Critique of "The Effect of ChatGPT on Students' Learning Performance, Learning Perception, and Higher-Order Thinking: Insights from a Meta-Analysis"

Introduction

This meta-analysis, published in 2025, evaluates ChatGPT's impact on student learning performance, perception of learning, and higher-order thinking skills. The authors reviewed 51 experimental or quasi-experimental studies published between November 2022 and February 2025 (Nature). The study's scope aims to clarify whether ChatGPT reliably supports educational outcomes across diverse settings.

Summary of the Article

Analyzing aggregated data, the authors report a large effect size ($g = 0.867$) for learning performance and moderate effects ($g \approx 0.456–0.457$) for learning perception and higher-order thinking. They conducted moderator analyses showing that the impact on performance varies by course type, learning model, and usage duration; perception outcomes depend on duration; and higher-order thinking varies with course type and ChatGPT's role. The study also addressed publication bias with funnel plots and fail-safe N tests, concluding the results are stable and reliable. The authors recommend structured educational frameworks, broader implementation across grades, integration into varied learning models, sustained use over 4–8 weeks, and flexible roles for ChatGPT (tutor, partner, tool).

Evaluation of Methods and Findings

The use of meta-analysis and inclusion of moderator analyses give the study depth. The authors' handling of risk of bias via funnel plots and fail-safe tests adds robustness. Yet, details on how studies were selected or their quality assessed remain limited, which makes assessing the overall reliability of included evidence difficult.

Effect sizes are clearly reported, but the absence of variation metrics (e.g., confidence intervals range for each outcome) leaves understanding of result variability incomplete. The moderator findings are informative, but further interpretation regarding why certain learning models or course types amplify outcomes would enhance practical value.

Strengths and Limitations

Strengths include the large dataset, quantified effect sizes, and thoughtful bias assessment. The study's structured recommendations add value for educators and policymakers.

Limitations lie in the insufficient transparency of inclusion criteria and insufficient depth in reporting variance and context-specific explanations. Without clarity on how studies were appraised for quality, confidence in generalizing these findings weakens.

Conclusion

Wang and Fan deliver compelling quantitative support for ChatGPT's positive educational effects, especially on performance. Their large-scale meta-analysis offers a promising foundation for integrating AI tools in education. Strengthening future work requires deeper scrutiny of study

selection, variance reporting, and exploration of contextual effects. Despite limitations, the study stands as a meaningful contribution toward understanding how generative AI supports learning.