Exploratory investigation of electrodermal activity in learning from a large language model versus from curated texts

Kenneth Y T Lim[a], Yue Heng Wong[b], Duc Nam Tran[b], Edrik K X Lee[b], Minh Tuan Nguyen Thien[b], Minh Anh Nguyen Duc[b], Alan J H Tan[b]

[a] *National Institute of Education, Singapore* kenneth.lim@nie.edu.sg

[b] *independent scholar, Singapore*

1. Introduction

Understanding how different types of instructional content impact student learning and arousal is critical in this era of advanced artificial intelligence (AI). Rochelle et al. [1] have highlight-ed the potential of LLMs to provide tailored learning experiences for individual students, allowing students to learn at their own pace.

However, most studies study the effect of LLMs on learning over long periods of time (e.g. over months at a time). As such, there exists a lack of research into the *in situ* effect of LLMs on learning performance. Studying the effect of LLMs *in situ* allows more extensive data to be collected from participants, opening avenues for greater study into the intricacies of learning aided by LLMs.

In the study, we investigated the effect of LLMs on learning performance via the analysis of Electrodermal Activity (EDA) over the short term. EDA, which measures the electrical conductance of the skin in response to sweat secretion, is linked to the sympathetic nervous system and offers valuable data on mental and physiological states [2].

In addition, EDA is tightly intertwined with one's emotions [2], which are involved every aspect of the learning process [3]. Positive emotions, such as joy, encourage the learning process, while negative emotions, like fear, inhibit the learning process [4]. As such, investigation of EDA during learning provides real-time insights into the cognitive and emotional stresses of learning *in situ* for a deeper understanding of the learning process when using LLMs.

2. Methodology

In August 2024, a total of seven participants volunteered in the experiment. The participants were 16- to 17-year-old males, recruited by convenience through the peer network of the authorial team.

Various topics drawn from the disciplinary domains in both the humanities and STEM fields chosen as the topic to be learned via the two learning methods as this covered a large spectrum of topics and expertise. This reduces bias which could occur if participants had prior knowledge about the topic.

A human-written curated text about the topic was prepared prior to the experiment. It was prepared in the tone of an academic article.

For the LLM, we used the official ChatGPT-3.5 interface. We used the CO-STAR framework [5] when designing the prompt such that the response given by the LLM would be helpful and relevant to the topic. The curated text was included in the prompt as a guide.

DIY EDA sensors, first introduced in Lim et al. [6], were used for the collection of EDA data from the participants. The EDA sensors were designed and assembled following work by Zangróniz et al. [7] and verified against research-grade equipment by Lim et al. [6].

All data was anonymized, and no personally identifiable information was collected.

The quiz varied according to the topic, and was administered online via a form.

Before the start of the experiment, the sensors were attached the fingers of the participant's non-dominant hand. The sensors would only be removed at the end of the experiment.

The experiment included baseline measurements at the start of the experiment, after the quiz, and at the end of the experiment. This was to allow for normalization of the EDA data between individuals.

The experiment included two 10-minute learning windows for participants to learn about the topic via one of the two learning methods (LLM / curated text). For curated text, participants used a laptop to view an online copy of the file. For LLM, participants were given a laptop with access to the LLM with the prompt and were encouraged to ask the LLM further questions in addition to the content generated by the prompt. Participants had the sensors measuring their EDA throughout to allow for the analysis of changes in EDA during learning.

All participants had to learn from both methods. The order of which the participants learned from the different learning methods were randomized among participants.

After the first learning window, the quiz was given to participants via an online form, regardless of the learning method they used.

The EDA data collected was processed in Python. The data was first resampled to 25Hz. We then performed outlier detection via the z-score method, removing outliers with an absolute z-score greater than 3. The data was then normalised and filtered using a low-pass filter (1.5Hz Butterworth, 8th Order) to remove artifacts. We per-formed further artifact detection and removal using the LSTM-CNN model provided by Llanes-Jurado et al. [8].

Afterwards, windowing of the data was performed, with a window length of 2-3 minutes corresponding to baseline length. This is done to ensure the changes in the indicators of EDA studied are sustained during the learning phase.

Within each window, five indicators of EDA were calculated: NSSCR per minute, mean SCR, mean SCL, TVSymp, and EDASymp.

NSSCR per minute was calculated through the peak detection algorithm introduced by Taylor et al. [9]. For mean SCR and SCL, we used the convex optimization approach (cvxEDA) by Greco et al. [10] to decompose the EDA data into tonic and phasic components, corresponding to SCL and SCR respectively. TVSymp was calculated according to processes by Posada-Quintero et al. [11]. Finally, for

Exploratory investigation of electrodermal activity in learning from a large language model versus from curated texts

Kenneth Y T Lim[a], **Yue Heng Wong**[b], **Duc Nam Tran**[b], **Edrik K X Lee**[b], **Minh Tuan Nguyen Thien**[b], **Minh Anh Nguyen Duc**[b], **Alan J H Tan**[b]

[a] *National Institute of Education, Singapore* kenneth.lim@nie.edu.sg

[b] *independent scholar, Singapore*

EDASymp, we downsampled the EDA data to 2 Hz and implemented the methods in Posada-Quintero et al. [12] to obtain the value of the indicator. Indicator values from the LLM and Text learning phases were then normalized by subtracting the value of the indicator from the learning session with that of the baseline. A clustered Wilcoxon signed-rank test was then performed on each indicator. The test was chosen to account for the effects of clustering due to segmentation.

3. Results and Discussion

Clustered Wilcoxon signed-rank tests showed that individuals using LLM did not elicit any statistically significant ($p < 0.1$) increase in any of the five indicators in comparison to individuals using Text. For the test (post-learning quiz results), on average, individuals performed better (Median = 7.0) for those who used LLM, compared to those who used Text (Median = 6.0). A Wilcoxon signed-rank test showed that this improvement is statistically insignificant ($W = 3.0$, $z = 0.0$, p-value = 0.5).

Through this study, we have analysed the differences in EDA between LLM and Text in learning, aiming to find differences in learner's cognitive and emotional stresses.

According to a study in 2025 by Gerlich on AI reliance and critical thinking, as people rely more on AI technology, they are less likely to engage in critical thinking and could resort to AI for cognitively demanding tasks. This phenomenon, termed "cognitive offloading," is detrimental to learners' efficacy [13]. This is congruent with the results indicated negligible differences across EDA features and test results, suggesting that participating in pedagogical activities which involve gleaning information from Generative AI may not significantly enhance learning.

4. Concluding remarks

This paper reports a recent iteration of investigation in the science of learning, arising from a trajectory of work by the authors dating from 2021 at the intersection of neuroergonomics and data science.

The work applied a frame of making and citizen science to the design of learning environments in which students sought to understand their own physiological responses as they participate in activities of learning in contexts authentic to them-selves, as opposed to lab-based studies.

We acknowledge that the preceding analyses of the results from this small-scale pilot study do not suggest there to be significant differences between learning from LLMs as opposed to learning from more traditional pedagogical means such as from texts curated by domain experts or peers.

Since ChatGPT became widely available to the general public in November 2022, it – and similar LLMs – have polarized opinions regarding, inter alia, their use in contexts of teaching and learning. Curriculum developers, school administrators and policy makers have had two years to develop, iterate and defend their respective stances, and a number of frameworks – not least being the UNESCO frameworks for both teachers and students launched in the summer of 2024 – have served to structure these sometimes heated conversations. The UNESCO frameworks in particular call for more moderate and nuanced approaches to the application of Generative AI and LLMs to teaching and learning, with their distinguishing hallmark being student-centricity. Notwithstanding its inherent limitations which we readily acknowledge, we see the present study – with its ambivalent suggestions as to the efficacies of LLMs in teaching and learning – as a small part of a nascent but rapidly growing body of literature which has been emerging over the past two years to inform policy and to suggest ways forward as school leaders and teachers seek to navigate this evolving landscape. We also acknowledge the limited generalizability of this small-scale study. The work described in this paper was conducted under the relevant IRB of the Nanyang Technological University, Singapore. The authors plan to continue extending this line of study, resources permitting.

References

[1] Rochelle, S., & Sushith. (2024). Exploring the AI era: A comparative analysis of AI-driven education and traditional teaching methods. International Journal For Multi-disciplinary Research, 6(4). https://doi.org/10.36948/ijfmr.2024.v06i04.24635

[2] Rahma, O. N., Putra, A. P., Rahmatillah, A., Putri, Y. S. K. A., Fajriaty, N. D., Ain, K., & Chai, R. (2022). Electrodermal Activity for Measuring Cognitive and Emotional Stress Level. Journal of medical signals and sensors, 12(2), 155–162. https://doi.org/10.4103/jmss.JMSS_78_20

[3] Schutz, Paul & Lanehart, S.L.(Eds.). (2002). Emotions in education [special issue]. Educational Psychologist. 37. 67-134.

[4] Osika, A., MacMahon, S., Lodge, J. M., & Caroll, A. (2024, October 1). Emotions and learning: What role do emotions play in how and why students learn?. THE Campus Learn, Share, Connect. https://www.timeshighereducation.com/campus/emotions-and-learning-what-role-do-emotions-play-how-and-why-students-learn

[5] Prompt engineering playbook. (2023, August 30). https://www.developer.tech.gov.sg/products/collections/data-science-and-artificial-intelligence/playbooks/prompt-engineering-

Exploratory investigation of electrodermal activity in learning from a large language model versus from curated texts

Kenneth Y T Lim[a], Yue Heng Wong[b], Duc Nam Tran[b], Edrik K X Lee[b], Minh Tuan Nguyen Thien[b],Minh Anh Nguyen Duc[b], Alan J H Tan[b]

[a] *National Institute of Education, Singapore* kenneth.lim@nie.edu.sg

[b] *independent scholar, Singapore*

playbook-beta-v3.pdf
[6] Lim, K. Y. T., Nguyen Thien, M. T., Nguyen Duc, M. A., & Posada-Quintero, H. F. (2024). Application of DIY Electrodermal Activity Wristband in Detecting Stress and Affective Responses of Students. Bioengineering (Basel, Switzerland), 11(3), 291. https://doi.org/10.3390/bioengineering11030291
[7] Zangróniz, R., Martínez-Rodrigo, A., Pastor, J. M., López, M. T., & Fernández-Caballero, A. (2017). Electrodermal Activity Sensor for Classification of Calm/Distress Condition. Sensors (Basel, Switzerland), 17(10), 2324. https://doi.org/10.3390/s17102324
[8] Llanes-Jurado, J., Carrasco-Ribelles, L. A., Alcañiz, M., Soria-Olivas, E., & Marín-Morales, J. (2023). Automatic artifact recognition and correction for electrodermal ac-tivity based on LSTM-CNN Models. Expert Systems with Applications, 230, 120581. https://doi.org/10.1016/j.eswa.2023.120581
[9] Taylor, S., Jaques, N., Weixuan Chen, Fedor, S., Sano, A., & Picard, R. (2015). Auto-matic identification of artifacts in Electrodermal Activity Data. 2015 37th Annual In-ternational Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 1934–1937. https://doi.org/10.1109/embc.2015.7318762
[10] Greco, A., Valenza, G., Lanata, A., Scilingo, E., & Citi, L. (2016). Cvxeda: A convex optimization approach to electrodermal activity processing. IEEE Transactions on Biomedical Engineering, 1–1. https://doi.org/10.1109/TBME.2015.2474131Author, F.: Article title. Journal 2(5), 99–110 (2016)
[11] Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, Á. D., & Chon, K. H. (2016). Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity. American journal of physiology. Regulatory, inte-grative and comparative physiology, 311(3), R582–R591. https://doi.org/10.1152/ajpregu.00180.2016
[12] Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., & Chon, K. H. (2016). Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment. Annals of biomedical engineering, 44(10), 3124–3135. https://doi.org/10.1007/s10439-016-1606-6
[13] Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Fu-ture of Critical Thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006