

# Comparative Impact of ChatGPT and Conventional Search Tools on Clinical Reasoning Performance: A Randomized Crossover Study in Preclinical Medical Students [Response to Letter]

Adisak Nartthanarung<sup>1</sup>, Komson Plangsiri<sup>2</sup>, Pinkawas Kongmalai <sup>1</sup>

<sup>1</sup>Department of Orthopedics, Faculty of Medicine, Kasetsart University, Bangkok, Thailand; <sup>2</sup>Department of Orthopaedics, Faculty of Medicine, Srinakharinwirot University, Nakhon Nayok, Thailand

Correspondence: Pinkawas Kongmalai, Department of Orthopedics, Faculty of Medicine, Kasetsart University, Bangkok, Thailand, Email pinkawass@hotmail.com

## Dear editor

We thank Kalra for the thoughtful comments regarding our article. We appreciate the opportunity to clarify the interpretation of our findings and to place the study's limitations in appropriate context.

First, we agree that carryover effects are an important consideration in crossover educational studies. In this setting, a washout period cannot fully remove learning once it has occurred. For that reason, our 60-minute washout period should be interpreted as a practical measure to reduce immediate recall and tool-related priming, rather than as a guarantee of complete elimination of residual learning effects. We therefore agree that carryover cannot be fully excluded. However, this does not invalidate the study. Rather, it means the findings should be interpreted cautiously as short-term, within-learner comparisons in a structured educational setting. This interpretation is consistent with published guidance for randomized crossover trials, which emphasizes transparent acknowledgment of period and carryover effects rather than assuming that any washout period can fully resolve them.<sup>1</sup>

Second, we agree that the single-institution setting and modest sample size limit generalizability. This was an educational study conducted within one authentic classroom cohort, and we do not claim broad external validity. At the same time, the randomized crossover design was chosen precisely because it improves internal efficiency by allowing participants to serve as their own controls. In that context, the design remains appropriate for an initial study of short-term educational performance, even though confirmation in larger and multicenter cohorts is still needed.<sup>1</sup>

Third, we agree that the eight-point rubric would be strengthened by additional psychometric evidence. Clinical reasoning is a complex construct, and rubric-based assessment should ideally be supported by broader validity evidence and reliability reporting. At the same time, the use of a structured rubric is not, in itself, a methodological weakness. On the contrary, rubric-based assessment is a recognized approach in medical education when grounded in relevant domains and applied transparently. Our intention was to use a practical structured measure aligned with the objectives of case-based learning, not to claim that this single rubric fully captures the entire construct of clinical reasoning. Future work should expand this by including formal inter-rater reliability and broader validity evidence.<sup>2</sup>

Fourth, regarding statistical analysis, we agree that repeated comparisons in crossover studies should be interpreted carefully. For that reason, our findings should not be read as proof of cumulative superiority across phases without reservation. However, we respectfully disagree that the statistical approach was inappropriate. Paired analysis is a reasonable method for within-subject comparisons in this educational design, and effect sizes were reported to complement p-values. The main point of our results was not to make an inflated confirmatory claim, but to show that



performance improved across both learning conditions in the short term. We agree that future studies may benefit from more advanced crossover-specific modeling and clearer prespecification of multiplicity handling.<sup>3</sup>

Fifth, we agree that variability in prompting is relevant. Prompt formulation can influence large language model output, and this should be recognized when interpreting reproducibility. However, we deliberately allowed students to use self-directed prompting because our aim was to reflect authentic educational use rather than an artificial, tightly scripted interaction. We therefore view prompt variability not only as a limitation, but also as part of the educational realism of the study. Future research should compare standardized and self-directed prompting strategies more directly. Recent reporting guidance for studies involving large language models supports more explicit documentation of prompts, model versions, and human oversight, and we agree that this is an important direction for the field.<sup>4</sup>

Sixth, we agree that hallucinations and inaccurate AI-generated content are important risks. However, we do not agree that this concern negates the value of the study. Our conclusions do not endorse uncritical use of AI outputs. On the contrary, we explicitly emphasize verification, critical appraisal, and faculty oversight. This cautious position is aligned with current international guidance, which highlights the risks of false or incomplete outputs and the need for human supervision when generative AI is used in health-related contexts.<sup>5</sup>

Finally, we agree that the study evaluated short-term outcomes only. We therefore do not claim long-term retention, transfer to clinical practice, or broader curricular effectiveness. Instead, we view this study as an initial contribution showing that both ChatGPT-assisted and conventional search-supported learning can improve short-term clinical reasoning performance in a structured preclinical setting. That question remains educationally relevant, even while longer-term and multicenter studies are clearly warranted.

We thank the author again for the constructive comments. We believe these points help refine the interpretation of our results, while also supporting the value of the study as an early controlled evaluation of AI-supported learning in medical education.

## Disclosure

The authors report no conflicts of interest in this communication.

## References

1. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. *BMJ*. 2019;366:14378. doi:10.1136/bmj.14378
2. Smith S, Kogan JR, Berman NB, Dell MS, Brock DM, Robins LS. The development and preliminary validation of a rubric to assess medical students' written summary statements in virtual patient cases. *Acad Med*. 2016;91(1):94–100. doi:10.1097/ACM.0000000000000800
3. Bender R, Lange S. Adjusting for multiple testing—when and how? *J Clin Epidemiol*. 2001;54(4):343–349. doi:10.1016/S0895-4356(00)00314-0
4. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–16. doi:10.1016/j.amjmed.2005.10.036
5. Organization WH. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. 2024.

Dove Medical Press encourages responsible, free and frank academic debate. The content of the Advances in Medical Education and Practice 'letters to the editor' section does not necessarily represent the views of Dove Medical Press, its officers, agents, employees, related entities or the Advances in Medical Education and Practice editors. While all reasonable steps have been taken to confirm the content of each letter, Dove Medical Press accepts no liability in respect of the content of any letter, nor is it responsible for the content and accuracy of any letter to the editor.

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

<https://doi.org/10.2147/AMEPS615304>