most of these studies have focused on alpha activity measured in posterior areas of the brain. Our results suggest that increased frontal alpha activity is associated with solving problems with fewer errors. This aligns with recent findings linking frontal alpha activity to internally demanding processes and deeper understanding [26, 27].

Whereas EEG activity in the alpha band significantly predicted error numbers, it was not predictive of self-reported effort, duration, or number of hint requests. This is

surprising, considering prior research indicates that alpha activity is a marker of task difficulty and cognitive load [28]. It is possible, however, that when students report their effort level after a problem has been solved, their reports do not accurately reflect their cognitive load during problem solving [29]. Similarly, hint requests and problem duration might be influenced by ‘gaming the system’, not just by cognitive load [30]. Therefore, future analyses will incorporate gaming detectors.

Our analysis of eye-tracking data has, so far, yielded just one significant result: The proportion of time students spent looking at the hint window was predictive of hint requests. Based on prior research, other eye-tracking metrics (e.g., transitions, fixations, gaze dispersion) may yield more insight [31–33].

With regard to predicting learning outcomes with ITS, our results indicate that other than pretest performance, the only significant predictors of posttest performance were hint requests and number of errors. This replicates a common finding in ITS research: hint requests and error rates during learning predict posttest performance when accounting for pretest [24]. Both log-based measures were negatively associated with posttest performance: students who made more errors and requested more hints performed more poorly on the posttest. Indeed, numerous hint requests are known to be an indicator of ‘gaming the system’ [30]. Interestingly, self-reported effort did not significantly predict posttest performance. This is surprising given prior research on cognitive load and learning outcomes [34]. However, as previously mentioned, it is possible that self-reported effort does not accurately capture cognitive load during problem solving.

Finally, our analysis suggests that incorporating EEG and eye-tracking measures did not improve the prediction of posttest performance. This analysis is preliminary as it only included one EEG measure (alpha power) and one eye-tracking measure (proportion of looking time per AOI). Consistent with our findings, recent research suggests that alpha power is a sensitive measure of moment-to-moment changes in engagement, not overall learning [21]. It is very possible that other EEG measures would be more predictive of learning outcomes. One such measure might be theta power, which has been associated with memory retention [28]. Similarly, more complex eye-tracking measures might have better predictive value.

Taken together, our results indicate that EEG measures, especially frontal alpha activity, relate to how students engage with ITSs. These findings suggest that neurophysiological measures could uncover aspects of ITS engagement that are not easily detectable using log-based data. This is important because according to the ICAP framework, there is a direct link between the level of cognitive engagement and learning outcomes. While it is assumed that the interactive design of ITSs promotes cognitive engagement, engagement with ITSs varies within and across students [35]. EEG measures could potentially help assess student cognitive engagement more accurately and ultimately contribute to the design of more effective ITSs. Subsequent analyses within the current study as well as future studies should examine a wider range of EEG and eye-tracking measures and investigate the unique contribution of these measures to our understanding of student engagement and learning with ITS.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. 2141139. We thank the participating schools, students, and families, as well as M. Coburn, B. Cary, A. Pradeep, and M. Salha for their help with data collection.

References

1. Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S.R., Rakhshan, M., Keikha, L., Ghazi Saeedi, M.: Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interact. Learn. Environ.* **29**, 142–163 (2021)
2. VanLehn, K.: The behavior of tutoring systems. *Int. J. Artif. Intell. Educ.* **16**, 227–265 (2006)
3. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ. Psychol.* **46**, 197–221 (2011)
4. D’Mello, S., Dieterle, E., Duckworth, A.: Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educ. Psychol.* **52**, 104–123 (2017)
5. Koedinger, K.R., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. *Educ. Psychol. Rev.* **19**, 239–264 (2007)
6. Chi, M.T.H., Wylie, R.: The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educ. Psychol.* **49**, 219–243 (2014)
7. Chi, M.T.H., et al.: Translating the ICAP theory of cognitive engagement into practice. *Cogn. Sci.* **42**, 1777–1832 (2018)
8. Schwartz, R.N., Plass, J.L.: Types of engagement in learning with games. In: Plass, J.L., Mayer, R.E. (eds.) *Handbook of Game-Based Learning*, pp. 53–80. MIT Press (2020)
9. D’Mello, S.K.: Improving student engagement in and with digital learning technologies. In: *Pushing the Frontiers with Artificial Intelligence. Blockchain and Robots*, pp. 79–104. OECD Publishing, Paris (2021)
10. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: potential of the concept, state of the evidence. *Rev. Educ. Res.* **74**, 59–109 (2004)
11. Clayton, M.S., Yeung, N., Cohen Kadosh, R.: The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* **19**, 188–195 (2015)
12. Jensen, O., Mazaheri, A.: Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* **4**, 186 (2010)
13. Baldwin, C.L., et al.: Detecting and quantifying mind wandering during simulated driving. *Front. Hum. Neurosci.* **11**, 406 (2017)
14. Compton, R.J., Gearinger, D., Wild, H.: The wandering mind oscillates: EEG alpha power is enhanced during moments of mind-wandering. *Cogn. Affect. Behav. Neurosci.* **19**, 1184–1191 (2019)
15. Casarotti, M., et al.: Paying attention through eye movements: A computational investigation of the premotor theory of spatial attention. *J. Cogn. Neurosci.* **24**, 1519–1531 (2012)
16. Risko, E.F., Kingstone, A.: Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* **73**, 291–296 (2011)
17. Geisler, W.S., Cormack, L.K.: Models of overt attention. In: Liversedge, S., Gilchrist, I.D., Everling, S. (eds.) *The Oxford Handbook of Eye Movements*, pp. 439–454. Oxford University Press, Oxford (2011)
18. McLaren, B.M., et al.: The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Comput. Hum. Behav.* **55**, 87–99 (2016)
19. McLaren, B.M., Lim, S.-J., Koedinger, K.R.: When is assistance helpful to learning? In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) *Intelligent Tutoring Systems*, pp. 677–680. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69132-7_75
20. McLaren, B.M., Isotani, S.: When is it best to learn with all worked examples? In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *Artificial Intelligence in Education*. Springer, Heidelberg, pp. 222–229 (2011). https://doi.org/10.1007/978-3-642-21869-9_30
21. Davidesco, I., et al.: Detecting fluctuations in student engagement and retention during video lectures using electroencephalography. *Br. J. Educ. Technol.* **54**, 1895–1916 (2023)

22. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004)
23. Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M.: FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, Article ID 156869 (2011)
24. Alevén, V., Koedinger, K.R.: Investigations into help seeking and learning with a cognitive tutor. In: *Papers of the AIED-2001 Workshop on Help Provision and Help Seeking in Interactive Learning Environments*, pp. 47–58 (2001)
25. Boudewyn, M.A., Long, D.L., Swaab, T.Y.: Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Front. Psychol.* **4**, 60 (2013)
26. Farkish, A., et al.: Evaluating the effects of educational multimedia design principles on cognitive load using EEG signal analysis. *Educ. Inf. Technol.* **28**, 2827–2843 (2023)
27. Pi, Z., et al.: Learning by explaining to oneself and a peer enhances learners' theta and alpha oscillations while watching video lectures. *Br. J. Educ. Technol.* **52**, 659–679 (2021)
28. Antonenko, P.D.: Educational neuroscience: exploring cognitive processes that underlie learning. In: Parsons, T.D., Lin, L., Cockerham, D. (eds.) *Mind, Brain and Technology*. Springer, Cham, pp 27–46 (2019). https://doi.org/10.1007/978-3-030-02631-8_3
29. Davidesco, I., et al.: Neuroscience research in the classroom: Portable brain technologies in education research. *Educ. Res.* **50**, 649–656 (2021)
30. Paquette, L., et al.: Cross-system transfer of machine learned and knowledge engineered models of gaming the system. In: Ricci, F., Bontcheva, K., Conlan, O., Lawless, S. (eds.) *User Modeling, Adaptation and Personalization*. Springer, Cham, pp 183–194 (2015). https://doi.org/10.1007/978-3-319-20267-9_15
31. Kardan, S., Conati, C.: Comparing and combining eye gaze and interface actions for determining user learning. In: Carberry, S., Weibelzahl, S., Micarelli, A., Semeraro, G. (eds.) *User Modeling, Adaptation, and Personalization*. Springer, Berlin, Heidelberg, pp 215–227 (2013). https://doi.org/10.1007/978-3-642-38844-6_18
32. Lallé, S., et al.: The impact of student individual differences and visual attention to pedagogical agents during learning with MetaTutor. In: André, E., Baker, R., Hu, X., Rodrigo, M.M.T., Du Boulay, B. (eds.) *Artificial Intelligence in Education*. Springer, Cham, pp 149–161 (2017). https://doi.org/10.1007/978-3-319-61425-0_13
33. Zhang, H., et al.: Wandering eyes: eye movements during mind wandering in video lectures. *Appl. Cogn. Psychol.* **34**, 449–464 (2020)
34. Sweller, J.: Cognitive load theory: recent theoretical advances. In: Plass, J.L., Moreno, R., Brünken, R. (eds.) *Cognitive Load Theory*, pp. 29–47. Cambridge University Press, Cambridge (2010)
35. Mills, C., et al.: Mind wandering during learning with an intelligent tutoring system. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.F. (eds.) *Artificial Intelligence in Education*, vol 9112. LNCS. Springer, Cham, pp 267–276 (2015). https://doi.org/10.1007/978-3-319-19773-9_27