Personalised Holistic Learning Performance Assessment: A Multimodal BCI Approach

1. Introduction

Learning performance, defined in this report as how well one is able to acquire knowledge or skill by instruction or study [1], is an important feature that needs to be evaluated of a student. In assessing students' learning performance, we are concerned not just with grades, but also with students' engagement level. Student engagement refers to the degree of attention, curiosity, interest, optimism and passion that students show when they are learning [2]. Learning performance improves when students are inquisitive, interested, inspired, and suffers when students are bored, dispassionate, disaffected or otherwise "disengaged".

Traditionally, teachers evaluate student engagement by interacting with students, observing them, and monitoring their grades. However, electronic methods of measuring student engagement are required in the e-learning context. One electronic method of measuring engagement is by measuring the electrical activity of students' brains, called electroencephalography (EEG). EEG signals are a mixture of several underlying base frequencies, which are considered to reflect certain cognitive affective or attentional states [3]. The different frequency bands are the delta, alpha, beta and gamma bands. The beta band can be further classified into Sensorimotor Rhythm (SMR) waves (12 - 15 Hz) and Mid beta (15 - 20 Hz) waves [4]. Both engagement indices, namely beta/(alpha + theta), beta/alpha, or 1/alpha [5], and the Concentration Power Index, calculated (SMR + Mid Beta)/Theta [6, 7], measure mental engagement. Frontal asymmetry, calculated $\log(\frac{Power_{left}-Power_{right}}{Power_{left}+Power_{right}})$, measures emotional engagement, where higher band power in the left versus the right frontal cortex indicates positive feelings, engagement and motivation [8]. However, the disadvantage in using EEG alone to determine engagement is that EEG cannot determine if students are engaged in e-learning or in other irrelevant activities. For example, a student engaged in reading an irrelevant article while accessing the e-learning material would be wrongly assessed as being engaged. Other physiological senses need to be measured alongside EEG signals in order to increase accuracy of engagement measurement.

Proposed Solution

Given that current measures of engagement lack comprehensiveness and hence accuracy, we propose an improved Engagement Index (fig. 1) which assesses student engagement using a multimodal brain-computer interface (BCI). The modalities include EEG signals, eye gaze, galvanic skin response (GSR) and heart rate variability. Each modality contributes different insights on student engagement, and when integrated, provides a comprehensive measure of student engagement in terms of both attention levels and affective states.

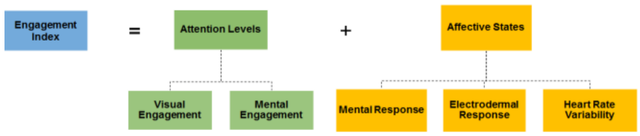


Fig. 1: Proposed Engagement Index

We hypothesize that our proposed Engagement Index is a more holistic and accurate measurement of student engagement, and will function as a strong indicator of learning performance. Ultimately, this will aid in helping educators to better assess and understand students' response to learning materials and their learning performance, enabling more effective intervention and meaningful improvements in learning material.

2. Methodology

2.1 Experimental Setup

10 healthy students (aged 17-22, 7 male and 3 female) took part in our experiment. Each session lasted approximately 1 hour. Students' EEG signals, eye gaze coordinates, heart rate and GSR were recorded throughout the experimental session.

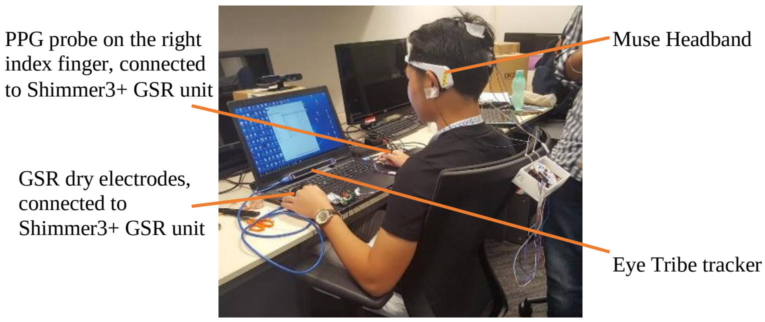


Fig. 2: Diagram showing experimental setup

Modality	Device Used	Raw Variables Measured	Recording software
EEG signals	Muse	Alpha, beta and theta absolute power spectrum measured from the frontal lobe	MuseLab

Eye gaze	Eye Tribe tracker	Eye gaze coordinates	Eye Tribe software developme nt kit
Heart rate variabilit y	Photopleth ysmogram (PPG) probe on the right index finger, connected to Shimmer3 + GSR unit	Interbeat (R-R) Interval	Shimmer Capture
GSR	GSR dry electrodes on the left index and middle fingers, connected to Shimmer3 + GSR unit	Skin conductanc e	

Table 1: Devices and software used to measure and log raw variable data of multiple modalities

2.2 Experiment Design

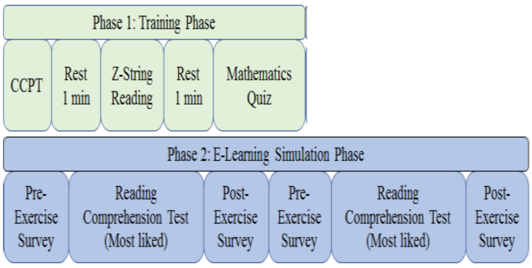


Fig. 3: Flow of Events in One Experimental Session

<u>Phase 1: Training Phase</u>

In Phase 1, participants' data is used to train the Support Vector Machine (SVM) classifier. In the Conjunctive Continuous Performance Test (CCPT), shapes of different colours would flash on screen. Participants had press the spacebar as fast as they could whenever they saw a red square. The purpose of this test was to measure sustained attention through reaction time. In the Z-String Reading task, participants had to pretend to read a paragraph of "text" comprised entirely of the letter z. The purpose of this task was to simulate mindless reading. Lastly, in the mathematics quiz, participants had to answer 10 difficult mental sums within 10 seconds each. The purpose of this was to simulate response to high cognitive workload and invoke stress.

<u>Phase 2: E-Learning Simulation</u>

Task	**Purpose and Description**
Pre-Exercise Survey	In this survey, participants indicated their most and least liked reading comprehension passages out of 4 prepared passages. They were also asked about their reading habits and disposition towards reading to find out if these affected their performance and behaviour in completing the reading comprehension tests.
Reading Comprehension Tests	Each test had 10 MCQs. Participants rated how confident they were of each answer. They were asked to do their most and least liked tests in order to find out how students' like or dislike for learning material would affect engagement level and performance.
Post-Exercise Survey	In this survey, participants self-reported their affective states, intrinsic motivation levels and mind wandering frequency. This was compared against extracted data from the various physiological sensors to determine if they matched.

Table 2: Purpose and Description of Tasks in Phase 2 of the Experiment

2.3 Data Analysis

<u>Data Pre-Processing</u>

Outliers were identified and removed through a one-class SVM with nonlinear kernel (RBF). All data was normalized to [-1,1] for standardisation.

<u>Data Labelling</u>

EEG Signals: EEG signals were labelled using participants' reaction times as they carried out the CCPT. From the reaction time, we gathered participants' EEG band power ratios at the 66th and 33rd percentile which provided us with participants' benchmarks for high, medium and low engagement. These benchmarks were later utilised to classify EEG signals during the reading comprehension tests.

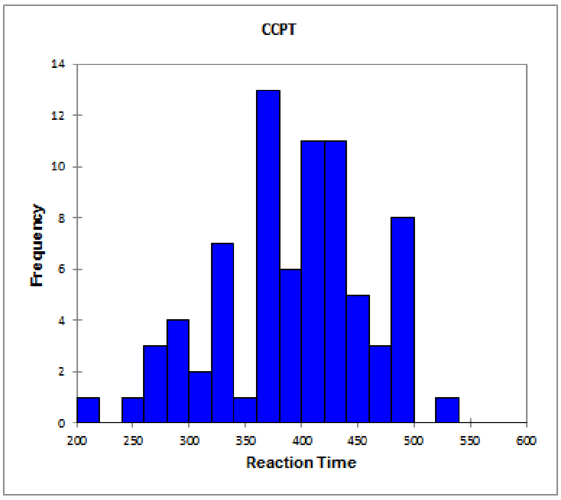


Fig. 4: Histogram showing reaction times of participant during CCPT

Band Power Ratio	Reaction Time	Engagement Level	Percentile
1/α, β/α, β/(α+θ)	<350 ms	High	$> 66^{th}$
1/α, β/α, β/(α+θ)	350-450 ms	Medium	33^{rd} - 66^{th}
1/α, β/α, β/(α+θ)	>450ms	Low	$< 33^{rd}$

Table 3: Engagement Level Labelling Scheme

Eye Gaze: Eye gaze coordinates collected when participants carried out the Z-String Reading task revealed that mindless reading, or lack of visual attention, was characterised by long fixations followed by high-speed saccades. To identify instances of mindless reading during the reading comprehension tests, consecutive eye gaze coordinates <2 cm apart were labelled as fixations, eye movements > 5 m/s were labelled as high-speed saccades, and all others were labelled as normal reading. To prove that the 2 groups of eye gaze coordinates labelled as fixations and normal reading were statistically different, a two-tailed t-test was used. Results revealed significant difference between the two groups, $t(3436) = 99.114$; $p < 0.001$. Scatter plots of eye gaze coordinates were made using widgets in Orange3 to visualise eye movements.

<u>Data Classification</u>:

Data was classified using a SVM with RBF in Orange3. EEG signals went through feature extraction using Python and MatLab. MatLab was used to plot spectrograms of EEG signals. Python was used to obtain line graphs of frontal asymmetry and CPI from Fourier transformed data. Using Kubios HRV, a Poincaré plot of RR interval was generated to study heart rate behaviour while the participant completed the reading comprehension tests.

3. Results

Participant's data was analysed. Using data from the CCPT as the training dataset, the following classification accuracy was achieved. This was later utilised for machine learning.

Experiment Section	CCPT	Most Liked Passage	Least Liked Passage
Classification Accuracy	93.3%	90.9%	91.6%
F1	93.3%	90.8%	91.7%
Precision	93.4%	90.9%	91.3%
Recall	93.3%	91.0%	91.8%

Table 4: Classification Accuracy Achieved Across Experiment

Visual Engagement: Number of instances of mindless reading were 95.1% greater (41 versus 80 instances) when the participant was reading the comprehension passage she least liked versus the one she most liked. The average duration of each instance of mindless reading also increased by 27.1%, and she took 16.7% longer a duration to finish reading the passage she least liked despite the two passages being of the same length.

Fig. 5: Scatter plot of eye gaze coordinates during Z-String Reading

Fig. 6: Scatter plot of eye gaze coordinates during least liked passage

Fig. 7: Scatter plot of eye gaze coordinates during most liked passage

Mental Engagement: Comparing the participant's CPI as she completed the reading comprehension test she most and least liked, we found that her CPI values were not statistically different, $t(172)$: -1.404; $p > 0.005$.

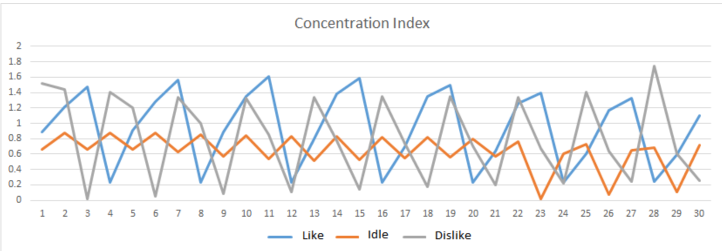


Fig. 8: Line Graph of CPI Values During Most Liked Passage, Least Liked Passage and when Idle

A similarity in frequency density of different engagement levels can be seen between the most liked and least liked passage.

Reading Passage	Most Liked	Least Liked
Frequency Density of High Engagement Level	23.4%	21.6%
Frequency Density of Medium Engagement Level	57.1%	55.7%
Frequency Density of Low Engagement Level	19.5%	22.7%

Table 5: Frequency Density of Various Engagement Levels

Mental Response: Frontal asymmetry calculations showed that the participant had higher band power in the left frontal cortex as she completed the reading comprehension test she most liked, indicating positive feelings, engagement and motivation. However, the participant had higher band power in the frontal cortex as she completed the reading comprehension test she least liked, indicating negative feelings, lack of engagement and lack of motivation.

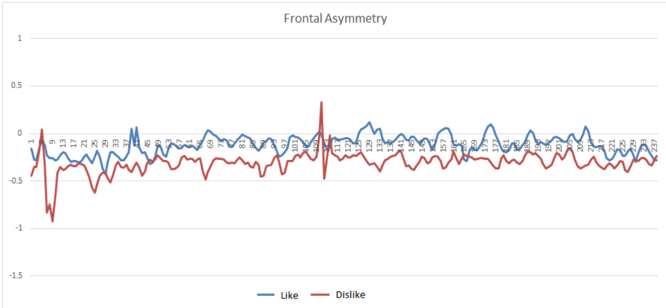
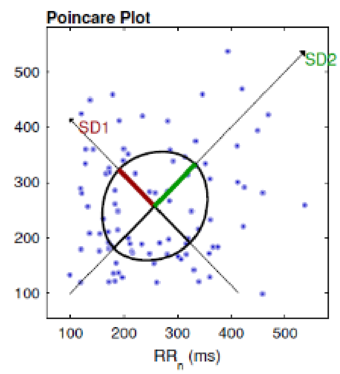


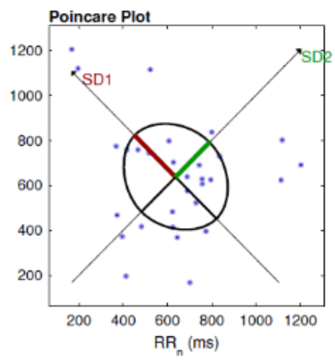
Fig. 9: Line Graph of Frontal Asymmetry During Least Liked and Most Liked Passage

Heart Rate Variability: From the Poincaré plot, it can be inferred that the participant experienced more negative emotions when completing the passage she disliked. The SD1 and SD2 values in the Poincaré plot for the disliked passage were 64.2% and 50.8% lower than that in the Poincaré plot for the liked passage respectively, indicating greater fatigue of both the parasympathetic and sympathetic nervous systems [9].



SD1: 93.6

SD2: 107.0



SD1: 261.4

SD2: 217.5

Fig. 10: (Bottom) Poincaré Plot of RR Interval During Most Liked Passage (Top) Poincaré Plot of RR Interval During Least Liked Passage

Engagement Index: Taking into consideration each modality, namely visual engagement, mental engagement, heart rate variability and mental response, a holistic engagement level score can be calculated. The lower the score, the higher the engagement level. This was done by assigning a score value to each modality and summing up the total score. For the participant in our paper, her overall engagement level score as she did the passage she most liked would be a 4, whereas that of the passage she least liked would be 6, where 5 - 6 is high, 7 - 9 is medium and 10 - 11 is low engagement level.

4. Discussion

While the participant's mental engagement as she completed her most and least liked passages were not statistically different, her mental response, heart rate variability, and visual engagement results all showed significantly different results between the two passages. True to the limitation of using EEG signals to measure engagement level as mentioned in the introduction, mental engagement alone is unable to distinguish between whether a participant is concentrating on the assigned task, or in an irrelevant task such as in mind wandering. Therefore, while the participant seemed to be equally attentive during both passages as shown by her mental engagement results, the participant was actually significantly less attentive as shown by her visual engagement results. This tells us that students' attentiveness decreases along with their preference for learning material. This eventually also affects students' efficiency in studying, seeing as the participant's increased mind wandering during the passage she least liked caused her to complete reading it in a longer amount of time. Without measuring visual and mental engagement in conjunction, such insight would not have been found. This shows that different physiological responses need to be monitored of a student for corroboration purposes to strengthen accuracy and reliability of measured engagement levels. By measuring visual engagement, educators will also be aware of just when exactly students' are losing attention, and be able to intervene accordingly.

In addition to visual engagement, mental response and heart rate variability also show similarly that there was greater engagement during the passage the participant preferred compared to the one she least liked. Mental response and heart rate variability both showed more negative feelings, lack of motivation, lack of engagement, and increased stress during the passage that she liked least compared to the passage she liked most, despite showing similar mental engagement results for both passages. This tells us that assessing affective states along with attention levels is important for a holistic understanding of students' response to learning materials. Affective states gives us insight to students' attitude towards learning, and long-term monitoring would be important if we were to aim to cultivate a positive learning attitude in students. Together, our proposed engagement index, which assesses both attitude and attention towards learning material, serves as a stronger learning performance indicator than existing ones.

Although the final calculated engagement level score for both passages were under the high engagement level range, this could be due to the fact that our engagement level score is based on a mere summation of assigned value scores for each modality rather than a weighted calculation where modalities more crucial to engagement level are assigned greater weightage to the final engagement level score.

5. Conclusion

As per our hypothesis, our proposed Engagement Index is indeed a more holistic and accurate measurement of student engagement. We found that where mental engagement alone failed to distinguish the difference in attention levels and affective states while the participant responded to passages she liked versus disliked, the other modalities included in the Engagement Index, namely visual engagement, mental response and heart rate variability, made up for it by showing significantly different results between the two passages. This confirms that a multimodal approach is required for corroboration purposes to strengthen reliability and accuracy of Engagement Level measures. If our Engagement Index were to be used over long-term periods, it can function as a strong indicator of learning performance by assessing both participant's attitude towards the learning material and their actual attention levels as they complete tasks. The fact that our Engagement Index is digital will also make it viable for use on e-Learning platforms, so long as we work on making our system calculate Engagement Level values in real-time so as to provide real-time feedback to educators online for them to monitor students' engagement during e-Learning. Other future work also includes analysing GSR data to determine its role in engagement levels. We would also like to weigh how strongly mental, visual engagement, heart rate variability and mental response contributes to engagement levels through correlation tests in order to come up with a weighted overall engagement level score, rather than a simplistic summation of each factor that our engagement level score is now.

References

[1] Learning [Def. 2]. (n.d.). In *Merriam Webster Online*, Retrieved December 16, 2016, from http://www.merriam-webster.com/dictionary/learning

[2] Hidden curriculum (2014, August 26). In S. Abbott (Ed.), The glossary of education reform. Retrieved from http://edglossary.org/hidden-curriculum\

[3] iMotions. (2016). *EEG* [Pocket Guide]. N.P.: iMotions.

[4] Dössel, O. (2009). *World Congress on Medical Physics and Biomedical Engineering* (1st ed., p. 538). [Berlin ; Heidelberg ; New York, NY]: Springer.

[5] Affect and Mental Engagement: Towards Adaptability for Intelligent Systems Maher Chaouachi, Pierre Chalfoun, Imène Jraidi and Claude Frasson, Université de Montréa

[6] S-Y Lee and C. Lee "Research on the Game for increasing intensive power using EEG signal", Department of Computer Science, The Graduate School of Industrial information, Woosong University , pp. 7-41, January 2009

[7] Y. Sung, K. Cho, and K. Um, "A Framework for Processing Brain Waves Used in a Brain-computer Interface", Journal of Information Processing Systems, vol.8, no.2, June 2012.

[8] Farnsworth, B. (2015). *Frontal Asymmetry 101 - Motivation and Emotions from EEG*. *iMotions*. Retrieved 21 November 2016, from https://imotions.com/blog/frontal-asymmetry-101-get-insights-motivation-emotions-eeg/

[9] A. Couzens (2015). *Geeking out on HRV: The advanced stuff*. Retrieved 14 June 2017, from https://www.alancouzens.com/blog/overtraining_HRV3.html/

[10] T. J. Choi, J. O. Kim, S. M. Jin, G. Yoon (2014), Determination of the concentrated state using multiple EEG channels, International Journal of Computer, Information, Systems and Control Engineering, World Academy of Science, Engineering and Technology, 8:8:1215- 1218, August.