Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2728–2738
https://doi.org/10.54941/ahfe1008497

Evaluation of ChatGPT-Generated vs. Self-Created Study Reviewers: A Case Study Comparing Student Test Scores, Efficiency and Cognitive Workload

Aliyah Maxine P. Mallari, Yngrid Bless Z. Cadena, Aleeza Fiona R. Montinola, and Lorelie C. Grepo

Department of Industrial Engineering and Operations Research, University of the Philippines Diliman, Philippines

ABSTRACT

This study compared ChatGPT-generated and self-created reviewers in terms of engineering students' test scores, time efficiency, and perceived cognitive workload. Using a within-subjects mixed-methods design, 30 engineering students participated. Each subject experienced using both a ChatGPT-generated reviewer and a self-created reviewer. Quantitative data were analyzed using descriptive statistics, ANCOVA, and correlation analysis, while qualitative responses were examined thematically. Results showed that ChatGPT-generated reviewers were associated with lower perceived cognitive workload and shorter average study time, while self-created reviewers produced slightly higher average test scores. ANCOVA results showed that the reviewer method significantly affected test scores, NASA-TLX workload, and time efficiency, even after controlling for covariates such as AI usage frequency, note-taking frequency, reviewer-making habits, mental exhaustion, and reviewer satisfaction. Note-taking frequency was also a significant predictor of test scores, with students who reported taking notes more frequently generally achieving higher test scores regardless of the reviewer method used. The reviewer-making habit was also a significant predictor of NASA-TLX scores. Participants who regularly create reviewers as part of their study routine reported higher perceived workload across both methods, showing differences in participants' study habits. For time efficiency, none of the covariates were significant. Qualitative findings supported these results, showing that ChatGPT was preferred for speed and reduced effort, while self-created reviewers were valued for deeper understanding and retention. Recommendations for future studies include increasing the sample size, involving a wider population, and exploring a hybrid study approach.

Keywords: Student learning, Reviewer method, Generative AI, Workload, Efficiency

INTRODUCTION

The proliferation of Artificial Intelligence (AI) such as ChatGPT, Claude, and Google Gemini, has significantly transformed the educational landscape. Traditionally, the creation of study guides or reviewers represents a critical phase of learning, characterized by active engagement and repetition that fosters long-term familiarity with the subject matter (Zhou et al., 2024).

Received August 25, 2026; Revised September 26, 2026; Accepted October 7, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Conversely, the capacity of AI to generate comprehensive review materials within seconds offers substantial time-saving benefits, particularly in high-pressure scenarios where examinations are imminent or academic workloads have become overwhelming.

AI integration has introduced a disruption to manual study paradigms. Gandhi et al. (2023) suggest that AI-driven automation mitigates mental workload and helps prevent academic burnout, as the labor-intensive preparation of materials typically consumes a significant portion of a student's productive time. However, this convenience may bypass the fundamental encoding processes required for the transition of information into long-term memory. Furthermore, student engagement may diminish, facilitating a transition from active cognitive participation to passive consumption through rapid automation.

Student perceptions of AI platforms evolve as they become accustomed to these functionalities. Statistics indicate that 86% of students worldwide now utilize AI tools. ChatGPT is the most prevalent platform, used by 66% of the student population (Digital Education Council, 2024). Given the structured and clear presentation of AI outputs, students may develop an automation bias, where the ease of readability is conflated with actual information retrieval and mastery, potentially impacting test performance.

While AI-generated reviewers reduce the time needed for study preparation and lessen cognitive workload, they may also reduce students' active engagement with the material, which can affect understanding and test performance. In contrast, self-created reviewers may promote active recall and deeper cognitive processing, but may require more time and mental effort. Consequently, students face the challenge of choosing a reviewer method that balances time efficiency, cognitive workload, and test performance without compromising active recall.

This study seeks to investigate the relationship between the use of ChatGPT-generated versus self-created reviewers and student test scores, time efficiency, and cognitive workload. It examines how the increasing adoption of AI tools affects cognitive processing and academic outcomes. While AI-generated reviewers enhance efficiency and reduce mental strain, they may concurrently encourage shallow processing, diminished critical thinking, and an over-reliance on automated systems. By comparing these two methods, the study aims to understand how different methods of creating reviewers influence students' memory encoding, academic performance, and mental workload.

RELATED LITERATURE

Students have their own interpretations of study materials, which they make sense of when creating reviewers. During this process, they extract text from their study materials and summarize them into their own words or in ways that allow them to better understand the lesson. Literature suggests that students possess unique interpretations and understanding of information, which was described as time-consuming, and may result in inability to complete their study materials on time or lack time to understand and

revise their notes (Sajjacholapunt & Joy, 2014). On the contrary, another study revealed that note-takers achieved superior recall results on a post-test, although the process still required a significant amount of time and cognitive effort to prepare (Kulhavy et al., 2014). Additionally, Kieft (2006) showed that the interpretation skills of students significantly assist in writing and understanding review materials, which, in turn, leads to better recall of information and contributes to good future test scores.

Given the limitations of creating reviewers manually, alternative approaches have emerged, particularly through the use of generative AI programs such as ChatGPT. Such systems are capable of summarizing study materials relatively quickly, offering a potential solution to the inefficiencies associated with manually constructed reviewers (Kasneci et al., 2023; Noy & Zhang, 2023). Existing literature suggests that AI-assisted learning can support academic performance if used appropriately. Kung et al. (2023) demonstrated that AI-generated responses were comparable to the quality of answers provided by students on standardized assessments. However, reduced direct engagement with the material may compromise an understanding and long-term retention of the lessons (Fiorella & Mayer, 2015), as opposed to self-created reviewers.

This trade-off extends to time efficiency. The use of AI-generated reviewers reduces the time allocated for preparing study materials, allowing students to effectively review without the burden of manual labor (Kasneci et al., 2023). Findings from human–AI interaction studies also indicate that AI assistance enables faster completion of summarization tasks, and directly mitigates the time constraints inherent in traditional note-taking (Noy & Zhang, 2023). Nevertheless, this streamlined process may come at the expense of disengaged students since they are less involved in organizing and interpreting the material's core contents. Similarly, AI tools further influence learning outcomes by redistributing the experienced workload. Grounded in Cognitive Load Theory, lowering such load can enhance learning efficiency by minimizing unnecessary mental efforts (Paas & van Merriënboer, 1994). However, research on generative learning emphasizes that a deeper understanding is solidified when structuring and summarizing information manually (Fiorella & Mayer, 2015).

Overall, there exists crucial gaps despite the growing literature on self-created and AI-generated study materials. Most studies examine these approaches separately, with limited comparison of overall test scores, time efficiency, and cognitive workload. While AI tools have been shown to enhance speed and support short-term understanding, there is still insufficient evidence on how these gains translate into actual academic performance. With these, it is essential to conduct studies that directly compare AI-generated and self-created reviewers across different measures to better gauge their overall effectiveness.

METHODS

This study used a mixed-methods research design. A within-subjects experiment was conducted where each participant experienced both AI-generated reviewer and self-created reviewer. This allowed direct

comparison of the two methods within the same participants, which controlled for the variation due to individual differences and made the comparison of test scores, perceived mental workload, and time efficiency more reliable. To reduce possible order effects, administration of the reviewer methods were counterbalanced.

The qualitative component supplemented the quantitative findings. Before the experiment, participants completed a questionnaire about prior AI use and study habits. They also provided short feedback after each reviewer method and an overall exit reflection on their experiences, including what they found easy or difficult, what helped or hindered their studying, their confidence in the reviewer used, and which method they perceived as more efficient, mentally demanding, or helpful for understanding which provided context for the quantitative results.

Students from the College of Engineering at University of the Philippines Diliman, recruited using convenience sampling, participated in the experiment conducted at the Kaizen Laboratory (KIM). Prior to participation, all interested students were informed about the nature, purpose, and procedure of the study and were asked to sign an informed consent form.

The data collected included test scores to measure performance, perceived cognitive workload when making the reviewer using the NASA-TLX, and the time it took to generate the reviewer (time efficiency) for both AI-generated and self-created reviewers. In addition, demographic information and covariates such as note-taking frequency when reviewing, frequency of generative AI use for academic purposes, perceived significance of creating a reviewer, perceived exhaustion when making reviewers, and satisfaction with the created reviewer for both methods were collected and analyzed. After each reviewer method, participants answered a questionnaire about their experience with it, specifically the perceived mental effort, ease of use, organization, completeness of information, satisfaction, recall, confidence, understanding, efficiency, and challenges encountered. Participants then compared the ChatGPT-generated and self-created reviewer methods in terms of efficiency, mental demand, usefulness for understanding, usefulness under time pressure, and overall preference.

Data were analyzed using descriptive statistics and Analysis of Covariance (ANCOVA) to examine if the method significantly affects test scores, perceived cognitive workload, and time efficiency after controlling for participant-related factors. The covariates included in the ANCOVA are note-taking frequency while reviewing materials, frequency of ChatGPT or similar AI tool use for academic purposes, significance of creating study reviewers as part of the participant's learning style, mental exhaustion after creating a study reviewer, and satisfaction with the quality of self-created reviewers compared to AI-generated reviewers.

The qualitative responses from the post-test feedback forms and exit reflection questionnaire were analyzed using thematic analysis. Responses were reviewed and grouped into recurring themes. These qualitative findings were used to support and explain the quantitative results by providing insight into participants' experiences with both reviewers.

RESULTS

The study involved 30 UP Diliman engineering students, comprising 16 male (53.33%), 13 female (43.33%), and 1 who preferred not to say (3.33%). 17 participants were second-year students (56.67%), while 13 were third-year students (43.33%). The participants' were aged 18 to 23, with most aged 21 (36.67%) and 22 (40.00%) years old.

Descriptive statistics were used to summarize the participants' test scores, NASA-TLX workload scores, and study time under the ChatGPT-generated and self-created reviewer methods. Results in Table 1 showed that participants obtained a higher average test score in the self-created reviewer method (M = 11.03, SD = 2.03) than in the ChatGPT-generated reviewer method (M = 9.97, SD = 2.14). This suggests that self-created reviewers may be associated with slightly better test performance. However, the self-created reviewer method also produced higher perceived workload. Participants reported a higher average NASA-TLX score when using self-created reviewers (M = 6.28, SD = 1.49) than when using ChatGPT-generated reviewers (M = 4.12, SD = 1.46). Students reported manually creating reviewers was more mentally demanding than using AI-generated reviewers. In terms of time efficiency, participants spent less time in the ChatGPT-generated reviewer (M = 21.43 minutes, SD = 9.40) than in the self-created reviewer (M = 25.97 minutes, SD = 9.83). This suggests that ChatGPT-generated reviewers may reduce study preparation time and improve efficiency. Overall, the descriptive results show a trade-off between the two methods: ChatGPT-generated reviewers reduced study time and perceived workload, while self-created reviewers produced slightly higher test scores.

Table 1: Descriptive statistics results.

Reviewer Method	Variable	Mean	Standard Deviation	Minimum	Maximum	Range
ChatGPT-generated	Test Score	9.966	2.141	6	13	7
	NASA-TLX score	4.122	1.464	1.83	7	5.16
	Study Duration	21.43	9.401	7.00	40.00	33.00
Self- created	Test Score	11.03	2.025	8	15	7
	NASA-TLX score	6.283	1.486	3	8.83	5.83
	Study Duration	25.97	9.827	3.00	40.00	37.00

For test scores, the ANCOVA showed a significant main effect of the reviewer method (F = 5.421, p = .0295) after controlling for the covariates. Self-created reviewers were associated with slightly higher adjusted test scores than ChatGPT-generated reviewers. Among the covariates, only note-taking frequency significantly predicted test scores (F = 6.997, p = .0142), suggesting that students who engaged more frequently in note-taking tended to perform better regardless of the method. Other covariates, AI-use frequency, reviewer-making habit, mental exhaustion, and satisfaction with self-created reviewers, were not significant.

For perceived workload, the ANCOVA also showed a significant main effect of the reviewer method ($F = 48.84$, $p < .001$). Participants reported lower perceived workload under the ChatGPT-generated reviewer method than the self-created reviewer method. Among the covariates, the significance of creating reviewers as part of the participant's learning style was significant ($F = 5.799$, $p = .0241$), suggesting that students who considered reviewer-making an important part of their learning style experienced higher perceived cognitive workload. The remaining covariates were not significant.

For time efficiency, the ANCOVA showed a significant main effect of the reviewer method ($F = 16.44$, $p < .001$), indicating that the time required to complete the study task differed between the ChatGPT-generated and self-created reviewer methods. None of the covariates were significant.

Overall, the ANCOVA results indicate that the reviewer method significantly influenced all outcomes measured: test scores, perceived cognitive workload, and time efficiency. ChatGPT-generated reviewers were more advantageous in reducing workload and study time, while self-created reviewers were associated with better test performance. The results also show that note-taking habits remain important for academic performance, and reviewer-making preference may influence the perceived cognitive workload experienced when making the reviewer.

Pearson's correlation analysis was used to examine the relationships among study time, test scores, and NASA-TLX workload scores. For study time and test scores, the results showed no significant relationship in either method. The ChatGPT-generated and self-created reviewers showed almost no correlation with $r = 0.089$ and $r = -0.0076$, respectively. This suggests that spending more time studying did not necessarily correspond to higher test scores for either method.

For study time and NASA-TLX scores, the self-created reviewer method showed a weak positive correlation ($r = 0.1773$), suggesting that longer time spent manually creating or studying reviewers was slightly associated with higher perceived workload. In contrast, the ChatGPT-generated reviewer method showed almost no relationship between study time and workload ($r = 0.0026$) indicating that perceived workload remained relatively stable regardless of study duration.

For NASA-TLX scores and test scores, both methods showed negative correlations. The ChatGPT-generated reviewer method showed a weak negative relationship ($r = -0.2117$), while the self-created reviewer method showed a very weak negative relationship ($r = -0.1217$), suggesting that higher perceived workload was associated with lower test performance, although the strength of the relationship was weak.

Overall, the correlation results show that study duration alone did not predict academic performance. Instead, cognitive workload appeared to have a more relevant relationship with test scores, as higher workload tended to correspond with lower performance.

The qualitative responses were analyzed using thematic analysis resulting in five major themes: efficiency, mental demand, helpfulness for understanding, usefulness under time pressure, and reviewer method preference. First, participants viewed ChatGPT-generated reviewers as more efficient because

they allowed faster summarization and organization of study materials. Many participants noted that ChatGPT helped reduce preparation time and gave them more time to focus on understanding the content. However, some participants also considered self-created reviewers efficient in a different way because manually creating notes helped them absorb and recall the material better. Second, self-created reviewers were generally described as more mentally demanding. Participants reported that creating their own reviewers required more effort because they had to read, understand, organize, and type the material themselves, supporting the reported NASA-TLX workload scores being higher in the self-created reviewer method. Third, self-created reviewers were perceived as more helpful for deeper understanding. Participants explained that the act of creating and digesting their own notes forced them to actively engage with the material, explaining why self-created reviewers produced slightly higher test scores. Meanwhile, ChatGPT-generated reviewers were considered helpful for simplifying difficult topics and breaking down complex concepts into more understandable explanations. Fourth, ChatGPT-generated reviewers were preferred under time pressure. Participants found AI-generated reviewers useful when they had limited time because ChatGPT could quickly summarize the material and present the main ideas in an organized format. Lastly, participants showed context-dependent preferences. Self-created reviewers were preferred for deeper learning, retention, and mastery while ChatGPT-generated reviewers were preferred for speed, convenience, and cramming situations. This indicates that neither method was viewed as universally superior. Each method served a different learning purpose.

DISCUSSION

ChatGPT-generated reviewers produced significantly lower NASA-TLX scores, reducing students' mental effort (Deng et al., 2025). ChatGPT can lower cognitive workload by reducing the effort required to manually summarize, organize, and structure study materials. The present study also found that ChatGPT-generated reviewers were more time-efficient, aligning with Abbas et al.'s (2024) study, reporting that students are more likely to use ChatGPT when they experience high workload and time pressure. In the present study, participants similarly perceived ChatGPT-generated reviewers as faster and more useful when studying under limited time. However, Abbas et al. (2024) also reported that excessive ChatGPT use may be associated with procrastination, memory loss, and lower academic performance, especially when students rely on it instead of engaging in independent learning. This helps explain why ChatGPT-generated reviewers reduced cognitive workload but did not outperform self-created reviewers in test scores.

Findings for time efficiency are consistent with previous research on generative AI and productivity. Noy and Zhang (2023) found that generative AI significantly improved writing task completion speed while improving output quality, and Lee et al. (2024) reported that AI tools can assist users in completing writing and summarization tasks more efficiently. These findings support the present study's result that the AI-generated reviewer method

reduced the time required to prepare and study review materials. However, it was shown that saving time does not automatically lead to higher test performance, showing that ChatGPT is useful for efficiency, but students still need active learning strategies to strengthen understanding and recall.

The slightly higher test scores in the self-created reviewer method may be explained by generative learning theory. Roelle et al. (2023) explain that generative learning activities require learners to organize, elaborate, and make sense of information, which supports the construction of coherent mental representations. In the present study, creating one's own reviewer required students to actively select, rewrite, and organize the material, which may have supported deeper understanding. In contrast, ChatGPT-generated reviewers reduced workload by providing already organized material, but this may have limited the learner's active engagement with the content. Salame et al. (2024) also emphasized that note-taking supports learning through both encoding and external storage. The encoding function helps students process and organize information while creating notes, while the storage function allows them to review the material later. This supports the present study's finding that note-taking frequency significantly influenced test scores. Since self-created reviewers require students to select, organize, and rewrite important information, they may promote deeper processing and better recall compared to ChatGPT-generated reviewers, where much of the organization is completed by the AI. The higher NASA-TLX scores in the self-created reviewer method may be explained by the cognitive demands of note-taking. Piolat et al. (2005) described note-taking as a complex activity involving comprehension, selection of information, and written production. Since self-created reviewers require students to read, organize, and rewrite materials, the process naturally requires greater cognitive effort than using an already generated reviewer, similar to the present study's finding that self-created reviewers were more mentally demanding and required more time, while still showing potential benefits for deeper understanding and test performance.

CONCLUSION

This study compared the impact of ChatGPT-generated and self-created reviewers on UP Diliman engineering students' test scores, time efficiency, and perceived cognitive workload. Using a within-subjects design, findings showed a clear trade-off between the two approaches. ChatGPT-generated reviewers were associated with lower perceived cognitive workload and shorter average study time. This suggests that ChatGPT can help students study more efficiently by reducing the effort needed to summarize, organize, and prepare review materials. On the other hand, self-created reviewers produced slightly higher test scores and were perceived as more helpful for deeper understanding. This may be because manually creating reviewers requires active reading, selecting, rewriting, and organizing the material, thereby promoting deeper cognitive processing.

The study also found that note-taking frequency influenced test scores, while the significance of creating reviewers as part of one's learning style

influenced NASA-TLX workload scores. This indicates that students' learning habits still matter even when AI tools are available. Overall, the results suggest that ChatGPT-generated reviewers are useful for speed, convenience, and reduced cognitive load, while self-created reviewers remain important for active learning, deeper understanding, and retention.

Based on the findings, a hybrid study approach may be effective. ChatGPT may be used as an initial study-support tool to generate summaries, organize long readings, define key terms, and simplify complex concepts. This is especially useful when students have limited time or need a quick overview of the topic. However, students should not use ChatGPT-generated reviewers as the final study material without verification. Since AI may oversimplify information, overlook assumptions, or present incomplete explanations, students should compare AI-generated content with reliable, accurate sources. Next, students should still create their own notes despite using ChatGPT: they could generate an initial reviewer using ChatGPT, verify the information, rewrite key concepts in their own words, and create a personalized reviewer. This allows students to benefit from ChatGPT's speed while still gaining the learning benefits of active note-making. Next, engineering students should pair AI-generated reviewers with practice problems. Since engineering courses often involve formulas, assumptions, diagrams, and problem-solving steps, students should not only read summaries. They should solve sample problems, explain solution steps, identify common mistakes, and test themselves after reviewing. Lastly, educators may guide students on responsible AI use. Instead of banning ChatGPT, instructors can require students to submit evidence of active engagement, such as annotated AI-generated reviewers, corrected AI outputs, self-made concept maps, or worked practice problems. For example, an improved study workflow may require students to label AI-generated content. Overall, ChatGPT may be used as an efficiency tool, but not as a replacement for learning. An effective approach is to combine AI assistance with active study strategies such as note-taking, self-testing, and practice problem-solving.

This study has limitations. First, the sample size was 30 UP Diliman engineering students, which may limit the generalizability of the findings to other student populations, universities, or disciplines. Second, the study focused only on immediate learning outcomes (test scores) and did not measure long-term retention or delayed test performance. Third, the study materials were limited to selected Physics topics with conceptual and basic problem-solving questions, so the results may differ for more complex engineering topics. Fourth, only ChatGPT was used as the AI tool, so the findings may not apply to other generative AI platforms. Fifth, the self-created reviewer method was limited to digital note-taking through Google Docs to maintain consistency, which may not fully represent students who prefer handwritten reviewers or other note-taking tools. Lastly, although the study measured some covariates such as AI-use frequency, note-taking frequency, reviewer-making habit, mental exhaustion, and satisfaction with self-created reviewers, other factors such as prior knowledge, motivation, sleep quality, fatigue, test anxiety, and learning style were not included.

Future studies should include a larger, more diverse sample to improve generalizability, involving students from different year levels, programs, and universities to determine whether the findings remain consistent across academic contexts and to explore disciplines that cover diverse topics and skills. Future research should explore delayed post-tests to measure long-term retention. This would help determine which reviewers are more effective not only for immediate performance but also for remembering concepts over time. Other AI tools or different prompting strategies may be explored to determine whether the quality of AI-generated reviewers affects performance. Finally, future studies may test a third method: a hybrid reviewer method in which students first use ChatGPT to generate an initial reviewer, then are required to revise, annotate, verify, and add practice questions. This would directly test whether combining AI efficiency with actively self-created learning material produces better outcomes than using either method alone.

DECLARATION OF AI USE

ChatGPT, Gemini, and Grammarly were used to guide sentence structure and grammar.

REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. International Journal of Educational Technology in Higher Education, 21, Article 10. https://doi.org/10.1186/s41239-024-00444-7

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Digital Education Council. (2024). *What students want: Key results from DEC Global AI Student Survey 2024*. https://www.digitaleducationcouncil.com/newsroom/what-students-want-key-results-from-dec-global-ai-student-survey-2024

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Gandhi, T. K., Classen, D., Sinsky, C. A., Rhew, D. C., Vande Garde, N., Roberts, A., & Federico, F. (2023). How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open, 6*(3), ooad079. https://doi.org/10.1093/jamiaopen/ooad079

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2006). Writing as a learning tool: Testing the role of students' writing strategies. *European Journal of Psychology of Education, 21*(1), 17–34. https://doi.org/10.1007/BF03173567

Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test expectancy on the learning of text material. *The Journal of Educational Research, 68*(10), 363–365. https://doi.org/10.1080/00220671.1975.10884802

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education. *PLOS Digital Health, 2*(2), Article e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence, 6*, Article 100221. https://doi.org/10.1016/j.caeai.2024.100221

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291–312. https://doi.org/10.1002/acp.1086

Roelle, J., Endres, T., Abel, R., Obergassel, N., Nückles, M., & Renkl, A. (2023). Happy together? On the relationship between research on retrieval practice and generative learning using the case of follow-up learning tasks. *Educational Psychology Review, 35*, Article 102. https://doi.org/10.1007/s10648-023-09810-9

Sajjacholapunt, P., & Joy, M. (2014). *Exploring patterns of using learning resources as a guideline to improve self-revision*. WRAP. https://wrap.warwick.ac.uk/id/eprint/61848/

Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2024). *Generative AI usage and exam performance*. arXiv. https://arxiv.org/abs/2404.19699

Zhou, X., Zhang, J., & Chan, C. (2024). Unveiling students' experiences and perceptions of artificial intelligence usage in higher education. *Journal of University Teaching and Learning Practice, 21*(6), 126–145. https://doi.org/10.53761/54fs5e77