Brain Function Analysis Using EEG Evidence: New Insights into English Paper-Based versus Computer-Based Tests

Thapanee Khemanuwong
School of Liberal Arts
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
thapanee.kh@kmiltl.ac.th

*Thanate Angsuwatanakul
College of Biomedical Engineering
Rangsit University
Pathum Thani, Thailand
thanate.a@rsu.ac.th
*Corresponding author

Ekkapon Phairot
Faculty of Education
Songkhla Rajabhat University
Songkhla, Thailand
ekkapon.ph@skru.ac.th

Keiji Iramina
Graduate School of Systems Life Sciences
School of Interdisciplinary Science and Innovation, Kyushu University
Fukuoka, Japan
iramina@inf.kyushu-u.ac.jp

Tasawan Puttasakul
College of Biomedical Engineering
Rangsit University
Pathum Thai, Thailand
tasawan.p@rsu.ac.th

Nagorn Bunyarit
College of Innovation and Industrial Management
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
nagorn.bu@kmitl.ac.th

2024 16th Biomedical Engineering International Conference (BMEiCON) | 979-8-3315-0543-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/BMEICON64021.2024.10896313

***Abstract*— In today's disruptive era, where digital systems and the internet are central, assessment methods are transitioning from paper-based to computer-based tests. As digital technology becomes more accessible, there is growing interest in determining whether these new formats provide more effective means of evaluating learning outcomes. To investigate this, we examined performance and objective measures of brain function during both computer-based and paper-based reading comprehension tests. This study focuses on brain function analysis of computer-based and paper-based tests measured by EEG signals. Five healthy students at the B1 CEFR level voluntarily participated in two experimental conditions: paper-based testing (PBT) and computer-based testing (CBT). During the tests, EEG signals were recorded and analyzed using MATLAB to identify various features of brain activity. The results indicate that participants who performed better on paper-based tests showed greater familiarity with the test format. The power spectral density of EEG recordings, along with the average frequency of alpha and beta waves, were positively correlated with test familiarity. Specifically, the correlation coefficients were as follows: CBT-difficult ($r = 0.82$, $p < 0.05$), CBT-easy ($r = 0.82$, $p < 0.05$), PBT-difficult ($r = 0.82$, $p < 0.05$), and PBT-easy ($r = 0.65$, $p < 0.05$).**

***Index Terms*— Brain Function Analysis, Computer-Based Tests (CBT), Paper-Based Tests (PBT), Electroencephalography (EEG)**

I. INTRODUCTION

The benefits of computer-based tests (CBT) include quick result turnaround, reduced paper expenses, and heightened student engagement. These advantages have driven the transition from paper-based tests (PBT) to CBT. The rapid feedback provided by CBT allows educators and administrators to adjust their methods based on student needs [1]. The benefits of CBT appear to outweigh those of PBT, suggesting that CBT may become the preferred method in many settings. This shift is particularly significant in universities, where timely and consistent feedback is crucial. However, moving from PBT to CBT raises several issues that must be addressed to avoid negatively impacting students. Concerns include the availability of necessary technology, the reliability of these systems (e.g., potential technical issues), students' familiarity with the new formats, discrepancies in assessment outcomes between the two types, and the potential effects on student performance. Addressing these issues early in the transition is essential. Additionally, the different impacts of these test formats on various student groups may raise fairness concerns, and it is important to ensure that increased technological use does not exacerbate economic disparities among students.

Following the global COVID-19 pandemic, educators are adapting their teaching methods across all levels. Assessing the impact of digital technologies on student performance and success is now more critical than ever. In language learning, CBT has emerged as one of the quickest and most accessible ways for students to receive immediate feedback [2]. In some developing countries where English is taught as a foreign language, CBT is less prevalent, often due to a lack of familiarity with the necessary technology and procedures. Consequently, students in these regions might feel that their CBT results do not accurately reflect their language skills, primarily due to difficulties with the test format. This can pose challenges for teachers. Therefore, it is vital that educational institutions ensure that the adoption of CBT enhances efficiency without creating additional obstacles for students and teachers. Preparing students for CBT in advance is crucial to build their confidence and guarantee their success. Instructors and test developers should recognize the importance of this preparation and work to equip students with the required skills. Given the increasing popularity of CBT in Thai universities, there is a valuable opportunity to evaluate the technology and address any potential challenges, ensuring a smoother and more effective implementation of CBT on a broader scale.

II. Literature Review

A. CBT and PBT Tests

Research comparing computerized tests (CBT) and paper-and-pencil tests (PBT) [3, 4] has revealed notable differences between the formats. For example, students often prefer CBT due to its dynamic visuals, interactivity, flexibility, and real-time score reporting, which enhance the assessment experience compared to PBT. Experiments by Sawaki [5] and Butcher, Perry, and Atlis [6] found that students completed CBT faster, with no significant difference in scores. Isleem's study on technology education teachers showed that computer competence predicts positive attitudes towards CBT [7], and Albirini's research highlighted the role of computer competence in shaping teachers' ICT attitudes [8]. Other studies indicate that computer ownership improves teachers' technology competence and attitudes [9]. Differences between PBT and CBT can vary based on factors such as the test measure and technology used [10]. For example, screen size and resolution can affect CBT performance, though more experience with CBT may lessen these effects [11]. Yunus found a positive link between computer competency and educational technology attitudes [12]. The "test mode effect" [13], influenced by factors like computer familiarity and attitudes, can cause variability in test outcomes. For instance, Hosseini, Abidin, and Baghdarnia found higher scores on PBT compared to CBT [14], while another study reported the opposite for a dental hygiene course [15]. Some research shows no significant differences between formats [16]. Continued research is essential to refine practices for both CBT and PBT to ensure accurate assessments.

B. EEG as Measurement for Brain Activity of the Test Takers

Electroencephalography (EEG) is a non-invasive neuroimaging technique [17] that measures electrical activity in the brain via electrodes on the scalp. Its high temporal resolution allows for precise tracking of dynamic brain activity, making it useful for studying mental workload, as reflected in changes in brain wave patterns like alpha and beta waves [18].

Applying EEG practically to assess mental workload presents challenges. Current literature often records only a limited range of variables, complicating comparisons across studies. Additionally, many studies analyze data at the group level rather than individually, making personalized assessments difficult. Experimental methods can also be problematic, with confounding factors like body movements and visual processing affecting both EEG and related measures.

This study aims to address these issues by assessing EEG's effectiveness in measuring mental workload during reading comprehension exams, comparing paper-based tests (PBT) and computer-based tests (CBT) of varying difficulty. EEG has shown potential in revealing neurocognitive processes involved in reading [19]. By exploring and comparing brain function during CBT and PBT, we hope to gain insights into how different testing formats affect cognitive load and performance, ultimately aiding in optimizing test design and assessment practices.

III. Methodology

In this study, brain activity related to paper-based tests (PBT) and computer-based tests (CBT) was investigated using electroencephalography (EEG) data collected from English language students during reading comprehension assessments. The objective of this pilot study was to identify patterns of brain activity linked to these two types of assessments. The null hypothesis posited that there is no significant correlation among the seven features measured by EEG.

A. Participants

This pilot study involved five healthy young adult volunteers from Rangsit University, comprising two males and three females, with an average age of 22.2 years. Participants were chosen through purposive sampling based on specific criteria: they needed to be English learners at the B1 level of the CEFR, have normal color vision, and be free from memory disorders and any past neurological or psychological conditions.

B. Instrument

This study employed the Thai Reading Assessment and Decoding System (Thai-READS) for data collection in both computer-based test (CBT) and paper-based test (PBT) formats. The Thai-READS evaluates three reading comprehension skills: literal, re-organization, and inferential, which are effective for assessing undergraduate reading performance [20]. We assessed the reliability of each test format using the KR20 coefficient, finding high reliability for the PBT (KR20 = 0.91) and moderate reliability for the CBT (KR20 = 0.64).

Both CBT and PBT instruments contained 30 items each, divided into Difficult and Easy categories (15 items per category). The items were matched in length with varied text topics to introduce variability in reading effort. According to Nystrand et al., the time to complete a multiple-choice exam can differ among individuals [21], and Khemanuwong et al. recommend allowing 60 to 75 seconds per item [22]. Time spent reviewing questions before submitting was also considered.

Participants took two equivalent tests sequentially: one CBT in a computer lab and one PBT as outlined in Table I. They were instructed on how to complete the computerized questions and informed of their right to withdraw from the study, adhering to ethical guidelines.

TABLE I. Components of CBT and PBT Tests

Number of Comprehension Questions	Educational Levels	Weight age	Level of Difficulty	Allocated Time
15	Undergraduate	50%	Easy	60 seconds per test item
15	Undergraduate	50%	Difficult	

C. Procedures

Before the experiment, we explained the study's scope and procedures to participants, assured them of no health risks, and emphasized voluntary participation. Informed consent was obtained from each participant, and they received detailed written instructions on the experimental process. Following Straker et al.'s protocol [23], participants were seated in adjustable chairs about 50 cm from the computer screen. They rested for one minute, focusing on a mark on a black screen, before starting the computerized test (CBT), completing 30 items in 30 minutes and pressing a button to move to the next item. After a 15-minute rest, participants began the paper-based test (PBT).

For the PBT, participants started the test by pressing a button to transition from a blank page to a text page, advancing through the test by pressing a button after reading each page. They focused on a mark (X) on a black screen to prepare for the PBT, then completed 30 items in 30 minutes. EEG signals were recorded before, during, and after both tests. The experimental sequence is illustrated in Figure 1, and the study's concept block diagram is shown in Figure 2. Button presses sent trigger signals to synchronize EEG recordings with each exam question.

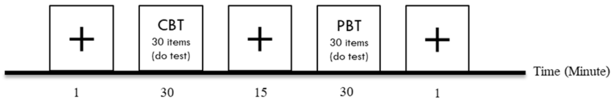


Fig. 1. Experiment process of CBT and PBT.

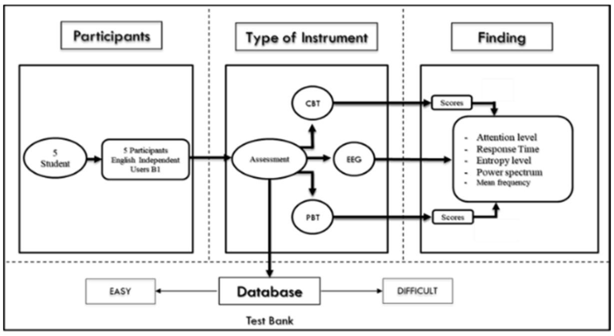


Fig. 2. Block Diagram of Experiment Procedure.

D. EEG Data Analysis

EEG data were recorded using the MindWave Mobile 2 (NeuroSky Inc.; San Jose, USA), an affordable and user-friendly biosensor for measuring brain electrical activity. This device, attached to the earlobes and forehead, amplifies and filters microvolt-level electrical signals from the scalp, mainly capturing signals from the prefrontal cortex (see Figure 3).

Data were acquired with Unity software (Unity Software Inc.; San Francisco, USA) and exported to Matlab (R2017b; The MathWorks Inc.; Natick, USA) for analysis. Features of EEG signals from both CBT and PBT sessions were computed, including mean alpha and beta power, mean alpha and beta frequency, 'attention level' (NeuroSky's proprietary algorithm), and sample entropy (see Figure 4). Response times for each question in both test formats were also recorded.

The extracted features were analyzed across four conditions: E1 (easy-correct), E0 (easy-incorrect), D1 (difficult-correct), and D0 (difficult-incorrect), with the first letter indicating difficulty and the second correctness. Each feature was standardized to have zero mean and unit variance for comparability.

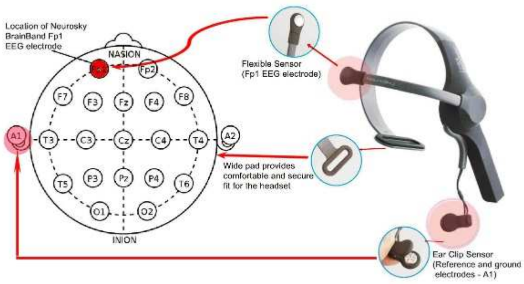
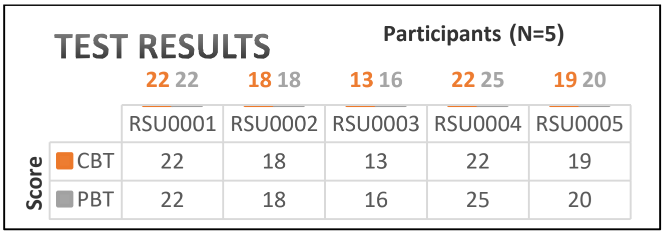


Fig. 3. Neurosky Mindwave Mobile [24].

IV. Results

As illustrated in Figure 4, the mean score for the computer-based test (CBT) was 18.80 ± 3.31 standard deviation (SD), while the mean score for the paper-based test (PBT) was 20.20 ± 3.12 SD. This comparison shows that the mean score for PBT was slightly higher (M = 20.20, SD = 3.12) than that for CBT. Despite this observed difference, it is important to note that the difference in mean scores between CBT and PBT was not statistically significant. This implies that while students performed somewhat better on the PBT, the variation in scores between the two testing formats may not be large enough to conclude that one format is definitively superior to the other.



	RSU0001	RSU0002	RSU0003	RSU0004	RSU0005
CBT	22	18	13	22	19
PBT	22	18	16	25	20

Fig. 4. CBT and PBT Descriptive Statistics Results.

A. Response Time

The response times of participants were recorded during both CBT and PBT assessments, and this data is analyzed and presented in Figure 5. The results indicate that, on average, participants took more time to complete questions on the CBT compared to the PBT. Furthermore, the analysis reveals that participants generally spent more time on difficult items than on easier ones. This pattern is consistent across both testing formats, suggesting that the complexity of the questions influences the time required for completion. Additionally, the data show that participants tended to spend less time on questions they answered correctly compared to those they answered incorrectly. This observation may imply that correct answers are often reached more quickly due to higher confidence or better understanding of the material, whereas incorrect answers may involve more deliberation or uncertainty.

These findings suggest that the CBT format might involve more cognitive processing time or require additional interaction time compared to the PBT format. The increased time for difficult questions and incorrect responses reflects the additional cognitive effort needed for more challenging or uncertain tasks.

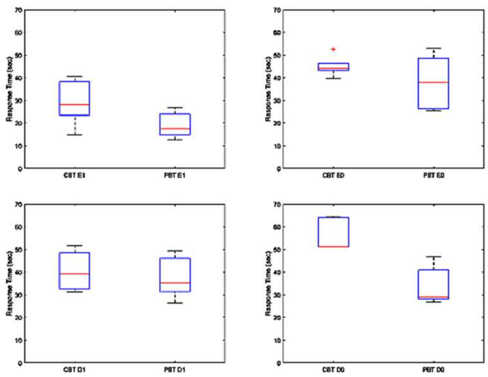


Fig. 5. Participants' Average Time Spent on CBT and PBT.

B. Entropy Level

The entropy level of EEG signals, which reflects the overall complexity of cognitive processing, was analyzed to understand its association with different test conditions [25]. As depicted in Figure 6, the results indicate that entropy levels were marginally higher for difficult questions compared to easier ones. This suggests that more complex cognitive processing is involved when tackling challenging questions, leading to greater signal complexity.

Interestingly, Figure 6 shows that the entropy level was also slightly higher during paper-based tests (PBT) compared to computer-based tests (CBT). This finding implies that the cognitive demands and complexity of processing information might be somewhat greater in the PBT format.

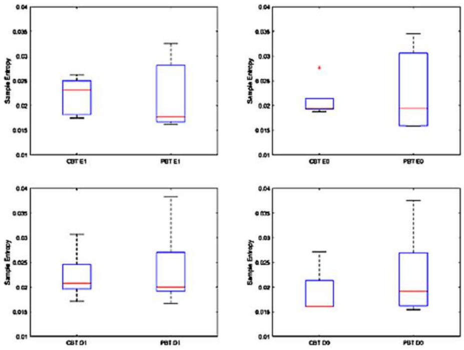


Fig. 6. Comparison of Entropy Level.

C. Attention Level

Using EEG to capture dynamic brain activity, the data presented in Figure 7 indicate that the paper-based test (PBT) appears to evoke higher levels of attention compared to the computer-based test (CBT). This observation suggests that participants may engage more attentively or exhibit greater neural activation during the PBT. However, test takers demonstrated a higher level of attention in CBT for D0. The findings suggest that test takers might maintain their concentration on CBT, even when faced with challenging questions or incorrect answers.

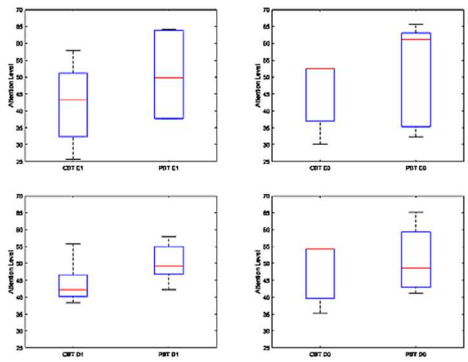


Fig. 7. Attention Level.

D. Feature Correlation Analysis

This study uses scatter plots to explore the relationships between pairs of features. A positive correlation occurs when both variables increase together (e.g., as y increases, x also increases), whereas a negative correlation is seen when one variable increases while the other decreases (e.g., as y increases, x decreases). Figure 8 displays a feature correlation matrix from CBT and PBT data, analyzing seven features, including EEG signals: attention level, response time, sample entropy, mean beta power, mean alpha power, mean beta frequency, and mean alpha frequency. Each feature was standardized to ensure comparability, with a mean of zero and a variance of one.

The matrix shows pairwise comparisons with scatter plots and histograms displaying feature value distributions. Analyzing these plots and histograms reveals the statistical significance of correlations. For instance, a cluster for response time and sample entropy indicates a significant relationship, while the correlation between response time and other features appears more linear. Additionally, linear correlations were found between mean beta power and mean alpha power, and between mean beta frequency and mean alpha power. These findings provide insights into how cognitive and EEG measures interact, highlighting key relationships and guiding further research to better understand feature interactions during CBT and PBT assessments.

E. Group-wise Evaluation

To investigate brain activity related to each experimental condition, feature correlation analysis was performed for each group, as shown in Figure 8. The plots display average data from all participants under four conditions: CBT-easy, CBT-difficult, PBT-easy, and PBT-difficult, categorized by test modality and question difficulty. Each condition included data from five participants who answered 15 questions per category, resulting in a sample size of (N = 75) for each condition. As shown in Figure 8, the results reveal similar patterns in dataset feature distribution across all groups, indicating comparable mental workload characteristics between the two assessment types. Pearson's correlation coefficient was used to quantify these correlations and analyzed with SPSS (Version 22; Inc.; New York, USA). The findings are summarized as follows:

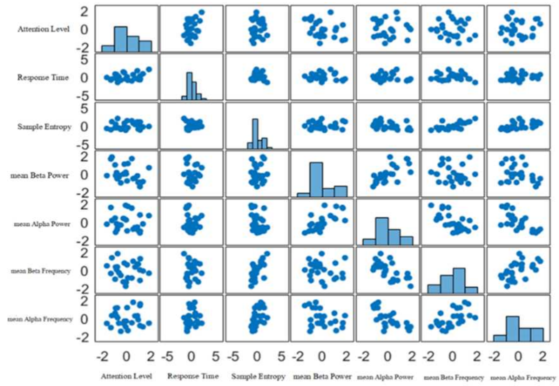


Fig. 8. Feature Correlation Analysis from Whole Study Group.

1. CBT-difficult Condition: Significant positive correlations ($p < 0.05$) were found between: a) Attention and: sample entropy, beta power, alpha power; b) Sample entropy and: beta power, alpha power, beta frequency; c) Beta power and: alpha power, beta frequency; d) Alpha power and: beta frequency, alpha frequency

2. CBT-easy Condition: Significant positive correlations ($p < 0.05$) were observed between: a) Attention and: sample entropy, beta power, alpha power, beta frequency; b) Sample entropy and: beta power, alpha power, beta frequency; c) Beta power and: alpha power, beta frequency; d) Alpha power and: beta frequency, alpha frequency

3. PBT-difficult Condition: Significant positive correlations ($p < 0.05$) were found between: a) Attention and: sample entropy, beta frequency; b) Sample entropy and: beta power, alpha power, beta frequency, alpha frequency; c) Beta power and: alpha power, beta frequency; d) Alpha power and: beta frequency, alpha frequency; e) Beta frequency and: alpha frequency

4. PBT-easy Condition: Significant positive correlations ($p < 0.05$) were observed between: a) Attention and: sample entropy; b) Sample entropy and: alpha power, beta frequency, alpha frequency; c) Beta power and: alpha power, alpha frequency; d) Alpha power and: beta frequency

These results highlight the relationships between various EEG features and their association with mental workload during different test conditions.

V. Discussion and Conclusion

This study provides initial evidence that EEG is a useful tool for assessing brain activity during Paper-Based Tests (PBT) and Computer-Based Tests (CBT). Unlike other neuroimaging methods like fMRI or MEG, EEG is non-invasive, cost-effective, and allows for natural movement, thus minimizing disruptions to behavior and performance [26, 27]. The analysis shows that EEG can distinguish brain functions across different testing modalities and difficulty levels. For instance, minimal variations in spectral power during the PBT-easy condition indicate a lower cognitive workload compared to more challenging conditions, highlighting EEG's utility in assessing cognitive load. Pairwise correlation analysis (see Figure 8) reveals that the mean EEG power spectrum is affected by the test format, rejecting the null hypothesis and showing significant positive correlations among the seven EEG features. Specifically, the power spectral density and average frequencies of alpha and beta waves were positively correlated with test modality, with correlations as follows: CBT-difficult ($r = 0.82$, $p < 0.05$), CBT-easy ($r = 0.82, p < 0.05$), PBT-difficult ($r = 0.82$, $p < 0.05$), and PBT-easy ($r = 0.65$, $p < 0.05$), resulting in an overall $r^2 = 0.59$, $p < 0.05$. These results suggest that EEG features could help identify test modalities and predict their impact on performance, though a larger sample is needed to confirm these findings and develop reliable predictive models.

As this study is a preliminary investigation, it cannot definitively determine the comparative effectiveness of PBT and CBT for measuring learning outcomes. The study's scope is too narrow to fully explore all factors affecting these assessments, especially in second language reading comprehension. Further research is needed to understand how PBT and CBT function in various contexts, as some assessments may be better suited to one modality over another, while others might require different approaches. Incorporating EEG data into test performance analysis can deepen our understanding of cognitive processes related to different testing modalities. Although this pilot study does not provide definitive evidence, it establishes a foundation for future research that could use EEG to enhance and tailor educational assessments based on brain function markers.

References

[1] L. M. Abrams, J. H. McMillan, and A. P. Wetzel, “Implementing benchmark testing for formative purposes: Teacher voices about what works,” Educ. Assess. Eval. Accountab., vol. 27, no. 4, 2015, pp. 347-375.

[2] G. J. Ockey, “Developments and challenges in the use of computer - based testing for assessing second language ability,” Mod. Lang. J., vol. 93, 2009, pp. 836-847.

[3] H. Khoshsima, M. Hosseini, and S. M. H. Toroujeni, “Cross-mode comparability of computer-Based testing (CBT) versus paper-pencil based testing (PPT): An investigation of testing administration mode among Iranian intermediate EFL learners,” Engl. Lang. Teach., vol. 10, no. 2, 2017, pp. 23-32.

[4] S. M. Bodmann, and D. H. Robinson, “Speed and performance differences among computer-based and paper-pencil tests,” J. Educ. Comput. Res., vol. 31, no. 1, 2004, pp. 51-60.

[5] Y. Sawaki, “Comparability of conventional and computerized tests of reading in a second language,” Lang. Learn. Technol., vol. 5, no. 2, 2001, pp. 38-59.

[6] J. N. Butcher, J. N. Perry, and M. M. Atlis, “Validity and utility of computer-based test interpretation,” Psychol. Assess., vol. 12, no. 1, 2000, pp. 6-15.

[7] M. I. Isleem, “Relationships of selected factors and the level of computer use for instructional purposes by technology education teachers in Ohio public schools: A statewide survey,” Ph.D. dissertation, Ohio State Univ., Columbus, OH, USA, 2003.

[8] A. Albirini, “Teachers' attitudes toward information and communication technologies: The case of Syrian EFL teachers,” *Comput. Educ.*, vol. 47, no. 4, 2006, pp. 373-398.

[9] T. Deutsch, K. Herrmann, T. Frese, and H. Sandholzer, “Implementing computer-based assessment–A web-based mock examination changes attitudes,” *Comput. Educ.*, vol. 58, no. 4, 2012, pp. 1068-1075.

[10] G. F. Shuster and M. Pearl, “Computer competency: A 7-year study to identify gaps in student computer skills,” *Int. Educ. Stud.*, vol. 4, no. 4, 2011, pp. 137-148.

[12] M. M. Yunus, “Malaysian ESL teachers' use of ICT in their classrooms: expectations and realities,” ReCALL, vol. 19, no. 1, 2007, pp. 79-95.

[13] J. M. Noyes, and K. J. Garland, “Computer-vs. paper-based tasks: Are they equivalent?,” Ergonomics, vol. 51, no. 9, 2008, pp. 1352-1375.

[14] M. Hosseini, M. J. Z. Abidin, and M. Baghdarnia, “Comparability of test results of computer-based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran,” Procedia-Soc. Behav. Sci., vol. 98, 2014, pp. 659-667.

[15] H. Jeong, “A comparative study of scores on computer-based tests and paper-based tests,” Behav. Inf. Technol., vol. 33, no. 4, 2014, pp. 410-422.

[16] B. Anakwe, “Comparison of student performance in paper-based versus computer-based testing,” J. Educ. Bus., vol. 84, no. 1, 2008, pp. 13-17.

[17] F. Pittau, et al., “The role of functional neuroimaging in pre-surgical epilepsy evaluation,” Front. Neurol., vol. 5, p. 31, 2014, pp.31-45.

[18] Jr. C. E. Stevens, and D. L. Zabelina, “Creativity comes in waves: An EEG-focused exploration of the creative brain,” Curr. Opin. Behav. Sci., vol. 27, no. 1, 2019, pp. 154-162.

[19] S. Weiss and H. M. Mueller, “The contribution of EEG coherence to the investigation of language,” Brain Lang., vol. 85, no. 2, 2003, pp. 325-343.

[20] K. S. Hui, K. M. Saeed, and T. Khemanuwong, “Reading Comprehension Ability of Future Engineers in Thailand,” *MEXTESOL J.*, vol. 44, no. 4, 2020, pp. 1-21.

[21] M. Nystrand, L. L. Wu, A. Gamoran, S. Zeiser, and D. A. Long, “Questions in time: Investigating the structure and dynamics of unfolding classroom discourse,” Discourse Process., vol. 35, no. 2, 2003, pp. 135-198.

[22] T. Khemanuwong, A. R. Mohamed, and S. A. M. M. Ismail, “Developing a Thai READS encoder to gauge EFL reading proficiency of Thai undergraduate students,” J. Teach. Learn. Engl. Multicult. Contexts (TLEMC), vol. 2, no. 1, 2018, pp. 23-34.

[23] L. Straker, C. Pollock, R. Burgess-Limerick, R. Skoss, and J. Coleman, “The impact of computer display height and desk design on muscle activity during information technology work by young adults,” J. Electromyogr. Kinesiol., vol. 18, no. 4, 2008, pp. 606-617.

[24] M. Alnemari, “Integration of a low-cost EEG headset with the Internet of thing framework,” Ph.D. dissertation, Univ. California, Irvine, CA, USA, 2017.

[25] T. Angsuwatanakul, J. O'Reilly, K. Ounjai, B. Kaewkamnerdpong, and K. Iramina, “Multiscale Entropy as a New Feature for EEG and fNIRS Analysis,” Entropy, vol. 22, no. 2, 2020, p. 189.

[26] S. J. Ruiz-Gómez et al., “Automated multiclass classification of spontaneous EEG activity in Alzheimer’s disease and mild cognitive impairment,” Entropy, vol. 20, no. 1, 2018, p. 35.

[27] T. Angsuwatanakul, K. Iramina, and B. Kaewkamnerdpong, “Multi-scale sample entropy as a feature for working memory study,” in 7th 2014 Biomed. Eng. Int. Conf., Fukuoka, Japan, Nov. 2014, pp. 1-5.