Generative AI Tools in Higher Education: A Meta-Analysis of Cognitive Impact

Xiaodong Qu
Computer Science Department
The George Washington University
Washington, District of Columbia, USA
x.qu@gwu.edu

Joshua Sherwood
Computer Science Department
The George Washington University
Washington, D.C., District of Columbia, USA
jsherwood@gwu.edu

Peiyan Liu
Computer Science Department
The George Washington University
Washington DC, District of Columbia, USA
peiyan.liu@gwu.edu

Nawwaf Aleisa
Computer Science Department
The George Washington University
Washington DC, District of Columbia, USA
nawwaf.aleisa@gwu.edu

Abstract

This meta-analysis examines the cognitive impact of Generative Artificial Intelligence (GenAI) tools on college students, focusing on various levels of Bloom's taxonomy. As GenAI integration in higher education grows, understanding its influence on critical thinking, problem-solving, and creativity is essential. Using a mixed-effects model, we synthesized data from quantitative studies to explore two moderators: cognitive skill level (e.g., understanding, applying, analyzing) and instructional context (instructed vs. unguided use). Our findings indicate that GenAI tools significantly enhance lower-order cognitive outcomes, particularly in understanding and applying concepts, with instructed use producing stronger positive effects than unguided use. However, their impact on higher-order cognitive skills, such as creating and evaluating, was minimal. These results highlight the importance of tailoring GenAI integration to task complexity and underscore the value of guided instruction in maximizing its educational benefits. Educators should prioritize instructional strategies that encourage active engagement with GenAI tools, particularly for fostering critical thinking and creativity.

CCS Concepts

• **Computing methodologies → Natural language processing**; **Artificial intelligence**; • **Human-centered computing → Interaction techniques**; **Empirical studies in HCI**; **HCI theory, concepts and models**.

Keywords

Generative AI, GenAI in education, Bloom's taxonomy, Education, Meta-analysis

ACM Reference Format:
Xiaodong Qu, Joshua Sherwood, Peiyan Liu, and Nawwaf Aleisa. 2025. Generative AI Tools in Higher Education: A Meta-Analysis of Cognitive Impact. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3706599.3719841

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719841

1 Introduction

In recent years, advancements in Generative Artificial Intelligence (GenAI) have demonstrated its ability to solve complex problems, offering new avenues for enhancing teaching and learning processes. These tools promise to bolster cognitive outcomes by fostering critical thinking, problem-solving, and creativity among college students. For instance, the introduction of OpenAI's ChatGPT (OpenAI, 2023) along with its successive versions including ChatGPT-4o, has highlighted its potential in providing formative/summative assessments and generating code, bringing these tools to the forefront for educators and policymakers. However, while the potential of GenAI in educational settings is widely recognized, there remains a significant gap in understanding how these tools impact different cognitive levels, as outlined in Bloom's taxonomy (Bloom, 1956/2001)[1]. Moreover, the effectiveness of these tools may vary depending on whether they are used in an instructed or unguided manner, yet this aspect has received limited attention in existing research. This meta-analysis aims to address these gaps by systematically reviewing and synthesizing empirical studies that investigate the impact of GenAI tools on college students' cognitive outcomes. Specifically, we focus on two key moderators: the cognitive level targeted (e.g., understanding, applying, analyzing) and the instructional context of GenAI tools use.

1.1 Research Questions

Our study seeks to answer the following research questions:

(1) What is the overall impact of GenAI tools on the cognitive outcomes of college students?
(2) How do these effects vary across different levels of Bloom's taxonomy?
(3) Does the guided use of GenAI tools enhance cognitive outcomes more effectively than unguided use?

By exploring these questions, this study contributes to the growing body of literature on AI in education and provides valuable insights for educators and policymakers aiming to leverage GenAI tools to their full potential.

2 Literature Review

GenAI tools, particularly ChatGPT, have emerged as transformative technologies in higher education. These tools provide automated feedback, generate solutions, and enhance learning experiences. Their use spans disciplines such as computer science, language learning, and creative arts, where they assist students in understanding and applying knowledge efficiently. For example, studies have shown that AI tools can reinforce foundational skills such as comprehension, grammar, and programming syntax. Svendsen et al. (2024)[16] demonstrated that pharmacy students using ChatGPT showed improvements in understanding complex drug mechanisms. Similarly, Lyu et al. (2024)[11] reported that AI tools helped novice programmers debug code and improve syntax knowledge, reinforcing the role of GenAI in supporting early learning stages.

Research examining the impact of GenAI tools often uses Bloom's taxonomy as a framework for evaluating cognitive outcomes. Bloom's taxonomy categorizes cognitive skills into lower-order skills (e.g., remembering, understanding, and applying) and higher-order skills (e.g., analyzing, evaluating, and creating). Multiple studies have demonstrated that GenAI tools are particularly effective for lower-order cognitive processes. Mahapatra (2024)[12] found that ESL students using ChatGPT received valuable formative feedback that improved their grammatical accuracy and organizational skills, helping them achieve better comprehension and application of writing principles. Similarly, Ododo, Essien, and Bassey (2024)[13] showed that ChatGPT-supported learning increased students' self-efficacy in applying basic programming concepts.

In contrast, the impact of GenAI tools on higher-order skills remains less consistent. Jošt, Taneski, and Karakatic (2024)[6] identified a negative correlation between reliance on AI tools and critical thinking skills in software development tasks, where deeper analysis and independent problem-solving were required. Urban et al. (2023)[17] observed modest improvements in creative tasks, such as idea generation, but noted that students struggled to critically evaluate or refine AI-generated outputs without guidance. These findings suggest that while GenAI tools excel at providing immediate support for foundational learning, their role in fostering higher-order cognitive skills requires further investigation.

The instructional context—whether students use GenAI tools with guidance or independently—has emerged as a key factor influencing learning outcomes. Research consistently shows that guided use of GenAI tools yields better results. Mahapatra (2024)[12] demonstrated that students in structured ESL writing activities achieved greater gains when specific prompts and feedback tasks were provided. In dental education, Kavadella et al. (2023)[7] found that students performed significantly better when instructors guided them to critically engage with ChatGPT-generated content for research purposes. These findings highlight the importance of scaffolding in helping students maximize the benefits of AI tools while maintaining active engagement.

In contrast, unguided use of GenAI tools can lead to suboptimal outcomes. Ododo, Essien, and Bassey (2024)[13] observed that students using ChatGPT without instructional support struggled with tasks requiring abstraction and higher-order problem-solving. Similarly, Jošt, Taneski, and Karakatic (2024)[6] noted that over-reliance on unguided AI use impeded students' ability to develop critical thinking skills independently. These studies emphasize the need for structured integration of GenAI tools into curricula, ensuring that students are encouraged to engage critically with AI-generated outputs.

3 Methods

3.1 Identification of Studies and Coding Schemes

A computerized database search generated a pool of articles for this meta-analysis using the query: *"college student AND LLM AND cognitive development AND learning OR ChatGPT OR GenAI"*. Broad terms ensured wide coverage, with limits set to studies on college students, published in English from January 2023 to September 2024, and available online. This timeframe captured recent empirical studies on generative AI and large language models (LLMs) in education. The initial search identified 1,305 peer-reviewed articles, which were screened for inclusion.

Screening criteria included: (a) title or abstract referencing college students, generative AI, and cognitive outcomes; (b) quasi-experimental design (e.g., pre/post or control groups); (c) descriptive statistics; and (d) sample size >10. Exclusions were based on: (1) systematic reviews or non-empirical reports; (2) opinion-based surveys; (3) lack of effect sizes or sufficient data; and (4) non-college populations. This process yielded 15 studies.

A back-searching strategy, reviewing references of selected studies, identified 24 additional articles, of which three met the inclusion criteria. The final sample included 18 studies (([2–4, 6–9, 11–21].). These studies, along with GenAI tools listed in Appendix 2, were included in the meta-analysis.

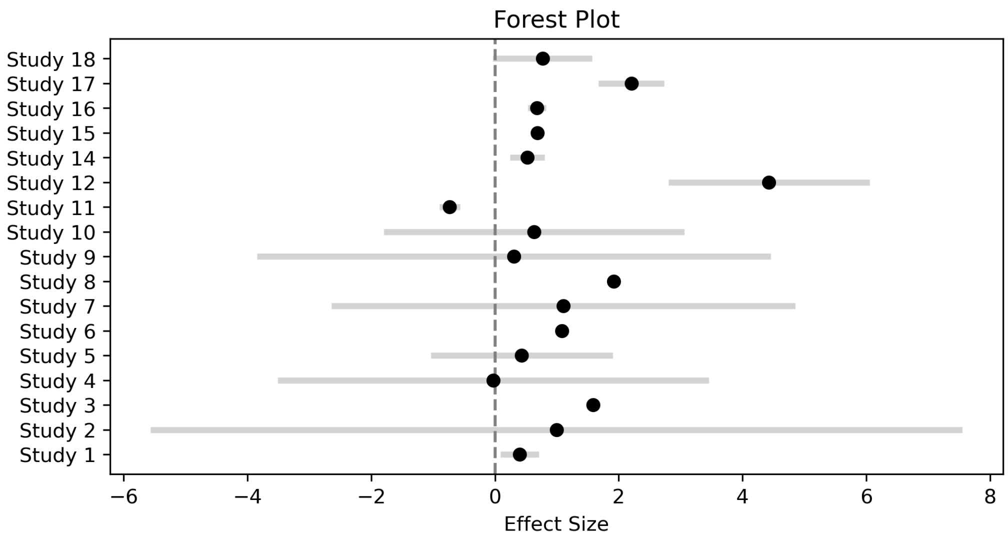
To examine the moderation effects of cognitive levels and instructional contexts in ChatGPT use, we reviewed 18 papers. Studies were categorized by cognitive skill level (lower-order: remembering, understanding, applying; higher-order: analyzing, evaluating, creating) and instructional context (instructed: guided use with prompt guidelines or designated points; unguided: free use without guidance).

Six STEM researchers with publication experience coded the papers, with each paper reviewed by three coders for validation. Final codes were based on majority consensus and used as input for the moderator analysis.

3.2 Effect Size Computation

Most studies in this meta-analysis reported means and standard deviations for experimental and control groups or pre- and post-conditions, allowing effect sizes to be calculated. Cohen's d, the standardized mean difference, was adjusted using Hedges' correction for small sample bias: $(1-\frac{3}{4N-9})\frac{M_e-M_c}{S_p}$, where N is the total sample size, M_e and M_c are the means for experimental and control groups, and S_p is the pooled standard deviation. When these values were unavailable, effect sizes were derived from other statistics, such as Pearson correlations, using Lipsey and Wilson's (2001)[10] formulas.

Figure 1: Forest Plot of the Mixed Effects Model



Effect sizes were corrected for bias. To address the upward bias of effect sizes in small samples, population effect sizes were estimated using Hedges' correction (Hedges, 1981)[5]. Additionally, inverse variance weights were calculated for each effect size, allowing the analyses to be weighted by inverse variance. This procedure gives larger samples more influence in the analyses than smaller samples (Lipsey & Wilson, 2001)[10].. In cases where a study provided more than one effect size for the same sample, specific criteria were applied to ensure consistency. For longitudinal studies, the effect size was calculated based on the most recent measurement. If a study provided multiple effect sizes from the same time point, the effect size with the largest sample size was selected and included in the analysis. Study# 13 (Yuan, 2024)[21] was excluded from the analysis due to the reported unrealistically substantial large effect size of 14.83. Such an extreme value could indicate a data entry error, or a methodological issue, and including it could distort the results.

3.3 Data Analysis

To address the research questions, data were analyzed using the mixed-effects model (Lipsey& Wilson, 2001)[10] and Python 3.0's statsmodels package. There was no missing data in the analyses. The mixed-effects model assumes that variability among effect sizes can be explained by both fixed effects (sampling error and moderators) and random effects (reliable deviation of a study from the mean of the population effect sizes distribution). This approach is preferable to using either fixed-effects or random-effects models alone, as it accounts for variability among effect sizes through both moderators and study-specific deviations, respectively.

To examine the heterogeneity of variance among the effect sizes, Q statistics were computed. Moderator analyses were conducted to determine whether the cognitive skill levels and/or the instructional context could explain the variability in effect sizes. These moderators were tested for significance using a multiple-regression-with-categorical-predictor analysis.

4 Results

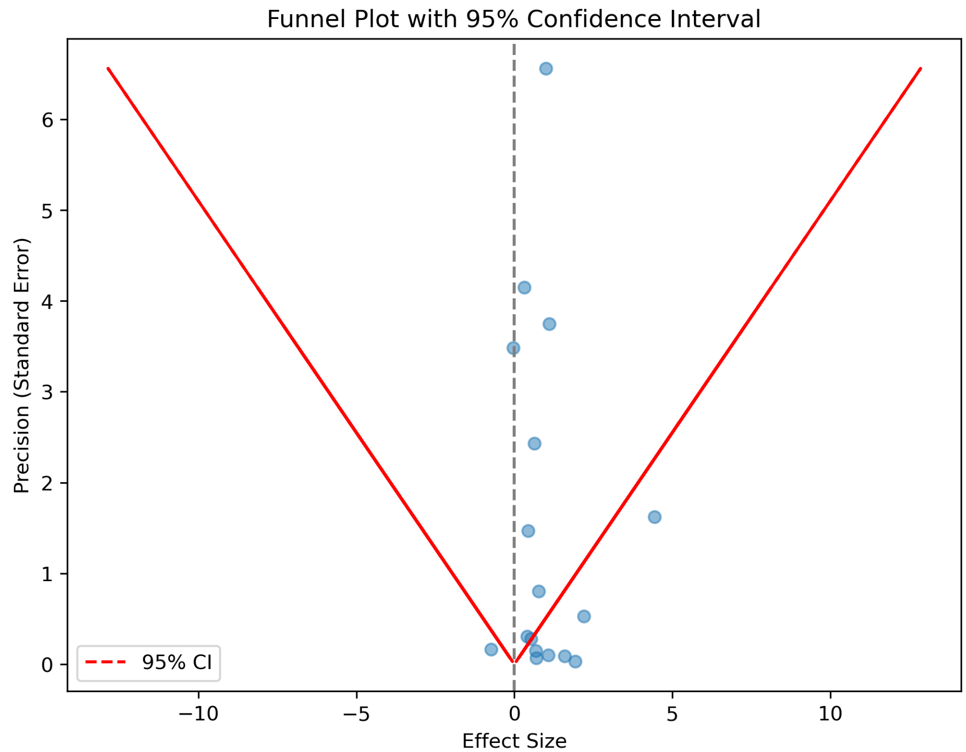
Although the test for homogeneity among effect sizes, $Q_{(16)}$ =20.82, p=0.19, indicated no significant heterogeneity, mixed-effects models with moderator analyses were still conducted. The fixed-effects model assumes that all studies estimate the same true effect size, but this assumption may not always hold in practice (Lipsey& Wilson, 2001)[10]. Mixed-effects models relax this assumption, allowing for the possibility that each study may have a different underlying true effect size, and that variability among studies could be influenced by potential moderators.

Figure 1 presents a forest plot of the data points, illustrating the effect sizes and relative weights of the 17 studies included in the analysis. The majority of the studies (15 out of 17) are positioned to the right of the no-effect line, indicating positive effect sizes. The effect sizes range from -0.73 to 4.43, with varying standard errors, represented by the grey horizontal lines. The length of each line reflects the uncertainty associated with each study's effect size, where shorter lines indicate more precise estimates. Overall, the plot highlights the consistency of positive effects across the studies, with only two studies showing negative effect sizes.

4.1 Mixed-effect Model with Bloom's Taxonomy Skill Level as One Moderator

A mixed-effects model estimated the overall effect size and examined Bloom's Taxonomy Skill Level as a moderator. The overall effect size was 1.34 (95% CI: 1.17, 1.50), indicating a significant positive impact of GenAI tools on college students' cognitive outcomes. Bloom's Skill Level (lower- vs. higher-order) significantly moderated the effect size ($\beta = -1.14, p < 0.001$), with smaller effects

Figure 2: Funnel Plot of the Studies



observed for higher-order skills. Random-effect variance (τ^2) was 0.52, reflecting moderate between-study variability.

Model fit indices (AIC: 55.34; BIC: 55.84) supported the inclusion of random effects. These findings highlight the consistently positive impact of GenAI tools on learning, moderated by cognitive skill levels. Statistics are summarized in Table 1.

4.2 Mixed-effect Model with Instructional Context as One Moderator

The second mixed-effects model included Instructional Context as a moderator. The overall effect size was 0.46 (95% CI: 0.39, 0.53), showing a significant positive impact of GenAI tools on learning. Instructional Context (instructed vs. unguided) significantly moderated the effect size ($\beta = 0.83, p = 0.02$), with guided use producing effect sizes 0.83 larger. Random-effect variance (τ^2) was 0.59, indicating moderate variability.

Model fit indices (AIC: 55.13; BIC: 57.63) supported random effects. These results suggest that guided use of GenAI tools enhances their positive impact on learning outcomes.

4.3 Mixed-effect Model with Bloom's Taxonomy Skills Levels and Instructional Context as Two Moderators

A mixed-effects model examined the impact of Bloom's Taxonomy Skill Level and Instructional Context on the effect size of integrating GenAI tools into college learning. The intercept was 0.93 (95% CI: 0.60, 1.27), indicating a significant positive effect of GenAI tools on cognitive outcomes.

Both moderators significantly influenced effect sizes. Bloom's Skill Level ($\beta = -0.95, p < 0.05$) showed smaller effect sizes for higher-order skills (analyzing, evaluating, creating) compared to lower-order skills (remembering, understanding, applying). Instructional Context ($\beta = 0.53, p < 0.05$) showed that structured guidance led to greater cognitive gains than unguided use.

Random-effect variance (τ^2) was 0.52, with fit indices (AIC: 53.75; BIC: 55.23) supporting random effects. The third model, including both moderators, had the best fit (AIC = 53.75, BIC = 55.23), outperforming models with individual moderators Table 1. These results highlight the importance of structured guidance and lower-order skills in maximizing the benefits of GenAI tools.

5 Discussion

5.1 Impact on Lower-Order Cognitive Skills

The findings confirm that GenAI tools are particularly effective at improving lower-order cognitive skills. This result aligns with multiple studies where tools like ChatGPT provided immediate, corrective, and tailored feedback. Svendsen et al. (2024)[16] demonstrated that pharmacy students using ChatGPT to explain drug mechanisms showed a significant improvement in comprehension compared to students in the control group. Similarly, Lyu et al. (2024)[11] reported that students using CodeTutor, a ChatGPT-powered assistant, improved their syntax comprehension and debugging abilities

Table 1: Mixed-effect Model Selection

Model	Intercept	β		Sig.(0.05)		τ^2	AIC	BIC
1	1.34	-1.14		$P < 0.001$		0.52	55.34	55.84
2	0.46	0.83		$P < 0.05$		0.59	55.13	57.63
3	0.94	$\beta_{Bloom's}$ -0.95	$\beta_{Instructional}$ 0.53	$P_{Bloom's}$ 0.05	$P_{Instructional}$ < 0.05	0.52	53.75	55.23

in introductory programming courses. The tool's ability to deliver real-time feedback enabled students to apply programming concepts with greater precision.

In language learning, Mahapatra (2024)[12] found that ESL students experienced significant gains in grammar, vocabulary, and organizational skills when they received structured formative feedback through ChatGPT. This support reduced cognitive load and allowed students to grasp foundational writing concepts more efficiently. Additionally, Ododo, Essien, and Bassey (2024) [13]observed increased programming self-efficacy among students in JAVA programming courses, where ChatGPT facilitated task comprehension and basic application of coding principles. These studies collectively reinforce the conclusion that GenAI tools excel at reinforcing foundational knowledge and helping students achieve lower-order cognitive outcomes.

5.2 Challenges with Higher-Order Cognitive Skills

While GenAI tools demonstrate clear benefits for lower-order tasks, their impact on higher-order cognitive processes is more limited. Jošt, Taneski, and Karakatic (2024)[6] reported a negative correlation between reliance on LLMs and performance in tasks requiring critical thinking, such as debugging and code generation. Students who relied heavily on ChatGPT for advanced problem-solving demonstrated poorer final grades, indicating that AI tools alone may not support the deeper cognitive engagement required for higher-order learning. Similarly, Urban et al. (2023)[17] found that ChatGPT effectively supported idea generation in creative problem-solving tasks but fell short in helping students evaluate and refine those ideas independently. Without structured intervention, students struggled to move beyond surface-level outputs generated by the AI.

Yuan (2024)[21] echoed these concerns in a study on AI-assisted music composition. While ChatGPT boosted students' motivation and initial idea generation, instructor support was essential for meaningful integration and refinement of AI outputs. These findings highlight a key limitation: GenAI tools can initiate and support higher-order processes but fail to replace the need for critical reflection and instructor guidance. Without active learning strategies, students risk becoming passive recipients of AI-generated content rather than active participants in the learning process.

5.3 The Role of Instructional Context: Guided vs. Unguided Use

Instructional context emerged as a key factor in moderating the effectiveness of GenAI tools. Studies consistently demonstrate that guided use—where students receive structured prompts, tasks, or instructional support—produces stronger cognitive gains compared to unguided use. Kavadella et al. (2023)[7] found that dental students using ChatGPT under instructor supervision achieved significantly better outcomes in research assignments. Structured prompts enabled students to engage critically with AI outputs, ask precise questions, and validate responses.

Similarly, Lyu et al. (2024)[11] reported that students in programming courses achieved better problem-solving outcomes when provided with scaffolded coding tasks and targeted feedback. In this guided setting, students were encouraged to actively analyze and refine ChatGPT suggestions rather than accepting them passively. Mahapatra (2024)[12] demonstrated a similar effect in ESL writing classrooms, where students provided with structured feedback prompts showed more significant gains than those using ChatGPT independently.

Conversely, unguided use of GenAI tools can limit learning outcomes, particularly in tasks requiring abstract or higher-order thinking. Ododo, Essien, and Bassey (2024)[13] observed that students who engaged with ChatGPT independently struggled with algorithm design and complex problem-solving tasks. Without clear instructional guidance, students may lack the critical scaffolding needed to bridge the gap between AI-generated suggestions and deeper cognitive processes. These findings highlight the importance of thoughtful instructional strategies to ensure students interact meaningfully with GenAI tools.

5.4 Over-Reliance on AI and Critical Thinking Risks

The potential for over-reliance on GenAI tools presents a major challenge. Jošt, Taneski, and Karakatic (2024)[6] reported that students who frequently relied on ChatGPT for critical tasks exhibited poorer performance, suggesting that excessive dependence on AI may undermine the development of critical thinking and independent learning. This concern was echoed by Lyu et al. (2024)[11], where students initially relied heavily on CodeTutor for programming tasks but eventually preferred traditional teaching assistants for complex problem-solving. Students cited the need for nuanced, human-driven feedback, which AI alone could not provide.

This shift highlights a key risk: while GenAI tools streamline learning, they may inadvertently discourage deeper cognitive engagement if students are not encouraged to critically evaluate and refine AI-generated outputs. Educators must carefully balance AI integration with activities that promote independent thought and reflective learning to prevent the passive acceptance of AI suggestions.

5.5 Disciplinary Variability in Cognitive Outcomes

The variability in outcomes across disciplines underscores the context-sensitive nature of GenAI tool integration. In language learning, Mahapatra (2024)[12] demonstrated that ChatGPT provided substantial benefits, particularly in improving organizational skills and grammatical accuracy. Similarly, Yuan (2024)[21] found strong motivational gains when ChatGPT was used to assist creative tasks such as music composition. These studies suggest that GenAI tools excel in disciplines where foundational knowledge, idea generation, and iterative refinement play central roles.

However, technical disciplines such as programming present a more nuanced picture. While tools like CodeTutor improve basic syntax comprehension and debugging skills, students often struggle with higher-order problem-solving tasks when relying on AI alone. This discrepancy emphasizes the need to tailor instructional strategies to the specific demands of each discipline. In technical fields, structured guidance becomes particularly critical to ensure that students develop both foundational knowledge and independent problem-solving abilities.

5.6 Publication bias

The funnel plot (Figure 2) shows that about a half of the studies fall within the 95% confidence interval, and an imbalance in the distribution—more studies clustered on the right side of the non-effect line—suggests the potential presence of bias. The result of Kendall's tau rank correlation (τ = -0.03, p = 0.87) suggests symmetry, but it is premature to rule out bias given the small sample size of 17 studies published within a narrow timeframe (January 2023 to June 2024). Future meta-analyses should include a broader and more diverse sample of studies to ensure a more comprehensive and balanced assessment of GenAI's impact.

6 Conclusion

This meta-analysis provides crucial insights into how GenAI tools like ChatGPT can significantly improve cognitive outcomes in college students, with an overall effect size of 0.94. The results clearly demonstrate that these tools excel in enhancing lower-order cognitive skills, such as understanding and applying, by reducing cognitive load and providing instant, tailored feedback. For example, studies like Mahapatra (2024)[12] and Lyu et al. (2024)[11] showcased notable improvements in ESL writing and programming skills when structured prompts and guidance were provided.

However, the impact of GenAI tools on higher-order skills—such as analyzing, evaluating, and creating—remains limited. Overreliance on AI tools, as reported in studies like Jošt et al. (2024)[6], can hinder the development of critical thinking and problem-solving abilities. This distinction emphasizes the need for educators to balance AI integration with traditional methods that foster deeper cognitive engagement.

The instructional context emerges as a critical factor in maximizing the benefits of GenAI tools. Guided use, where students receive structured prompts, feedback, or instructor support, yields significantly stronger outcomes than unguided, free use. For instance, Kavadella et al. (2023)[7] demonstrated that structured research tasks enabled dental students to engage more critically with AI-generated content, leading to improved learning outcomes.

These findings underscore the importance of aligning GenAI tool use with specific cognitive goals and instructional strategies. For foundational knowledge, GenAI tools offer a scalable and effective means of reinforcing concepts, particularly in resource-constrained settings. For higher-order skills, GenAI should serve as a supplementary aid, encouraging students to critically evaluate, refine, and extend AI-generated outputs. Educators must design blended learning environments that leverage the strengths of AI tools while mitigating their limitations.

While the transformative potential of GenAI tools is vast, their success depends on thoughtful, deliberate integration into curricula. Policymakers should develop frameworks that emphasize guided use and ensure alignment with cognitive skill development goals. Future research should focus on long-term impacts, exploring how GenAI can support not just foundational learning but also creativity, critical thinking, and problem-solving in diverse disciplines.

References

[1] Benjamin S Bloom, Max D Englehart, Edward J Furst, Walker H Hill, David R Krathwohl, et al. 1956. Taxonomy of educational objectives, handbook I: the cognitive domain. New York: David McKay Co.
[2] Harry Barton Essel, Dimitrios Vlachopoulos, Albert Benjamin Essuman, and John Opuni Amankwa. 2024. ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence* 6 (2024), 100198.
[3] Ying Guo and Daniel Lee. 2023. Leveraging chatgpt for enhancing critical thinking skills. *Journal of Chemical Education* 100, 12 (2023), 4876–4883.
[4] Philipp Haindl and Gerald Weinberger. 2024. Students' Experiences of Using ChatGPT in an Undergraduate Programming Course. *IEEE Access* 12 (2024), 43519–43529.
[5] Larry V Hedges. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *journal of Educational Statistics* 6, 2 (1981), 107–128.
[6] Gregor Jošt, Viktor Taneski, and Sašo Karakatič. 2024. The Impact of Large Language Models on Programming Education and Student Learning Outcomes. *Applied Sciences* 14, 10 (2024), 4115.
[7] Argyro Kavadella, Marco Antonio Dias Da Silva, Eleftherios G Kaklamanos, Vasileios Stamatopoulos, Kostis Giannakopoulos, et al. 2024. Evaluation of ChatGPT's real-life implementation in undergraduate dental education: mixed methods study. *JMIR Medical Education* 10, 1 (2024), e51344.
[8] Tomaž Kosar, Dragana Ostojić, Yu David Liu, and Marjan Mernik. 2024. Computer Science Education in ChatGPT Era: Experiences from an Experiment in a Programming Course for Novice Programmers. *Mathematics* 12, 5 (2024), 629.
[9] Jian Liao, Linrong Zhong, Longting Zhe, Handan Xu, Ming Liu, and Tao Xie. 2024. Scaffolding Computational Thinking with ChatGPT. *IEEE Transactions on Learning Technologies* (2024).
[10] Mark W Lipsey. 2001. Practical meta-analysis. *Thousand Oaks* (2001).
[11] Wenhan Lyu, Yimeng Wang, Tingting Chung, Yifan Sun, and Yixuan Zhang. 2024. Evaluating the effectiveness of llms in introductory computer science education: A semester-long field study. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*. 63–74.
[12] Santosh Mahapatra. 2024. Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments* 11, 1 (2024), 9.
[13] Emmanuel Philip Ododo, Nseabasi Peter Essien, and Ani Etim Bassey. 2024. Effect of Generative artificial intelligence (AI)-based tool utilization and students' programming self-efficacy and computational thinking skills in JAVA programming course in Nigeria Universities. *International Journal of Contemporary Africa Research Network* 2, 1 (2024).
[14] Basit Qureshi. 2023. Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv preprint arXiv:2304.11214* (2023).
[15] Dan Sun, Azzeddine Boudouaia, Chengcong Zhu, and Yan Li. 2024. Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 14.

[16] Kristian Svendsen, Mohsen Askar, Danial Umer, and Kjell H Halvorsen. 2024. Short-term learning effect of ChatGPT on pharmacy students' learning. *Exploratory Research in Clinical and Social Pharmacy* 15 (2024), 100478.
[17] Marek Urban, Filip Děchtěrenko, Jiří Lukavskỳ, Veronika Hrabalová, Filip Svacha, Cyril Brom, and Kamila Urban. 2024. ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education* 215 (2024), 105031.
[18] Tugba Aydın Yıldız. 2023. The Impact of ChatGPT on Language Learners' Motivation. *Journal of Teacher Education and Lifelong Learning* 5, 2 (2023), 582–597.
[19] Ramazan Yilmaz and Fatma Gizem Karaoglan Yilmaz. 2023. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence* 4 (2023), 100147.
[20] Jiaqi Yin, Tiong-Thye Goh, Bing Yang, and Yi Hu. 2024. Using a chatbot to provide formative feedback: A longitudinal study of intrinsic motivation, cognitive load, and learning performance. *IEEE Transactions on Learning Technologies* (2024).
[21] Na Yuan. 2024. Does AI-assisted creation of polyphonic music increase academic motivation? The DeepBach graphical model and its use in music education. *Journal of Computer Assisted Learning* (2024).

Appendix

Study #	Title	Year	Tool Used	Key Functionalities
1	Using a Chatbot to Provide Formative Feedback: A Longitudinal Study of Intrinsic Motivation, Cognitive Load, and Learning Performance	2024	Chatbot	• Text generation • Formative feedback
2	Exploring the Use of ChatGPT as a Tool for Learning and Assessment in Undergraduate Computer Science Curriculum: Opportunities and Challenges	2023	ChatGPT	• Text generation • Learning support
3	Leveraging ChatGPT for Enhancing Critical Thinking Skills	2023	ChatGPT	• Enhancing critical thinking through structured tasks
4	Computer Science Education in ChatGPT Era: Experiences from an Experiment in a Programming Course for Novice Programmers	2024	ChatGPT	• Programming assistance for novice learners
5	Scaffolding Computational Thinking With ChatGPT	2024	ChatGPT	• Scaffolding computational thinking
6	Students' Experiences of Using ChatGPT in an Undergraduate Programming Course	2024	ChatGPT	• Student learning feedback in programming courses
7	The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation	2023	ChatGPT	• Improving computational thinking and programming skills
8	The Impact of ChatGPT on Language Learners' Motivation	2023	ChatGPT	• Language learning support • Motivational enhancement
9	Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study	2024	ChatGPT	• Programming mode facilitation • Debugging
10	ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs)	2024	ChatGPT	• Immediate responses for cognitive skill reinforcement
11	The Impact of Large Language Models on Programming Education and Student Learning Outcomes	2024	LLM tools	• Programming education • Debugging
12	Effect of Generative Artificial Intelligence (AI)-based tool utilization and Students' Programming self-efficacy and Computational Thinking skills in JAVA programming course in Nigeria Universities	2024	ChatGPT	• Improving programming self-efficacy • Computational skills

13	Does AI-assisted creation of polyphonic music increase academic motivation? The DeepBach graphical model and its use in music education	2024	AI-music tool (DeepBach)	• Creative support for music composition
14	A mixed-methods evaluation of ChatGPT's real-life implementation in Undergraduate Dental Education	2023	ChatGPT	• Dental education research • Knowledge reinforcement
15	Evaluating the Effectiveness of LLMs in Introductory Computer Science Education: A Semester-Long Field Study	2024	LLM tool (CodeTutor)	• Learning programming language syntax
16	ChatGPT Improves Creative Problem-Solving Performance in University Students: An Experimental Study	2023	ChatGPT	• Creative problem-solving performance enhancement
17	Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study	2024	ChatGPT	• ESL writing skill feedback
18	Short-term learning effect of ChatGPT on pharmacy students' learning	2024	ChatGPT	• Short-term knowledge reinforcement in pharmacy