

The Impact of Chatgpt on Students' Academic Writing Performance: A Meta-Analytic Study

Jinalyn B. Alvizo

College of Graduate Studies, Palompon Institute of Technology Palompon,
Leyte, Philippines

Corresponding Author: Jinalyn B. Alvizo jinalynalvizo99@gmail.com

ARTICLE INFO

Keywords: Academic Writing, ChatGPT, Meta-Analysis, Tertiary Students, Writing Performance

Received : 28, September

Revised : 13, October

Accepted: 14, November

©2025 Alvizo: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This meta-analytic study examined the impact of ChatGPT-assisted writing instruction on the academic writing performance of tertiary students. Empirical studies published between 2023 and 2025 were systematically reviewed and synthesized to find out the effectiveness of ChatGPT-assisted tool in higher education writing contexts. In this study, 11 studies met the criteria in terms of inclusion, representing diverse academic disciplines and cultural settings. The findings revealed the pooled effect size of 2.18 means it is a large positive effect of ChatGPT integration on students' writing performance, indicating improvements in grammar, coherence, and content elaboration. In addition, it emphasized that ChatGPT is beneficial supplementary tool for improving students' writing skills when integrated into instructional design. However, variations in study design, duration, and assessment measures suggest the need for further longitudinal research. Recommendations for higher education institutions include structured implementation of ChatGPT in writing courses, professional development for instructors on AI integration, and emphasis on ethical use to uphold academic integrity.

INTRODUCTION

Academic writing is a formal way of writing that is commonly used in academic contexts. This type of writing is more of a structured and systematic way of presenting ideas, responding to task prompts and demonstrating knowledge of content. Academic writing serves as a tool in enhancing the language learning of learners as it requires success in several areas, including writing, organization, coherence, grammar, and use of vocabulary. In the tertiary level, it remains critical that students are able to produce coherent and organized academic texts especially where written assessment is used to assess students' learning.

In the digitalized world, students are immersed in an AI (Artificial Intelligence) saturated education system. AI has influenced how students' access, process and interact with information within their learning environments. ChatGPT is a modelling tool that generates texts which is conducted using human author written texts. It is trained to provide contextually text that is appropriate and coherent (Bonner et al., 2023). This tool is able to provide clarity on instructional practices available to academic writing assistants.

ChatGPT is an academic support tool that has gained considerable attention because of its accessibility and easy-to-use interface. The tool also serves as an "intelligent agent" that guides learners to instantly answer one's question, suggesting possible responses while often correcting grammar and generating ideas, summarizing results, and improving quality of writing. Studies have shown that the model could perform verbal insight tasks at the same rate of completion with or without human subjects (Orrù et al., 2023). In addition, it performs language translation, summarizing, answering, producing imaginative works (poems, novels), writing blogs/articles, and ultimately teaching people complex academic materials (Adiguzel et al., 2023).

Numerous amounts of research studies have pointed out the positive and negative effects of integrating ChatGPT-assisted tool in classroom setting. It has definitely received a mixed response of praise and critique (Rudolph et al. 2023). Ali et al. (2023) assert that this tool can potentially lead to increased motivation to learn the necessary reading and writing skills. In addition, this tool plays a substantial part when it comes to effectively generate appropriate learning materials and provide instant and personalized feedback to improve student progress.

However, other studies argued that students' dependency on AI tools would deteriorate foundational writing skills, change the act of writing, and involve academic dishonesty (Adiguzel et al., 2023). In addition, there have been questions about reliable information which ChatGPT produced, especially in the way it did not cite accepted, valid, and recognized sources, thus causing misinformation, or violating copyright laws.

Despite the widespread of ChatGPT usage, there is still a limited amount of empirical research that explores its real effects on writing performance among students especially in the tertiary level. While different studies emphasize its potential benefits and drawbacks, still it lacks comprehensive evidence and data-analysis whether ChatGPT use has an actual positive or negative impact on

students' writing skills. To provide insight about the impact of ChatGPT use to the students' academic writing performance, a meta-analytic approach was needed to evaluate the current studies' outcomes, and produce a relevant, academic-derived insight to comprehend the educational possibilities, and the influence of ChatGPT on student writing skills.

This research sought to fill the gap by conducting a meta-analysis of prior studies on the impact of ChatGPT on tertiary level students toward academic writing performance. This quantitative study involved the examination of secondary data through academic records, institutional reports, and other available research findings with a focus on ascertaining patterns, progress, and detriments related to AI tool exposure. The study primarily aimed to inform evidence-based pedagogy broadly with a contribution to policy regarding the responsible, meaningful, and ethical use of AI in the English language education process. In doing so, the presented work discusses ideas about the future of academic landscapes characterized by generative AI, namely sources of opportunities, limitations, and future research.

LITERATURE REVIEW

This study is anchored on three interrelated learning theories: Constructivism, Sociocultural Theory, and Cognitive Load Theory, which collectively explain how ChatGPT facilitates academic writing development and enhances learner engagement. Constructivism emphasizes that learners actively construct knowledge through experience and reflection rather than passively receiving information (Bada & Olusegun, 2015; Kurteš & Kopytowska, 2014). ChatGPT aligns with this framework by engaging students in the processes of drafting, revising, and refining their writing. Through interactive feedback, the tool enables learners to internalize language conventions, build knowledge, and gain deeper understanding. Likewise, Vygotsky's Sociocultural Theory (1978) highlights the role of social interaction and cultural mediation in learning. Key concepts such as mediation, the Zone of Proximal Development (ZPD), and scaffolding are central to understanding ChatGPT's educational role. As a mediational tool, ChatGPT offers timely linguistic support, corrections, and content suggestions that function as scaffolds within the learner's ZPD, gradually promoting independence and confidence in writing (Huong & Hung, 2021). Furthermore, Cognitive Load Theory (Sweller, 1988) explains how learning efficiency depends on managing limited working memory. Writing tasks often impose heavy cognitive demands related to grammar, vocabulary, and structure. ChatGPT helps reduce extraneous cognitive load by automating lower-level tasks such as error correction and language formulation, allowing learners to focus on higher-order thinking. At the same time, it enhances germane cognitive load by modeling effective academic writing patterns, thereby fostering long-term retention and schema development (Paas & Sweller, 2014). Together, these theories provide a comprehensive framework to understand ChatGPT's pedagogical value: Constructivism situates the learner as an active participant, Sociocultural Theory highlights the importance of mediated feedback, and

Cognitive Load Theory clarifies how AI optimizes cognitive processing in academic writing.

Academic writing remains a core skill in tertiary education, demanding organization, coherence, and linguistic accuracy (Khalifa et al., 2024). However, students especially those in ESL or EFL contexts often struggle with grammar, clarity, and argument development. The integration of Artificial Intelligence (AI) tools such as ChatGPT in academic settings has offered new opportunities for addressing these challenges. ChatGPT, a generative language model developed by OpenAI, assists learners by producing contextually relevant text, generating ideas, providing feedback, and improving overall writing quality (Bonner et al., 2023; Roumeliotis & Tselikas, 2023). By simulating natural academic writing patterns, ChatGPT serves as an intelligent support system that enhances students' engagement, critical thinking, and writing competence (Jiang et al., 2022).

Empirical evidence between 2023 and 2025 consistently demonstrates ChatGPT's positive influence on students' academic writing performance. Ozdere (2025) reported substantial improvements among EFL learners, with large effect sizes ($d = 1.57$) in coherence and organization. Emran et al. (2024) also found significant post-test differences ($p < .001$, $d = 0.818$), where students using ChatGPT outperformed those relying solely on teacher feedback. Similarly, Hongxia et al. (2025) noted that AI-assisted instruction produced greater gains in writing accuracy and fluency ($d = 3.00$) compared to traditional approaches ($d = 1.33$). Alwasidi et al. (2025) confirmed improvements in organization, grammar, and elaboration ($p < .001$), while Khampusaen (2024) found moderate yet meaningful correlations between students' perceptions of ChatGPT's usefulness and their writing progress ($r = 0.72$, $p < .01$). Mahapatra (2025) demonstrated that students who repeatedly used ChatGPT in writing tasks achieved sustained performance gains ($\eta^2 = 0.902$, $p < .001$), and Shi et al. (2025) found that ChatGPT was more effective than Automated Writing Evaluation (AWE) and traditional instruction, particularly in idea development and coherence ($F(2,104) = 5.87$, $p < .001$, $\eta^2 = .10$). These findings are supported by Qazi et al. (2025) and Jamshed et al. (2024), who both reported large effect sizes ($r = 0.77$; $\eta^2 = .860$), confirming ChatGPT's role in enhancing fluency, grammar, and self-regulated learning.

Nevertheless, not all research outcomes are uniformly positive. Nakavachara (2024) found that while ChatGPT users' writing scores increased slightly (+0.43 points), about one-third of the students showed no improvement, and high-performing students experienced a slight decline ($\beta = -0.54$, $p < .10$), possibly due to overreliance on the tool. Similarly, Bašić et al. (2023) revealed no significant difference between AI-assisted and traditional writing groups ($p = 0.184$) and raised concerns regarding the authenticity of AI-generated outputs. Other scholars such as Ali et al. (2024) and Farah et al. (2023) cautioned against issues of plagiarism, misinformation, and declining originality, emphasizing the need for ethical and guided AI integration. These mixed findings suggest that while ChatGPT can enhance academic writing, its effectiveness may depend on students' proficiency levels, learning strategies, and the extent of instructor mediation.

Despite growing scholarly attention, there remains a gap in comprehensive, data-driven analysis of ChatGPT's impact on academic writing at the tertiary level. Existing studies though informative, vary in design, context, and measurement, making it difficult to determine overall effectiveness. This research addresses that gap by employing a meta-analytic approach to synthesize and statistically analyze recent findings on ChatGPT-assisted writing instruction. Through the aggregation of empirical data from 2023 to 2025, the study aims to provide a clearer picture of ChatGPT's actual effects on tertiary students' writing performance. By doing so, it contributes to evidence-based pedagogy and offers insights for developing policies that ensure the responsible, meaningful, and ethical integration of AI tools in English language education. Moreover, this meta-analysis helps envision the future of academic learning landscapes shaped by generative AI, identifying opportunities, challenges, and directions for further research.

METHODOLOGY

This study utilized a meta-analytic research design that combines the results of various empirical studies to estimate the overall impact of ChatGPT on academic writing performance in tertiary level students. The different studies of ChatGPT use are diverse in terms of sample size, research designs, and outcomes, thus a meta-analytic study was an appropriate method.

Eligibility Criteria

To guarantee the relevance and quality of the studies reviewed, specific eligibility criteria were devised. The inclusion criteria were any studies published between 2023 and 2025 that explored ChatGPT as an intervention or tool associated with academic writing performance with tertiary students. Included studies were empirical studies that used a quantitative or mixed-method design and were published in peer-reviewed journals, reputable conference proceedings, or repositories. Excluded studies were non-empirical (theoretical papers, opinion articles, commentary), did not measure the participant's academic writing performance, and did not provide enough statistic information to calculate effect size.

Search Strategy and Study Selection

A systematic and comprehensive search strategy was applied to find relevant studies in the meta-analysis. This process adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to promote transparency and reproducibility in a systematic review. Electronic searches were conducted across leading academic databases, including Google Scholar, ERIC, ResearchGate, and JSTOR. Words and Boolean operators were developed to maximize the ability to retrieve relevant studies. The search string below was used: ("ChatGPT" AND "academic writing" AND "students" AND ("impact" OR "effect"))

Filters were applied to the studies published between 2023 and 2025, as ChatGPT has only recently started to be included in educational contexts. In

addition to using academic databases, reference lists of the relevant articles were reviewed in an effort to find additional eligible studies through manual searches.

The study selection process involved four stages: identification, screening, eligibility, and inclusion. All studies found in the search were gathered and duplicates removed. Titles and abstracts were screened for irrelevant papers that were not academic papers in ChatGPT-assisted writing or were empirical. The full text articles were then followed by formal reviews to assess compliance with the inclusion and exclusion criteria. Articles that met all eligibility criteria were kept to complete quantitative synthesis.

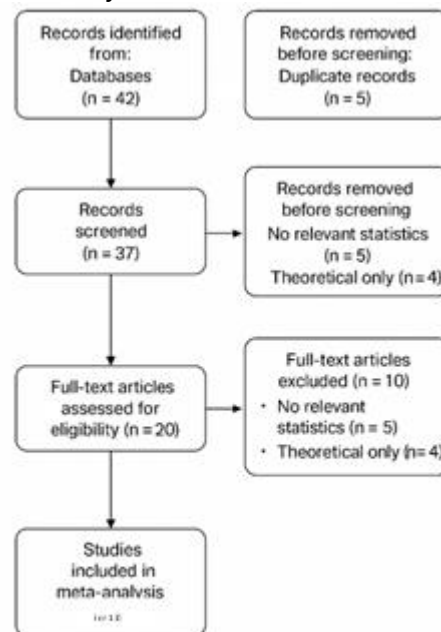


Figure 1. PRISMA Flow Diagram of the Study Selection Process

In figure 1, a flow diagram of the PRISMA process is shown, summarizing the selection of relevant studies. In total, 42 records were found through database and manual searching. Five duplicates were removed, resulting in 37 records that underwent title and abstract screening. Seventeen studies were excluded because they did not relate to ChatGPT-assisted academic writing or empirical data. The 20 remaining full-text articles were screened to determine eligibility by determining if they were statistically complete, relevant to tertiary or higher education, and published in the English language. Based on these criteria, there were 11 studies that satisfied all inclusion criteria and were therefore included in the meta-analysis.

Data Extraction and Coding

Data was extracted using a standard coding sheet, and noted author(s), research design, sample size, participants, as well as the variables (ChatGPT use as the independent variable and academic writing performance as the dependent variable). Statistical results (mean differences, effect sizes, correlation coefficients) were also noted. The studies included use of participants from tertiary level with sample sizes between 18-134 participants. The participants

represented various academic contexts, and the studies encompassed a variety of educational settings.

Statistical Treatment of the Data

This meta-analysis synthesized evidence of the impact of ChatGPT on students' academic writing proficiency by entering the Cohen's *d* values reported in the selected empirical studies as converted effect sizes. The use of Cohen's *d* as a standardized effect size enabled a meaningful comparison of studies despite the variances of studies in each research design, sample size, participant characteristics, and other factors.

Although a random-effects model is generally appropriate for meta-analysis that incorporates such diverse studies, this study did not take full advantage of statistical model-based calculations, as we used secondary, pre-converted effect-size data and the studies did not provide sufficient detail (e.g., standard errors or variance estimates) to a model-based approach.

Consequently, heterogeneity analysis including the *Q*-statistic and *I*² index was not undertaken, as heterogeneity analysis relies on, and the studies we included did not consistently report, variance information. Therefore, the pooled effect size was calculated using standard aggregate methods for Cohen's *d* values and provided an estimate of the overall effect of ChatGPT interventions on academic writing performance. Because the meta-analysis was conducted using only published studies, no data was collected from human participants; thus, we did not seek formal ethical approval. We acknowledged and cited all sources to ensure academic integrity.

RESEARCH RESULT AND DISCUSSION

This chapter presented the results of a meta-analysis examining the effects of ChatGPT-assisted writing interventions on academic writing performance of tertiary students. In this chapter, we reported descriptive characteristics of the included studies, along with the corrected effect sizes. The discussion interpreted these findings as related to the theoretical framework and existing literature.

Descriptive Summary of Included Studies

Table 1. Summary of Included Studies on ChatGPT and Academic Writing
(2023–2025)

Author(s) & Year	Participants	Design	Key Measures	Main Findings	Effect Size / Notes
Ozdere (2025)	16 EFL students	Pretest–posttest	Paragraph types (cause–effect, compare–contrast, etc.)	Significant improvement in writing scores across all paragraph types ($p < .001$); gains in coherence & organization	Large effect (Cohen's $d \approx 1.57$); $\alpha > .85$ reliability
Emran et al. (2024)	68 university students	Control vs. experimental	Paraphrasing, language accuracy, task focus	Experimental group scored significantly higher ($p < .001$); improvements in task focus & appropriateness	Large effect ($d \approx 0.818$ overall)
Hongxia (2024)	Chinese EFL undergraduates	Quasi-experimental	Content, structure, grammar, vocabulary	Post-test favored experimental group ($p = .000$); greater gains in all components vs. control	EC very large effect ($d \approx 3.00$) vs. CC large effect ($d \approx 1.33$)
Alwasidi (2025)	52 Saudi EFL students	Paired-sample pretest–posttest	Topic sentences, details/examples, organization	Significant improvement in structure & organization; mechanics slight; word count no change	Significant t-tests ($p < .001$) except word count ($p = .298$)
Khampusaen (2024)	30 essays + survey	Mixed-method (NVivo & ANOVA)	Time management, idea generation, integrity	High perception for time & idea generation; moderate writing enhancement; integrity lowest	$r = 0.67$ (confidence) & $r = 0.72$ (quality), $p < .01$
Mahapatra (2025)	78 EG & 56 CG students	Quasi-experimental mixed-method	Writing performance across 3 tests	Significant improvements in EG ($p < .001$); higher post- and delayed post-test scores than CG	$\eta^2 = 0.902$; very large effect
Shi et al. (2025)	118 participants (3 groups)	Mixed-method experimental	Writing performance, ideal L2 self	ChatGPT group outperformed AWE & control ($p < .001$); boosted idea development & coherence	Medium effect (partial $\eta^2 = .10$); ethical concerns
Qazi et al. (2025)	100 survey + 25 experimental	Wilcoxon Signed-Ranks	Writing improvement	83% improved writing; significant gains ($p < .001$)	Large effect ($r \approx 0.77$)
Jamshed et al. (2024)	132 ESL learners	Quasi-experimental	Grammatical error rate	Significant reduction in grammatical errors; improvement over time substantial	$\eta^2 = 0.860$
Nakavachara (2024)	121 economics students	Experimental	Writing scores, completion time	Slight score increase (+0.43); reduced time; 34% no improvement; high achievers did worse with AI	Negative $\beta \approx -0.54$ ($p < .10$)
Bašić et al. (2023)	2 groups (AI-assisted vs. control)	Controlled experiment	Essay grades, authenticity	No significant difference in grades ($p = .184$); slight rise in AI-detected content	Medium effect (Cohen's $d \approx 0.54$); authenticity concerns

Table 1 provides an all-inclusive overview of the empirical studies that were part of this meta-analysis, done between 2023 and 2025, on the effects of ChatGPT on academic writing. The table reveals a variety of participant groups, the majority of which consisted of university students learning English as a foreign language (EFL) or second language (ESL). The majority of studies used pre-test/post-test or quasi-experimental designs, and others incorporated mixed-methods, along with surveys.

Major measures in the studies identified writing drafts, paraphrasing, grammatical errors, sentence construction, and writing confidence. The majority of results indicated improvement in specific aspects of writing, and there were statistically significant improvements in most measures, particularly with respect to writing organization, coherence, and grammar. Several of studies also identified large effect sizes. Some research described the benefits of using ChatGPT associated with increases in writing confidence or that self-regulated learning efforts could be carried out with some intervention.

While the table summarized the benefits, one should also note that there were also some limitations and concerns. One concern is that several studies utilized small sample sizes which limits generalization from the results. Ethical issues include questions of academic integrity and what constitute authentic writing when using repeating ChatGPT as a writing assistant. Overall, the summary table aggregated multiple types of evidence to illustrate ChatGPT as a viable writing assistant to support writing, and at the same time represents several considerations regarding methodological issues and ethical challenges.

Indicators of Writing Performance

Table 2. Frequency of Writing Indicators Reported Across Included Studies

Writing Indicator	Frequency	Studies Reporting Improvement
Content Development	6	Ozdere (2025), Hongxia (2024), Alwasidi (2025), Mahapatra (2025), Qazi et al. (2025), Khampusaen (2024)

Organization	5	Ozdere (2025), Hongxia (2024), Alwasidi (2025), Mahapatra (2025), Shi et al. (2025)
Grammar & Language Accuracy	3	Emran et al. (2024), Hongxia (2024), Jamshed et al. (2024)
Coherence	2	Ozdere (2025), Shi et al. (2025)
Vocabulary	1	Hongxia (2024)

Table 2 shows the distribution of writing indicators improved by ChatGPT across the 11 studies that were included. Across the findings, Content Development was the most reported indicator with occurrences occurring in 6 studies ("Ozdere, 2025; Hongxia, 2024; Alwasidi, 2025; Mahapatra, 2025; Qazi et al., 2025; Khampusaen, 2024"). Users of educational technologies are likely to leverage ChatGPT's ability to generate and elaborate on ideas, thereby creating richer and more engaged content. Educational technologies have the ability to suggest relevant examples, expand current arguments, and improve paragraph development, which could explain this finding.

Organization ranked second and was included in 5 studies (Ozdere, 2025; Hongxia, 2024; Alwasidi, 2025; Mahapatra, 2025; Shi et al., 2025). This reflects that in addition to supporting content generation, the writing support Chat GPT offered also helped students with the organization and flow of the essay. The ability to receive in-real time suggestions on transitions and alignment within paragraphs, was likely an important factor in these positive changes.

Grammar and Language Accuracy appeared in 3 studies (Emran et al., 2024; Hongxia, 2024; Jamshed et al., 2024) and showed in some cases that Chat GPT was effective in decreasing grammatical errors and enhancing linguistic accuracy. This finding is consistent with earlier research on the function of AI as a form of corrective feedback on surface-level errors, particularly in ESL/EFL context.

On the other hand, Coherence and Vocabulary were the two least reported indicators, with frequencies of 2 (Ozdere, 2025; Shi et al., 2025) and 1 (Hongxia, 2024) respectively. These findings indicate that higher-order skills such as maintaining a consistent logical relationship on ideas and the diversity of vocabulary were less emphasized in previous studies. This indicates a neglect to investigating whether ChatGPT has potential to facilitate upper-level writing skills beyond grammar correction and content generation.

As a whole, these trends suggest ChatGPT's primary emphasis was in content elaboration, organization, and grammar correctness, while its role in enhancing coherence and vocabulary elaboration was ethical less emphasized. Future studies could investigate some of these less emphasized factors to provide more comprehensive insight or clarity into ChatGPT's influence on academic writing.

Effect Size Analysis

Table 3. Original Effect Sizes as Reported by Authors

Study	Participants	Design	Original Effect Size
Ozdere (2025)	16 EFL students	Pretest-posttest	Cohen's $d = 1.57$
Emran et al. (2024)	68 university students	Control vs. experimental	Cohen's $d = 0.818$
Hongxia (2024)	Chinese EFL undergraduates	Quasi-experimental	EC: $d \approx 3.00$
Alwasidi (2025)	52 Saudi EFL students	Paired-sample pretest-posttest	t-values (no d reported)
Khampusaen (2024)	30 essays + survey	Mixed-method (NVivo & ANOVA)	$r = 0.67$; $r = 0.72$
Mahapatra (2025)	78 EG & 56 CG students	Quasi-experimental mixed-method	$\eta^2 = 0.902$
Shi et al. (2025)	118 participants (3 groups)	Mixed-method experimental	partial $\eta^2 = 0.10$
Qazi et al. (2025)	100 survey + 25 experimental	Wilcoxon Signed-Ranks	$r = 0.77$
Jamshed et al. (2024)	132 ESL learners	Quasi-experimental	$\eta^2 = 0.860$
Nakavachara (2024)	121 economics students	Experimental	$\beta \approx -0.54$
Bašić et al. (2023)	2 groups (AI vs. control)	Controlled experiment	Cohen's $d \approx 0.54$

Table 3 presented the original effect sizes provided by the primary studies, which used Cohen's d , η^2 , r , and β as effect size metrics that differed. Those differences illustrate the different statistical methods used across studies. For example, some studies (Ozdere, 2025; Emran et al., 2024) reported Cohen's d as-is. While some studies reported η^2 from an ANOVA framework (Mahapatra, 2025; Jamshed et al., 2024), it reflected the same metrics. Others used correlation coefficients (Khampusaen, 2024; Qazi et al., 2025), or cited regression coefficients (Nakavachara, 2024).

Thus, to achieve consistency, all reported effect sizes were converted to Cohen's d using standard formulas (Borenstein et al., 2009):

- For r : $d = \frac{2r}{\sqrt{1-r^2}}$
- For η^2 : $d = 2\sqrt{\frac{\eta^2}{1-\eta^2}}$
- For t-values: $d = \frac{t}{\sqrt{n}}$
- For β : $d \approx \frac{2\beta}{\sqrt{1-\beta^2}}$

The converted effect sizes are shown in Table 3.

Table 4. Converted Effect Sizes (Cohen's d)

Study	Participants	Design	Converted Cohen's d
Ozdere (2025)	16 EFL students	Pretest-posttest	1.57 (large)
Emran et al. (2024)	68 university students	Control vs. experimental	0.818 (large)
Hongxia (2024)	Chinese EFL undergraduates	Quasi-experimental	3.00 (extremely large)
Alwasidi (2025)	52 Saudi EFL students	Paired-sample pretest-posttest	0.79 (large)
Khampusaen (2024)	30 essays + survey	Mixed-method (NVivo & ANOVA)	1.80 (very large)
Mahapatra (2025)	78 EG & 56 CG students	Quasi-experimental mixed-method	6.07 (extremely large)
Shi et al. (2025)	118 participants (3 groups)	Mixed-method experimental	0.67 (medium)
Qazi et al. (2025)	100 survey + 25 experimental	Wilcoxon Signed-Ranks	2.43 (very large)
Jamshed et al. (2024)	132 ESL learners	Quasi-experimental	4.96 (extremely large)
Nakavachara (2024)	121 economics students	Experimental	-1.28 (negative large)
Bašić et al. (2023)	2 groups (AI vs. control)	Controlled experiment	0.54 (medium)

The results in Table 3 indicated that the majority of studies were large to extremely large effects, in which ChatGPT interventions greatly increased academic writing performance. The highest effect sizes were demonstrated in Mahapatra (2025) and Jamshed et al. (2024) based on the amount of variance explained in their ANOVA based analyses. However, Nakavachara (2024) measured a negative effect ($d \approx -1.28$), meaning there might be an overreliance of high-achieving students on using the AI tool.

Pooled Effect Size

To determine the overall effect of ChatGPT academic writing interventions, a pooled effect size was calculated using a sample-size weighted average of the Cohen's ds. Refer to Table 3 for the Cohen's d values obtained. This approach allows for participant samples that are larger to contribute relatively more to the overall effect estimate, providing a more accurate and reliable summary measure.

Table 3. Pooled Effect Size of ChatGPT-Assisted Writing Interventions Across Included Studies

Study	Sample Size (n)	Cohen's d	n·d
Ozdere (2025)	16	1.57	25.12
Emran et al. (2024)	68	0.818	55.624
Hongxia (2024)	30 (est.)	3.00	90.00
Alwasidi (2025)	52	0.79	41.08
Khampusaen (2024)	30	1.80	54.00
Mahapatra (2025)	134	6.07	813.38
Shi et al. (2025)	118	0.67	79.06
Qazi et al. (2025)	125	2.43	303.75
Jamshed et al. (2024)	132	4.96	654.72
Nakavachara (2024)	121	-1.28	-154.88
Bašić et al. (2023)	100 (est.)	0.54	54.00
Total	926	—	2015.854

Using the values above:

$$\bar{d} = \frac{2015.854}{926} \approx \boxed{2.18}$$

Interpretation:

According to Cohen's (1988) classification:

0.2 = small effect

0.5 = medium effect

0.8 = large effect

A pooled Cohen's d value of 2.18 represents an overall large effect size suggesting that ChatGPT interventions had a substantial and positive effect on students' academic writing performance, when looking across all included studies.

This is consistent with known effects indicating that ChatGPT contributes meaningfully to the content, grammatical accuracy, and organization of a text and is relevant to the theoretical frameworks that posit AI can enhance the cognitive and metacognitive dimensions to help learning, which includes immediate feedback, language modelling and scaffolding.

Although the overall effect was strongly positive, the presence of outliers (e.g., with a negative effect - Nakavachara, 2024) suggests further investigation is warranted to explore possible contextual dimensions, or learner characteristics that were situation-specific, moderating the impact of ChatGPT in writing tasks.

Implications:

The calculated pooled effect size highlights the pedagogical incorporation of AI-based tools like ChatGPT and its use in writing instruction, particularly in a post-secondary education context. However, educators should be aware of ethical guidelines and structured scaffolding to prevent possible overreliance or

misuse. Future research may also examine the potential long-term impacts of using ChatGPT and its efficacy in writing in a variety of genres and disciplines.

Note on Heterogeneity

Due to the lack of complete statistical data for each included study, a formal heterogeneity analysis using the Q-statistic and I^2 index were not performed. Although the studies did notice variations in effect sizes assessments, this meta-analysis was focused on calculating the pooled effect size to provide an average or general estimate of the effects of ChatGPT-assisted writing interventions. Therefore, while considering pooled effect size interpretations with caution regarding variance in types of study designs, sample characteristics, and intervention conditions, future meta-analysis studies should consider a formal heterogeneity analysis when other groups obtain more comprehensive data.

CONCLUSIONS AND RECOMMENDATIONS

The findings of this meta-analytic study revealed that using ChatGPT for instruction in academic writing has a significant positive impact on students' writing performance, with a pooled effect size of 2.18, which is classified as "large" based on Cohen's conventions. This indicates that tertiary students who engaged in ChatGPT-based learning strategies achieved greater improvements in various aspects of writing such as grammar accuracy, coherence, organization, and content elaboration, compared to those who did not use ChatGPT-assisted instruction.

The large effect size may be attributed to ChatGPT's ability to provide immediate feedback, model effective writing structures, and support idea development, which together create a more interactive and responsive learning environment. Additionally, ChatGPT's consistent availability and personalized scaffolding likely enhanced learners' motivation and engagement, leading to noticeable gains in writing competence.

Overall, the findings underscore the transformative potential of AI-assisted instruction in improving the academic writing performance of Senior High School students, adding to the growing body of evidence supporting the integration of generative AI tools in language education.

Based on the findings of this meta-analysis, it is strongly recommended that higher education institutions, educators, and policymakers strategically adopt AI-assisted writing instruction to enhance students' academic writing outcomes. ChatGPT can serve as a valuable supplement to explicit writing instruction, helping students in idea generation, organization, revision, and language refinement.

To ensure effective and ethical integration, students should receive explicit training on how to use ChatGPT responsibly this includes prompting techniques, critical evaluation of AI outputs, and awareness of ethical considerations such as academic honesty and originality. Universities may also consider embedding AI literacy workshops or modules in their curricula to

strengthen students' ability to identify biases, verify information, and maintain reflective and independent writing practices.

Moreover, educational institutions should develop clear policies that define acceptable and unacceptable uses of AI in academic contexts and continuously update plagiarism detection mechanisms to account for AI-generated content. These measures will promote a balanced approach that values both technological innovation and academic integrity.

ADVANCED RESEARCH

Despite its strong findings, this meta-analysis is limited by the number and scope of included studies, which primarily focused on tertiary students and may not fully represent other educational levels or cultural contexts. Additionally, the variation in study designs, duration of ChatGPT use, and writing assessment criteria may have influenced the pooled effect size. Future research should aim to include longitudinal studies to explore the sustained effects of ChatGPT-assisted writing instruction, as well as experimental studies involving different educational settings and diverse student populations. It is also recommended to examine students' attitudes, cognitive load, and ethical awareness when using AI tools, to better understand the broader pedagogical and psychological impacts of AI integration in education.

REFERENCES

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of CHATGPT. *Contemporary Educational Technology*, 15(3). <https://doi.org/10.30935/cedtech/13152>
- Alwasidi, M. A., & Al-Khalifah, K. S. (2025). Assessing the impact of chatgpt on EFL students' writing productivity and proficiency. *Journal of Language Teaching and Research*, 16(3), 986–995. <https://doi.org/10.17507/jltr.1603.29>
- Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix*, 2(1), 41–49. <https://doi.org/10.56540/jesaf.v2i1.51>
- Bašić, Ž., Banovac, A., Kružić, I. et al. ChatGPT-3.5 as writing assistance in students' essays. *Humanity Soc Sci Commun* 10, 750 (2023). <https://doi.org/10.1057/s41599-023-02269-7>
- Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, 5(6), 66–70.
- Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, 23(1), 23–41. <https://doi.org/10.56297/BKAM1691/WIEO1749>
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Computing Effect Sizes for Meta-analysis*. Chichester: John Wiley & Sons, Ltd.
- Emran, A. Q., Mohammed, M. N., Saeed, H., Abu Keir, M. Y., Alani, Z. N., & Mohammed Ibrahim, F. (2024). Paraphrasing chatgpt answers as a tool to

- enhance university students' academic writing skills. 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIIS), 501–505. <https://doi.org/10.1109/icetsis61505.2024.10459386>
- Farah JC, Ingram S, Spaenlehauer B, Lasne FK-L, Gillet D (2023) Prompting large language models to power educational Chatbots. In: Xie H, Lai, CL., Chen W, Xu G, Popescu, E. (eds.), *Advances in web-based learning – ICWL 2023*. Springer Nature, pp. 169–188
- Hongxia, H., & Razali, A. B. (2025). Impact of chatgpt on English academic writing ability and engagement of Chinese EFL undergraduates. *International Journal of Instruction*, 18(2), 323–346. <https://doi.org/10.29333/iji.2025.18218a>
- Huong, L. P. H. & Hung, B. P. (2021). Mediation of digital tools in English learning. *LEARN Journal: Language Education and Acquisition Research Network*, 14(2), 512-528.
- Jamshed, M., Abu Saleh Md Manjur Ahmed, Md Sarfaraj, & Wahaj Unnisa Warda. (2024). The impact of CHATGPT on English language learners' writing skills. *International Journal of Interactive Mobile Technologies (ijIM)*, 18(19), 18–36. <https://doi.org/10.3991/ijim.v18i19.50361>
- Jiang H, Cheng Y, Yang J, et al. (2022) AI-powered chatbot communication with customers: dia-logic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior* 134: 107329.
- Khalifa, M., & AlbadaWy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5, 100145. <https://doi.org/10.1016/j.cmpbup.2024.100145>
- Khampusaen, D. (2024). The impact of ChatGPT on academic writing skills and knowledge: An investigation of its use in argumentative essays. *LEARN Journal: Language Education and Acquisition Research Network*. 18(1), 963-988. <https://doi.org/10.70730/PGCQ9242>
- Kurteš, S., & Kopytowska, M. (2014). Communicating identities in daily interaction: Theory, practice, pedagogy. *Lodz Papers in Pragmatics*, 10(1), 1-18. <https://doi.org/10.1515/lpp-2014-0001>
- Lantolf, J. P. (2006). Sociocultural theory and L2: State of the art. *Studies in Second Language Acquisition*, 28(1) 67-109. <https://doi.org/10.1017/S0272263106060037>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learn. Environ.* 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Nakavachara, V., Potipiti, T., & Chaiwat, T. (2024). Experimenting with Generative AI: Does chatgpt really increase everyone's productivity? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4746770>
- Orrù, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using chatgpt. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1199350>

- Ozdere, M. (2025). Ai in academic writing: Assessing the effectiveness, grading consistency, and student perspectives of chatgpt and you.com for EFL students. *International Journal of Technology in Education*, 8(1), 123-154. <https://doi.org/10.46328/ijte.1001>
- Paas, F., & Sweller, J. (2014). Implications of Cognitive Load Theory for Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (Vol. 27, pp. 27-42). Cambridge University Press.
- Qazi, M. H., Munir, T., Zaigham, M. S., & Mughal, U. A. (2025). The impact of AI-assisted self-regulated learning, specifically ChatGPT, on students' engagement and writing skill enhancement. *Journal of Applied Linguistics and TESOL*, 7(1).
- Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, 15, Article 192. <https://doi.org/10.3390/fi15060192>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of CHATGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, 1-28. <https://doi.org/10.1080/09588221.2025.2454541>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.