World Journal of Applied Studies
www.worldpaperjournal.com
Online ISSN:
Received: 05-12-2025, Accepted: 05-01-2026, Published: 20-01-2026
Volume 2, Issue 1, 2026, Page No. 4-6

Scaffolding, not substitution: How instructional design shapes Generative AI's effect on student learnin

Mina Rui Zhang
Department of Education, Nanyang Technological University, Singapore

Abstract
Generative artificial intelligence (GenAI) tools are now embedded in higher education digital learning, yet the literature reports strikingly inconsistent effects on student outcomes, ranging from strong performance gains to evidence of reduced independent reasoning. This review argues that this inconsistency is not primarily a disagreement about GenAI's capabilities but a reflection of how it is instructionally deployed. Synthesizing recent meta-analyses and qualitative studies (2023–2026), the review distinguishes between "substitution" uses, in which GenAI directly completes a cognitive task for the student, and "scaffolding" uses, in which GenAI structures, questions, or supports the student's own reasoning process. Evidence indicates that substitution-oriented use correlates with cognitive offloading and weaker higher-order thinking, while scaffolding-oriented use correlates with stronger, more durable learning gains. The review examines this distinction across writing, programming, and critical-thinking tasks, and considers the institutional and assessment-design barriers that make scaffolded use harder to sustain than substitution at scale. It concludes that instructional design, not the technology itself, is the primary determinant of whether GenAI helps or harms student learning.

Keywords: Generative AI, instructional scaffolding, digital learning, higher education, cognitive offloading, critical thinking

Introduction

Since ChatGPT's public release in late 2022, generative AI (GenAI) has become a routine feature of digital learning in higher education. A puzzle has emerged in the resulting research literature: some studies and meta-analyses report substantial positive effects of GenAI use on academic performance, while others report cognitive offloading, diminished independent reasoning, and superficial engagement. Rather than treating these as competing claims about whether GenAI "works," this review proposes that the two bodies of evidence are largely compatible once instructional design is taken into account. It distinguishes GenAI use that substitutes for a student's own cognitive effort from GenAI use that scaffolds it, and asks what the literature indicates about how this distinction shapes learning outcomes across common digital-learning tasks.

Review Approach

This narrative review draws on recent PRISMA-based systematic reviews and meta-analyses of GenAI in higher education, supplemented by targeted searches for studies published between 2023 and 2026 addressing critical thinking, cognitive offloading, and instructional design. Search terms combined "generative AI" or "ChatGPT" with "scaffolding," "cognitive offloading," "critical thinking," and "learning outcomes." Sources were prioritized where they reported outcome data comparing different modes or intensities of GenAI use, or where they offered conceptual frameworks distinguishing productive from unproductive integration.

Two Modes of GenAI Use in Digital Learning

The literature on GenAI in higher education implicitly separates into two patterns of use. In the substitution pattern, students direct GenAI to complete a task end-to-end, such as generating a full essay draft, providing a direct answer to a problem, or producing working code, with limited independent processing of the underlying reasoning. In the scaffolding pattern, GenAI is used to support a task the student still performs themselves, such as posing clarifying questions, offering partial explanations, checking reasoning, or providing feedback on student-generated work. The same tool can be used in either mode depending on how a task is designed and how students are guided to use it.

This distinction maps onto findings across several domains. In academic writing, GenAI used to brainstorm, restructure arguments, or improve fluency functions as scaffolding, and is associated with reduced cognitive load during drafting without necessarily reducing the student's own analytical work. In programming education, GenAI used as a collaborative partner that explains errors and prompts revision has been linked to improved computational thinking and self-efficacy, whereas direct code generation without engagement with the underlying logic offers less durable benefit. In critical-thinking tasks, structured interventions that use GenAI as a dialogic partner posing questions and requiring justification have shown promise for supporting reasoning, in contrast to unstructured use in which GenAI simply supplies conclusions.

Evidence on Effectiveness

Meta-analytic evidence broadly supports positive average effects of GenAI use on academic performance, though effect sizes vary considerably across syntheses. One meta-analysis of 35 experimental studies found stronger effects on cognitive outcomes such as problem-solving than on noncognitive outcomes such as motivation. Another synthesis of 51 studies reported a large effect on learning performance, concentrated in skills-based courses with well-defined objectives; a comprehensive review of 69 experimental studies similarly found strong effects on achievement and higher-order thinking but weaker,

inconsistent effects on self-efficacy and mental effort. A meta-analysis of randomized controlled trials reported a large pooled effect with low evidence of publication bias.

Read through the substitution-scaffolding lens, a pattern emerges from the moderator analyses embedded in these syntheses: effects are consistently larger when GenAI supplements rather than replaces traditional instruction, and larger for tasks with well-defined procedural structure than for open-ended tasks lacking scaffolding. This is consistent with the interpretation that GenAI's average positive effect on performance is driven disproportionately by scaffolded or structured use cases, while more open-ended, substitution-prone use contributes the variability and occasional null or negative findings reported elsewhere in the literature.

Evidence specifically on higher-order thinking reinforces this picture. Experimental work applying Bloom's taxonomy to postgraduate business students found GenAI's benefit concentrated at lower cognitive levels such as recall, with weaker benefit for analysis and synthesis, the levels most vulnerable to substitution. Qualitative interview studies describe students who rely heavily on GenAI reporting diminished independent reasoning and reduced analytical persistence, a pattern researchers have linked to cognitive offloading, in which tasks such as idea generation and evaluation are transferred to an external tool rather than performed internally.

Barriers to Sustaining Scaffolded Use

If scaffolded use produces more favorable outcomes than substitution, a key implementation question is why substitution nonetheless remains common. Several barriers are evident in the literature. First, scaffolded use requires more deliberate task and prompt design than substitution, placing additional demands on instructors who may lack confidence or training in structuring GenAI-integrated assignments; survey-based work on instructor attitudes finds substantial variation in confidence and beliefs about GenAI, which shapes how consistently structured use is modeled for students.

Second, assessment structures built around finished products, such as take-home essays or standalone problem sets, are inherently easier to complete through substitution than assessments that require demonstrated process, such as staged drafts, oral defenses, or annotated reasoning. Academic integrity debates have intensified partly because conventional assessment design does not distinguish between these two modes of use, making scaffolded and substitution-based GenAI use difficult to differentiate after the fact.

Third, students themselves face incentives that favor substitution under time pressure, even when they recognize its limitations for learning; some scholars caution that educational technologies promoted as universally beneficial can carry detrimental effects when adopted without attention to how they reshape student effort and incentives. Finally, unequal access to structured AI-literacy instruction means that scaffolded, pedagogically guided GenAI use is more available to students at well-resourced institutions, while students elsewhere are more likely to encounter GenAI only in its unstructured, substitution-prone form.

Discussion and Implications

The evidence reviewed here suggests that the central question for digital learning is not whether GenAI helps or harms student learning in general, but under what instructional conditions each outcome occurs. This has direct implications for practice. Instructors can shift assignment design toward processes that make substitution less viable and scaffolding more natural, such as requiring students to submit and reflect on GenAI-generated drafts rather than submitting them directly, or using GenAI-generated explanations as a starting point for independent verification rather than a final answer. Explicit modeling of scaffolded prompting, in which students are taught to ask GenAI for questions, counterarguments, or partial guidance rather than complete solutions, appears to be a transferable and teachable skill rather than an incidental byproduct of tool access.

For institutional policy, this framing argues against treating GenAI governance purely as a detection-and-enforcement problem. Policies focused solely on identifying AI-generated text do not address the underlying incentive structure that makes substitution attractive, whereas assessment redesign and instructor support for scaffolded task design address the incentive directly. For research, the substitution-scaffolding distinction suggests a need for studies that experimentally manipulate mode of use, rather than only presence or absence of GenAI access, since much of the observed variability in effect sizes across the literature may be attributable to unmeasured differences in how GenAI was actually used within each study's intervention.

Conclusion

The apparent contradiction between studies showing GenAI enhances learning and studies showing it erodes independent reasoning can be substantially reconciled by distinguishing scaffolding from substitution as modes of use. Current evidence suggests scaffolded GenAI use, in which the technology supports rather than replaces student reasoning, is associated with stronger and more durable learning benefits, while substitution use is more closely linked to cognitive offloading and weaker higher-order thinking. Because current assessment structures and institutional incentives often make substitution the path of least resistance, sustaining scaffolded use will require deliberate assignment redesign, instructor support, and equitable access to AI-literacy instruction, rather than depending on the technology to produce beneficial outcomes on its own.

References

1. Ali JKM, et al. Potential benefits and challenges of using ChatGPT in educational contexts: A systematic review. [As cited in Nature Humanities and Social Sciences Communications, 2025], 2024.
2. Cong-Lem N, Nguyen TT, Nguyen KNH. Critical thinking in the age of generative AI: Effects of a short-term experiential learning intervention on EFL learners. International Journal of TESOL Studies, 2025.
3. Darwin, Rusdin D, Mukminatien N, Suryati N, Laksmi ED, Marzuki. Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. Cogent Education,2024:11(1):2290342.

4. Deng R, et al. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 2025.
5. Liu Y, et al. The impact of ChatGPT on students' academic achievement: A meta-analysis. Journal of Computer Assisted Learning, 2025.
6. Selwyn N. On the limits of artificial intelligence (AI) in education. Nordisk Tidsskrift for Pedagogikk og Kritikk,2024:10(1):3–14.
7. Sun T, Zhou Y. Meta-analysis of the impact of GenAI-supported programming instruction on university students' academic achievement. [As cited in Nature Humanities and Social Sciences Communications, 2026], 2024.
8. Vendrell M, Johnston S. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. Computers and Education: Artificial Intelligence,2026:10:100572.
9. Wang X, Fan J. Meta-analysis of randomized controlled trials on the efficacy of ChatGPT for students' learning outcomes. [As cited in arXiv preprint 2509.07171], 2025.
10. Wu F, Dang Y, Li M. A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education. Behavioral Sciences,2025:15(4):467. https://doi.org/10.3390/bs15040467
11. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments,2024:11(1):28.