

The impact of ChatGPT on students' academic writing skills: A scoping review



Hamdoni Pangandaman^{1*,A-F}, Mirella Veras^{2,D-F}, Samiel Macalaba^{1,C-F}, Abolbashar Mangontawar^{1,A,B,F}, Sittie Ainah Mai-Alauya^{1,D-F}, Saliha Janine Gubaten^{1,D-F}, Maslainie Aliola^{1,A,B,F}, Norhanie Lininding^{1,E,F}, Paida Abdulmalik^{1,E,F}, Sittie Hainieh Taratingan^{1,E,F}

¹ Mindanao State University, Main Campus, Marawi City, Lanao Del Sur, 9700, Philippines

² Department of Physical Therapy, College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, Canada

*Corresponding author: Hamdoni Pangandaman; Mindanao State University, Main Campus, Marawi City, Lanao Del Sur, 9700, Philippines; email: hamdoni.pangandaman@msumain.edu.ph

Received: 2026-03-23

Accepted: 2026-05-03

Published: 2026-07-01

A – Research concept and design

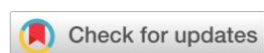
B – Collection and/or assembly of data

C – Data analysis and interpretation

D – Writing the article

E – Critical revision of the article

F – Final approval of article



This is an Open Access article distributed under the terms of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

ABSTRACT

Background: The integration of artificial intelligence (AI) in education has attracted increasing attention, particularly in supporting students' academic writing. ChatGPT, a generative AI tool, has been widely explored for its potential to assist writing processes, including feedback, revision, and language development. However, evidence regarding its effectiveness remains fragmented and varied.

Objectives: This scoping review aimed to map current evidence on the use of ChatGPT in writing-related educational contexts, focusing on its applications, outcomes, and existing research gaps.

Methods: A scoping review was conducted following a structured search across six academic databases (ScienceDirect, ProQuest, Sage Journals, Taylor & Francis, PubMed, and Google Scholar) for studies published between 2020 and 2025. Eligibility was determined using the Population–Concept–Context (PCC) framework. Of 19,251 identified records, nine studies met the inclusion criteria. Data were extracted using a standardized framework and synthesized according to Synthesis Without Meta-Analysis (SWiM) principles.

Results: The included studies examined diverse applications of ChatGPT, including writing assistance, feedback provision, and interactive learning support. Findings indicated that ChatGPT may enhance specific aspects of writing performance, particularly grammar, vocabulary, organization, and coherence. Positive learner-related outcomes, such as increased motivation, engagement, and writing self-efficacy, were also reported. Additionally, immediate feedback and reduced cognitive load emerged as key benefits of the process. However, findings were heterogeneous, with some studies reporting limited or no significant improvement in writing performance.

Conclusions: ChatGPT shows potential as a supplementary tool for supporting writing development, particularly in feedback and revision processes. Nevertheless, current evidence remains limited and heterogeneous. Effective implementation requires instructor guidance, digital literacy, and ethical integration. Further research is needed to establish stronger evidence and develop pedagogical frameworks for AI-assisted writing instruction.

Keywords: academic writing; artificial intelligence, ChatGPT, scoping review, students, writing skills.

How to cite this article: Pangandaman, H., Veras, M., Macalaba, S., Mangontawar, A., Mai-Alauya, S. A., Gubaten, S. J., Aliola, M., Lininding, N., Abdulmalik, P., & Taratingan, S. H. (2026). The impact of ChatGPT on students' academic writing skills: A scoping review. *Journal of Artificial Intelligence in Education & Learning Innovation*, 2(1), 14-31. <https://doi.org/10.56003/jaieli.v2i1.712>

INTRODUCTION

The rapid advancement of artificial intelligence (AI) has transformed multiple sectors, including education, where digital tools are increasingly integrated into teaching and learning processes (Adiguzel et al., 2023; Owan et al., 2023). Among these innovations, ChatGPT (Generative Pre-trained Transformer) has emerged as a prominent AI-powered language model that supports academic writing by generating text, providing feedback, and assisting with idea development. As writing proficiency remains a foundational academic skill across disciplines, there is growing interest in leveraging AI technologies to enhance students' writing performance and learning experiences (Amoush & Alhosban, 2025; Dergaa et al., 2023; Kim et al., 2025; Wang, 2025).

Traditional approaches to writing instruction typically rely on teacher feedback, peer review, and iterative revision processes (Peungcharoenkun & Waluyo, 2023; Yang & Xiaochen, 2022). While effective, these approaches are often time-intensive and may not provide immediate or individualized support. In contrast, AI-assisted tools such as ChatGPT offer real-time feedback, automated grammar correction, and structured writing guidance, potentially transforming how students engage with writing tasks (Amoush & Alhosban, 2025; Li et al., 2024; Wang, 2025). Existing studies suggest that ChatGPT can support various dimensions of writing development, including argument structuring, coherence, vocabulary expansion, and sentence refinement (Herbold et al., 2023; Mahapatra, 2024; Punar Özçelik & Yangın Ekşi, 2024). These affordances indicate that AI may function as a supplementary learning tool that enhances both writing processes and outcomes.

However, despite these promising applications, integrating AI into academic writing raises critical concerns. Scholars have highlighted risks related to over-reliance on AI-generated content, potential erosion of independent writing skills, and challenges to academic integrity and originality (Dergaa et al., 2023; Perkins, 2023; Pratiwi et al., 2025). These tensions underscore the need for a more critical understanding of how AI tools like ChatGPT are used in educational contexts and the implications for student learning.

Recent literature has begun to explore the role of ChatGPT in writing instruction, including empirical studies and emerging reviews that examine its impact on writing performance and learning engagement (Herbold et al., 2023; Pangandaman et al., 2025; Punar Özçelik & Yangın Ekşi, 2024). Existing studies vary widely in terms of population, research design, intervention approaches, and outcome measures. Moreover, prior reviews have largely focused on evaluating effectiveness or synthesizing findings within specific contexts, such as English as a Foreign Language (EFL) settings or higher education broadly, without systematically mapping the scope, characteristics, and gaps of the available evidence base.

However, previous studies have not systematically mapped the breadth of evidence on ChatGPT-assisted writing across diverse educational contexts, study designs, and outcome domains. In particular, there remains limited synthesis that identifies how ChatGPT is implemented, which aspects of writing it most influences, and where gaps persist in the literature. This lack of comprehensive mapping constrains educators' and researchers' ability to understand the full landscape of AI-assisted writing and identify priorities for future investigation.

To address this gap, this study adopts a scoping review to examine the existing literature on ChatGPT and students' writing skills. Rather than determining causal

effectiveness, this review aims to map the extent, range, and nature of research activity, identify key themes and patterns, and highlight areas requiring further empirical investigation. Specifically, this review synthesizes evidence on the use of ChatGPT in writing instruction, the types of writing outcomes reported, and the methodological characteristics of the included studies.

By providing a structured overview of current research, this study seeks to contribute to a clearer understanding of how AI-assisted writing tools are shaping educational practices. The findings are intended to inform educators, curriculum developers, and policymakers on the opportunities and limitations of integrating ChatGPT into writing instruction, while also identifying critical gaps that future research should address to support responsible and effective AI use in education.

METHODS

Study Design

This study employed a scoping review design to map the available evidence on the use of ChatGPT in supporting students’ academic writing skills. A scoping review was considered appropriate because the literature on ChatGPT-assisted writing is still emerging, methodologically diverse, and characterized by variations in study populations, educational contexts, intervention formats, and measured outcomes. Unlike a systematic review, which primarily assesses intervention effectiveness, this scoping review aimed to identify the extent, range, and nature of existing studies, clarify how ChatGPT has been applied in writing-related learning activities, and identify gaps requiring further empirical investigation.

The Population guided the review–Concept–Context (PCC) framework, a commonly used framework for scoping reviews (Table 1). The Population consisted of students engaged in ChatGPT-assisted writing activities; the concept focused on ChatGPT as an AI-supported writing tool; and the context covered educational settings where academic writing, composition, or writing-related learning outcomes were examined. This framework was used to ensure that the review question remained consistent with the broader mapping purpose of the study. The guiding review question was: What is known from existing studies about the use of ChatGPT in supporting students’ academic writing skills across educational contexts?.

Table 1. Description of PCC Framework

PCC Elements	Description
People/ Participants	Undergraduate Students
Concept	Use of Chat-GPT (Generative Pre-trained Transformer) in academic writing
Context	Traditional writing practices

Search Methods

A comprehensive literature search was conducted across seven academic databases to identify relevant studies examining the impact of ChatGPT on students' writing skills. The selected databases included ScienceDirect (867), ProQuest (3,092), Sage Journals (393), Taylor & Francis (522), PubMed (77), and Google Scholar (14,300), covering the publication period from 2020 to 2025 (N = 19,251).

To ensure a systematic and exhaustive search, Boolean operators (AND, OR) were utilized to refine and combine search terms. Quotation marks (") were applied for exact phrase searching, while parentheses () were used to group related concepts. The search terms were carefully selected to capture all relevant literature. They

included the following keywords: ("ChatGPT" OR "Generative Pre-trained Transformer") AND (students OR learners) AND ("writing skills" OR "academic writing" OR "composition skills") AND (use OR application OR integration).

Inclusion and Exclusion Criteria

As aligned with the PCC framework of this study, the population of interest included students engaged in writing-related academic tasks across various educational levels, including secondary, undergraduate, and graduate education. This broader inclusion was adopted to reflect the diversity of the existing literature and to address variations in study populations identified during the preliminary search.

The review focused on ChatGPT as an AI-assisted writing tool. Studies were considered eligible if they examined the use of ChatGPT to support writing-related processes, including grammar correction, vocabulary development, coherence, organization, feedback provision, and overall writing performance. The context encompassed educational settings where writing instruction or writing-related outcomes were evaluated, including academic writing courses, English as a Foreign Language (EFL) environments, and other writing-intensive learning contexts.

In terms of study characteristics, the review included empirical research employing quantitative, qualitative, or mixed-methods designs to allow for a comprehensive mapping of the available evidence. Only peer-reviewed journal articles published in English between 2020 and 2025 were considered eligible to ensure relevance to the emergence of generative AI technologies and to maintain a consistent time frame across the review.

Studies were excluded if they did not involve student populations or were conducted outside educational contexts, such as purely technical or computational studies of artificial intelligence. Articles that did not examine the use of ChatGPT or similar tools in relation to writing-related outcomes were also excluded. In addition, non-empirical publications, including literature reviews, conference proceedings, editorials, commentaries, protocols, theses, and dissertations, were not considered. Studies without accessible full-text versions or those lacking sufficient methodological detail for data extraction were likewise excluded.

Screening of Articles

The screening process was conducted using a systematic and multi-stage approach to ensure methodological rigor and consistency in study selection. All records identified through the database search were initially compiled and organized, and duplicate entries were removed before screening. The remaining articles underwent a two-stage screening process consisting of title and abstract review followed by full-text evaluation.

In the first stage, four independent reviewers (HKP, SPM, AMM, and SAMM) assessed the titles and abstracts of all retrieved studies to determine their relevance using predefined inclusion and exclusion criteria. This initial screening focused on identifying studies that examined the use of ChatGPT in relation to writing-related outcomes within educational contexts. Articles that did not meet the criteria were excluded at this stage. In the second stage, full-text versions of the remaining articles were retrieved and evaluated in detail to confirm their eligibility for inclusion in the review.

To enhance the reliability of the screening process, the reviewers selected independently and then compared their decisions. Inter-reviewer agreement was assessed, and any discrepancies in study selection were resolved through discussion and consensus. When disagreements persisted, an independent panel of reviewers (MV, SJDG, MDA, NAL, PPA, and SHMT) was consulted to provide further evaluation and ensure consistent application of the eligibility criteria. This collaborative process minimized potential selection bias and strengthened the transparency of the review. A detailed PRISMA flow diagram (Figure 1) illustrates the stages of identification, screening, eligibility assessment, and final inclusion of studies. In addition to the diagram, the numerical results of each screening stage are described in the Results section to provide a clear and transparent account of the study selection process.

Data Extraction

A structured data extraction process was undertaken to systematically capture relevant information from all studies included in the review. Following full-text screening and eligibility confirmation, the final set of articles was subjected to detailed data extraction using a standardized extraction framework developed by the research team. This framework was designed to ensure consistency in capturing key study characteristics and to facilitate comparison across studies with varying designs and contexts.

The extracted information included bibliographic details such as author and year of publication, as well as study-specific characteristics including country of origin, research objectives, study design, participant characteristics, and educational setting. In addition, data were collected on the nature of the ChatGPT intervention or application, the duration of implementation (when available), and the specific writing-related outcomes assessed, such as grammar, vocabulary, coherence, organization, academic expression, motivation, and writing self-efficacy.

The extraction process was conducted independently by multiple reviewers to enhance accuracy and reduce bias. Extracted data were cross-checked for consistency, and any discrepancies were resolved through discussion and consensus among the reviewers. The finalized data were organized into a comprehensive summary table to support the synthesis and interpretation of findings (Table 5).

Quality Appraisal and Risk of Bias

Given the methodological diversity of the included studies, a structured and design-specific approach to quality appraisal and risk of bias assessment was employed. As reflected in Table 5, the review incorporated studies with varied designs, including qualitative, quasi-experimental, randomized experimental, longitudinal intervention, and mixed-methods approaches. This heterogeneity necessitated the use of multiple appraisal tools tailored to the methodological characteristics of each study.

For quasi-experimental and longitudinal intervention studies, the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Quasi-Experimental Studies was used to evaluate methodological rigor (Table 2), including participant comparability, intervention integrity, and outcome measurement. Randomized and experimental studies were assessed across key domains of bias, including allocation procedures, blinding, completeness of outcome data, and selective reporting. These assessments were subsequently visualized using the Risk-of-Bias Visualization (ROBVIS) tool to

provide a structured summary of potential bias across studies (Table 3). For mixed-methods studies, the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Mixed Methods was applied to examine the integration of qualitative and quantitative components and the overall coherence of the research design (Table 4). Qualitative studies were appraised using relevant criteria within the JBI framework to assess credibility, dependability, and methodological transparency.

The appraisal process was conducted independently by designated reviewers, and the results were cross-checked to ensure consistency. Any discrepancies in ratings were resolved through discussion and consensus, with input from additional reviewers when necessary. The overall risk of bias for each study was categorized as low, moderate, or high based on the extent to which methodological criteria were satisfied.

Consistent with the purpose of a scoping review, the quality appraisal was undertaken to provide context for interpreting the findings rather than to exclude studies based on methodological limitations. The appraisal results were used to highlight variations in study quality, identify potential sources of bias, and inform interpretation of the evidence base. The assessment revealed variability in methodological rigor across studies, particularly regarding sample size, clarity of intervention procedures, and reporting of outcome measures, reflecting the evolving and heterogeneous nature of research on ChatGPT-assisted writing.

Table 2. Risk of Bias Assessment for Quasi Experiment Design

Author & Year [sample respondents']	JBI assessment tools										Interpretation ^b
	Q1 ^a	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	% Yes	
Alshalan & Alyousef, 2024 (n=38)	Yes	Yes	Unc	N	Yes	U	Yes	Yes	Yes	88%	Low risk of bias
Jamshed et al., 2024 (n=132)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100%	Low risk of bias
Hidayatullah et al., 2024 (n=68)	Yes	Yes	U	N	Yes	U	Yes	Yes	Yes	88%	Low risk of bias
Yanthi et al., 2024 (n=68)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100%	Low risk of bias

Notes:

^a Q1 – Q9 indicate questions 1 to 9 based on the JBI risk assessment

^b The risk of bias was ranked as high when the study reached up to 49% of “yes” scores, moderate when the study reached from 50 to 69% of “yes” scores, and low when the study reached more than 70% of “yes” scores

^c Not means “Unclear”

Table 3. ROBVIS Risk of Bias Tool for RCT

Author (s) & year	Sample size (n)	Allocation concealment	Blinding	Incomplete outcome data	Selective reporting	Other bias	Overall
Avello-Martínez et al., 2024	41	+	?	+	+	?	+
Song et al., 2024	50	+	+	+	+	?	+
Shahsavar et al., 2024	83	+	+	?	+	?	?

Note: (+) indicates a low risk of bias, (-) indicates a high risk of bias, (?) shows unclear risk of bias

Table 4. Joanna Briggs Institute (JBI) Critical Appraisal Tool for Mixed Methods

Author(s) & Year	Q1	Q2	Q3	Q4	Q5	Q6	Q7	% Yes	Overall Interpretation
Malik et al., 2024 (n=10)	Y	Y	U	Y	Y	Y	U	86%	Moderate risk of bias
Huang & Mizumoto, 2024 (n=327)	Y	Y	Y	Y	Y	Y	Y	100%	Low risk of bias

Note: (Y) criterion is fully met, indicating high methodological quality; (U) criterion is partially met or not explicitly stated, leading to some uncertainty; (N) criterion is not met, indicating a potential risk of bias or methodological weakness.

Data Analysis

The data analysis was conducted using a structured narrative synthesis approach guided by the principles of Synthesis Without Meta-Analysis (SWiM) to ensure transparency and systematic organization of findings across heterogeneous study designs. Given the variability in study methodologies, populations, interventions, and outcome measures, a meta-analysis was not feasible. Instead, the SWiM framework was applied to facilitate a consistent and rigorous synthesis of both quantitative and qualitative evidence.

The analysis began by organizing the extracted data into key domains, including study characteristics (author, year, country), research design, participant profile, educational setting, nature of the ChatGPT intervention, duration of implementation, and reported writing-related outcomes. These outcomes included both direct writing performance indicators, such as grammar, vocabulary, coherence, organization, and academic expression, as well as related constructs such as motivation, engagement, and writing self-efficacy.

Studies were then grouped according to similarities in design and outcome focus to allow for meaningful comparison across the evidence base. Rather than aggregating effect sizes, the synthesis focused on identifying patterns, consistencies, and variations in how ChatGPT was applied and in the types of outcomes reported. Particular attention was given to distinguishing among different categories of outcomes, including improvements in writing quality, changes in learner-related factors (e.g., motivation and self-efficacy), and process-related effects (e.g., feedback and cognitive load).

The synthesis also examined contextual factors influencing the reported findings, including differences in educational level, instructional approach, intervention duration, and level of instructor involvement. This approach allowed for a more nuanced interpretation of the evidence, recognizing that the impact of ChatGPT may vary depending on how it is implemented and the learning environment in which it is used.

In addition to identifying patterns, the analysis considered inconsistencies and limitations within the included studies. Variations in study design, sample size, outcome measurement, and reporting practices were critically examined to contextualize the findings and to avoid overgeneralization. This interpretive approach is consistent with the purpose of a scoping review, which is to map the existing body of evidence and highlight areas requiring further investigation rather than to establish definitive conclusions regarding effectiveness.

The synthesized findings were subsequently organized into a summary table to support comparison across studies and to enhance clarity in the presentation of results. This structured analytical approach enabled a comprehensive overview of the current evidence on ChatGPT-assisted writing while maintaining methodological transparency and analytical rigor.

RESULTS

Characteristics of the selected studies

A total of nine studies were included in this scoping review following a multi-stage screening process. Initially, 19,251 records were identified across the selected databases. After applying limiters and removing duplicates, 653 articles remained for title and abstract screening. Of these, 38 studies were assessed for full-text eligibility,

and 9 met the inclusion criteria after applying the predefined selection criteria and incorporating relevant articles identified through reference list screening. The study selection process is illustrated in Figure 1.

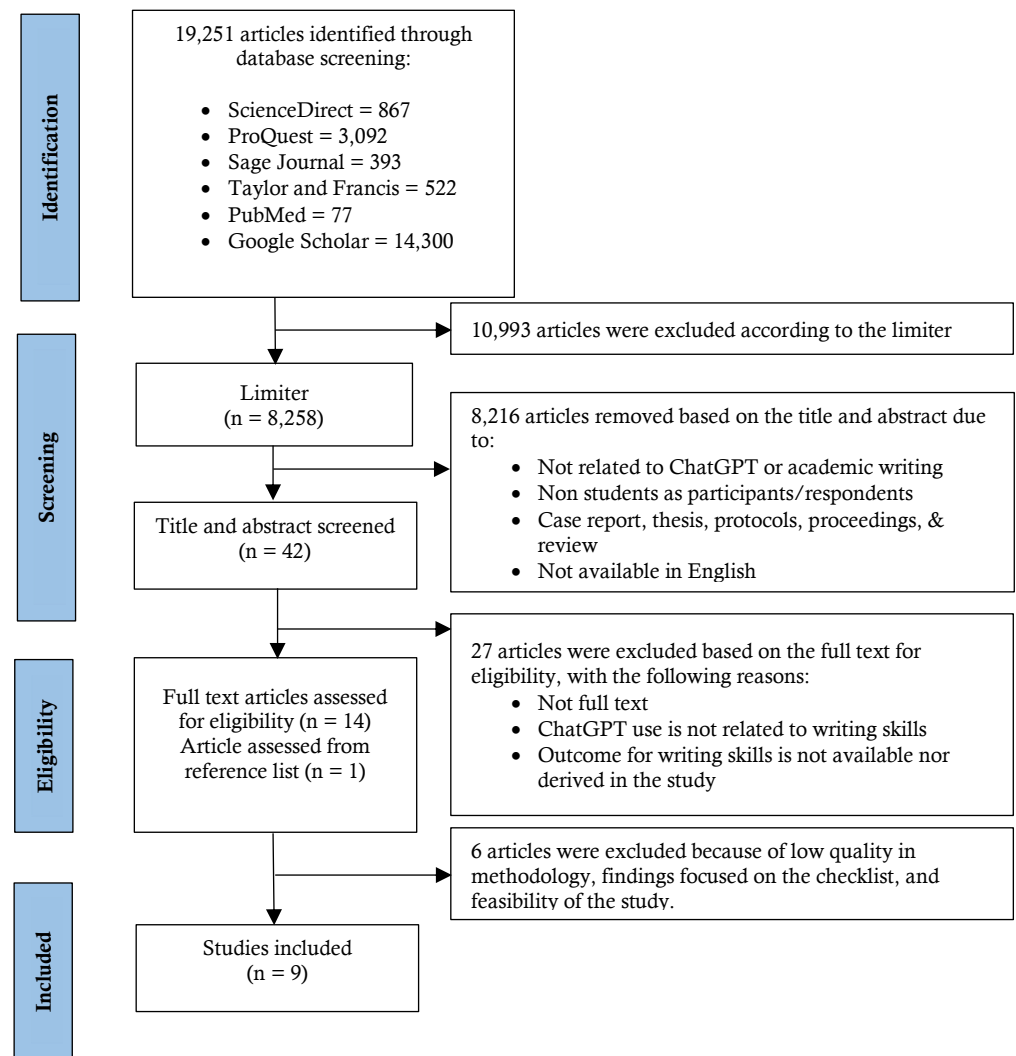


Figure 1. PRISMA flowchart

The included studies represent a geographically diverse evidence base, with research conducted across multiple countries, including Saudi Arabia, China, Ecuador, Indonesia, Iran, Pakistan, and India. Most studies were conducted in higher education contexts, although several included participants from secondary or vocational education settings, reflecting the variability of the populations examined. This diversity aligns with the broader inclusion criteria of the scoping review and highlights the range of educational contexts in which ChatGPT-assisted writing has been explored.

Methodologically, the studies demonstrated notable heterogeneity. The included research comprised quasi-experimental designs, randomized and non-randomized experimental studies, mixed-methods approaches, qualitative investigations, and longitudinal interventions. Sample sizes varied considerably, from small qualitative samples to large cohort-based studies, reflecting differences in study scope and methodological depth. The nature of ChatGPT integration also varied across studies. Several studies implemented structured instructional interventions that embedded

ChatGPT within writing activities, often supplemented with teacher guidance, feedback mechanisms, or digital learning platforms. Other studies examined more specific applications, such as ChatGPT-assisted feedback, mobile-based writing support, or its role in particular writing tasks, including argumentative essays, report writing, and digital storytelling.

The duration of interventions varied across studies, with some reporting short-term interventions lasting several weeks, while others involved longer instructional periods of up to 16 weeks. However, not all studies clearly specified the duration of exposure, reflecting inconsistencies in reporting practices within the current literature.

Participants and Study Settings

The studies included in this review involved participants from a range of educational levels and learning environments, reflecting the heterogeneous nature of the current evidence base. While several studies focused on undergraduate students in higher education institutions, others included participants from secondary and vocational education, as well as graduate-level learners. Sample sizes varied considerably across studies, ranging from small qualitative samples to larger cohort-based investigations. For instance, some studies involved fewer than 50 participants, while others included over 300 students, suggesting differences in study scale and statistical power. In addition, the participants' disciplinary contexts were diverse, including English as a Foreign Language (EFL) courses, medical education, business-related programs, and general academic writing classes. The studies were conducted in predominantly classroom-based settings, including university lecture courses, language classes, and structured writing programs. However, there were also variations in instructional delivery, with some studies incorporating hybrid or technology-mediated learning environments, such as mobile-based feedback systems or AI-integrated learning platforms.

Importantly, the diversity in participant characteristics and study settings introduces several interpretive limitations. Differences in educational level, prior writing proficiency, and digital literacy may influence how students interact with AI tools and the benefits they derive from their use. For example, learners with greater digital familiarity may be better positioned to use ChatGPT effectively. At the same time, those with limited exposure to AI technologies may face challenges interpreting and applying generated feedback. Similarly, the role of instructors varied across studies, with some interventions involving structured guidance and others relying on more autonomous student use.

Applications of ChatGPT in Writing Contexts

The included studies demonstrated a range of ChatGPT applications in writing-related educational contexts, reflecting variability in how the tool is integrated into learning environments. Rather than functioning uniformly as a structured instructional intervention, ChatGPT was used in diverse ways, including as a feedback provider, writing assistant, content generator, and interactive learning support tool.

Several studies incorporated ChatGPT into structured classroom activities, using it to support writing instruction through guided exercises, automated feedback, and iterative revision. In these contexts, ChatGPT was often embedded within broader pedagogical frameworks, such as constructivist learning models or AI-assisted instructional designs, allowing students to engage with writing tasks in a more

interactive and scaffolded manner. Other studies employed ChatGPT in a more targeted manner, focusing on specific functions such as grammar correction, vocabulary enhancement, or organization of written outputs.

In addition to structured use, some studies explored less formal or exploratory applications of ChatGPT, including its role in supporting independent writing, facilitating idea generation, or assisting with research-related writing tasks. These variations indicate that ChatGPT is not limited to a single mode of application but can be adapted to different instructional needs and learning environments.

The duration and intensity of ChatGPT use also varied across studies. While some research implemented short-term activities lasting several weeks, others examined longer-term integration within writing courses. However, inconsistencies in reporting intervention duration and usage frequency were observed, making it difficult to establish clear patterns regarding optimal exposure or implementation strategies.

Importantly, the role of instructors varied across studies, shaping how students used and interpreted ChatGPT. In some cases, instructors actively guided the use of ChatGPT, providing structure and feedback on AI-generated outputs. In other cases, students engaged with the tool more independently, which may have affected how effectively they applied the feedback or suggestions provided. This variation highlights the importance of instructional context in shaping the use and potential benefits of AI-assisted writing tools.

Outcomes of ChatGPT Use in Writing

The included studies reported a range of outcomes associated with ChatGPT use in writing-related contexts. These outcomes can be broadly categorized into three domains: writing performance outcomes, learner-related outcomes, and process-related outcomes. Distinguishing these categories is important, as not all reported effects relate directly to improvements in writing quality, and the findings across studies are not uniformly consistent.

In terms of writing performance outcomes, several studies reported improvements in specific aspects of writing, including grammatical accuracy, vocabulary use, coherence, organization, and overall structure. These improvements were particularly evident in studies where ChatGPT was integrated into structured instructional activities or used as a feedback tool to support revision. However, the extent of improvement varied across studies, and not all writing domains were consistently enhanced. For example, while gains in organization and mechanics were frequently reported, some studies noted limited or uneven effects on higher-order writing skills such as language use or critical argumentation. Moreover, at least one study found no significant improvement in writing performance during digital storytelling tasks, despite observing benefits in other areas. ChatGPT's impact on writing quality may depend on the type of writing task and how the tool is used.

Beyond direct writing outcomes, several studies examined learner-related outcomes, including motivation, engagement, and writing self-efficacy. The findings suggest that ChatGPT may increase student engagement and confidence in writing, particularly when used as an interactive and supportive learning tool. Students reported greater willingness to revise their work and experiment with writing when immediate feedback was available. However, these outcomes represent attitudinal or affective changes rather than direct measures of writing proficiency and should therefore be interpreted separately from performance-based improvements.

Table 5. Summary of Included Studies on ChatGPT Use in Writing Contexts

Study	Design	Sample	Application of ChatGPT	Outcome Domain	Key Effect(s)
Alshalan & Alyousef (2024), Saudi Arabia	Qualitative (corpus-based)	n=38 undergraduate business students	ChatGPT-assisted argumentative writing vs handwritten essays	Writing performance	Shift from descriptive to action-oriented writing; changes in linguistic patterns
Avello-Martinez et al. (2024), Ecuador & Poland	Randomized experimental	n=41 graduate students	ChatGPT in digital storytelling tasks	Writing performance; Process	No significant improvement in storytelling skills; reduced cognitive load
Hidayatullah & Untari (2024), Indonesia	Mixed methods	Vocational high school students	ChatGPT for interactive writing practice	Learner-related; Process	Increased engagement and interactivity in writing activities
Huang & Mizumoto (2024), Japan	Experimental	n=327 undergraduate EFL students	ChatGPT feedback vs peer feedback	Learner-related	Improved motivation and writing self-efficacy
Jamshed et al. (2024), India	Quasi-experimental	n=132 secondary ESL students	ChatGPT mobile-based feedback vs teacher feedback	Writing performance	Reduced grammatical errors; preference for AI feedback
Malik et al., (2024), Multi-country	Qualitative	n=10 researchers	ChatGPT in academic/research writing	Learner-related; Process	Identified benefits (efficiency, creativity) and risks (plagiarism, over-reliance)
Shahsavari et al. (2024), Iran	Longitudinal intervention	n=83 medical students	ChatGPT as writing assistant in academic writing course	Writing performance	Improved content, organization, vocabulary, and mechanics; limited effect on language use
Song & Song (2023), China	Mixed methods	n=50 undergraduate EFL students	ChatGPT-assisted instruction vs traditional	Writing performance; Learner-related	Improved organization, coherence, grammar, vocabulary, and motivation
Yanthi et al. (2024), Indonesia	Mixed methods	n=68 vocational students	ChatGPT-based constructivist learning model	Writing performance; Process	Improved writing skills; challenges in digital literacy and feedback interpretation

The studies also identified process-related outcomes, particularly regarding feedback mechanisms and cognitive load. ChatGPT was frequently described as providing immediate, accessible feedback that facilitated the writing process and supported iterative revision. In some cases, using ChatGPT was associated with reduced cognitive load, allowing students to focus more on content development rather than mechanical aspects of writing. At the same time, concerns were raised regarding the potential for over-reliance on AI-generated content and the risk that students may engage less deeply with the writing process if the tool is used without sufficient guidance.

Taken together, the findings indicate that ChatGPT has the potential to support multiple dimensions of writing-related learning, but the evidence is mixed and context-dependent. While many studies report positive trends, the variability in outcomes, study designs, and implementation approaches limits the ability to draw uniform conclusions. The distinction between writing performance, learner-related factors, and process-level effects is therefore essential for accurately interpreting the current evidence base and avoiding overgeneralization.

DISCUSSION

This scoping review mapped the current body of evidence on the use of ChatGPT in writing-related educational contexts, highlighting patterns in its application, associated outcomes, and gaps in the literature. The findings suggest that ChatGPT is increasingly integrated into writing instruction across diverse educational settings; however, the evidence base is heterogeneous and limited, warranting cautious, analytically grounded interpretation.

A central contribution of this review is the explicit distinction between writing performance outcomes, learner-related outcomes, and process-related outcomes, which are often conflated in existing literature. Regarding writing performance outcomes, several studies reported improvements in grammatical accuracy, vocabulary use, organization, and coherence. For example, studies by [Song & Song \(2023\)](#) and [Shahsavar et al. \(2024\)](#) demonstrated gains in organization, vocabulary, and mechanics, while [Jamshed et al. \(2024\)](#) found reductions in grammatical errors when ChatGPT was used as a feedback tool. These findings are consistent with prior research highlighting the role of AI-assisted feedback in enhancing lower- to mid-level writing skills ([Dong, 2023](#); [Veras et al., 2024](#)).

However, the evidence is not uniform. Some studies reported domain-specific or limited effects. For instance, [Shahsavar et al. \(2024\)](#) noted only minimal improvement in language use, and [Avello-Martínez et al. \(2024\)](#) found no significant improvement in digital storytelling skills despite reductions in cognitive load. These inconsistencies suggest that ChatGPT's effectiveness is influenced by task type, instructional context, and how the tool is integrated, rather than representing a universally effective solution for writing development.

In contrast, learner-related outcomes, including motivation, engagement, and writing self-efficacy, were more consistently positive. Huang and Mizumoto (2024) reported significant improvements in students' motivation and self-efficacy, while [Hidayatullah & Untari \(2024\)](#) and [Yanthi et al. \(2024\)](#) highlighted increased engagement and interactivity in writing activities. These findings align with studies indicating that AI-supported environments can foster learner confidence and encourage iterative writing practices ([Essel et al., 2024](#); [Harahap, 2024](#)). However, these outcomes reflect affective and behavioral changes rather than direct improvements in writing proficiency and should therefore be interpreted as supportive rather than definitive indicators of learning gains.

The review also identified important process-related outcomes, particularly regarding feedback and cognitive load. ChatGPT was frequently described as providing immediate, accessible feedback that facilitates revision and supports the writing process. For example, [Avello-Martínez et al. \(2024\)](#) reported reduced germane cognitive load, suggesting that AI assistance may help students manage the cognitive demands of writing tasks. Similarly, [Malik et al. \(2024\)](#) emphasized the

efficiency and time-saving potential of ChatGPT in academic writing. These findings are consistent with broader literature suggesting that AI tools can enhance writing efficiency and reduce mechanical burden (Abdulkhaled & Pangandaman, 2025; Imran & Almusharraf, 2023; Pratiwi et al., 2025).

At the same time, several studies raised concerns about over-reliance on AI-generated content, superficial engagement, and risks to academic integrity. Malik et al. (2024) highlighted issues related to plagiarism and skill degradation, while Yanthi et al. (2024) identified challenges in interpreting AI-generated feedback, particularly among students with limited digital literacy. These findings reinforce concerns in the literature about the potential unintended consequences of AI-assisted writing (Dergaa et al., 2023; Pangandaman, 2024; Silva et al., 2023), underscoring the need for careful, guided implementation.

A key insight emerging from this review is the role of contextual and moderating factors. Differences in digital literacy influence how effectively students engage with ChatGPT, with more digitally competent learners better able to interpret and apply AI-generated suggestions. The instructor's role is equally critical. Studies that incorporated structured guidance and pedagogical integration tended to report more meaningful outcomes, whereas unguided use was associated with variability in results (Al-Abdullatif & Alsubaie, 2024; Yang et al., 2025). This suggests that ChatGPT functions most effectively as a supplementary pedagogical tool, rather than a standalone solution, and that its impact depends heavily on instructional design and facilitation.

Despite these promising trends, the current evidence base presents several important limitations. The number of included studies is relatively small, limiting the strength of conclusions that can be drawn. Additionally, there is substantial heterogeneity in study populations, including undergraduate, secondary, and graduate learners, as well as variability in research designs, intervention approaches, and outcome measures. This diversity complicates direct comparison and reduces the generalizability of findings. Furthermore, inconsistencies in reporting—particularly regarding intervention duration, frequency of use, and implementation details—limit the ability to identify optimal practices. These limitations highlight the need for more standardized, methodologically rigorous, and context-specific research.

From a theoretical perspective, the findings can be situated within frameworks of technology-enhanced learning, constructivist pedagogy, and self-regulated learning. ChatGPT serves as cognitive scaffolding, supporting students in drafting, revising, and refining their writing through immediate feedback and iterative interaction. This aligns with theories of formative assessment and learner-centered instruction. However, the observed risks of over-reliance suggest a need to balance scaffolding with the development of independent critical thinking and writing skills, pointing to a tension that future theoretical and empirical work should address.

Generally, this review indicates that ChatGPT holds potential as an AI-assisted writing support tool, particularly in enhancing specific aspects of writing and promoting learner engagement. However, its impact is not uniform and should not be overstated. The effectiveness of ChatGPT is conditional, shaped by task type, instructional context, learner characteristics, and implementation strategies. Future research should focus on longitudinal, controlled studies, clearer reporting standards, and the development of pedagogical frameworks to guide the ethical and effective integration of AI into writing instruction.

LIMITATIONS OF THE STUDY

This scoping review is limited by the relatively small number of available studies on the impact of ChatGPT on students' academic writing skills. The heterogeneity of the included studies in terms of research designs, participant characteristics, educational contexts, and ChatGPT implementation approaches limited direct comparisons and the identification of optimal strategies for its use in writing instruction. Additionally, variations and insufficient reporting of learners' prior writing proficiency, digital literacy, and familiarity with artificial intelligence technologies limited a deeper understanding of their influence on writing outcomes. The rapidly evolving nature of generative artificial intelligence may also affect the applicability of current findings over time. Despite these limitations, this review provides a comprehensive overview of the current landscape of ChatGPT-assisted writing in education, identifies emerging patterns in its applications and outcomes, and highlights important gaps for future rigorous and context-specific research.

CONCLUSIONS

The mapped current evidence infers that ChatGPT is commonly used as a writing support tool, particularly for feedback, revision, and language-level improvements. Across the included studies, there is a consistent indication that ChatGPT can support specific aspects of writing performance, such as grammar, vocabulary, organization, and coherence, while also enhancing learner engagement and writing self-efficacy. However, evidence does not show consistent improvement across all areas of writing, especially higher-order skills such as critical thinking and argumentation. The current evidence points to potential benefits rather than definitive effectiveness of ChatGPT in writing instruction.

Moreover, the scientific conclusions of this study have a significant weakness and limited generalizability due to the limited number of studies, mixed participant populations, and differences in research designs and outcome measures. However, in practice, ChatGPT is best used as a complementary tool to support feedback, drafting, and revision processes rather than as a replacement for traditional instruction. Its effectiveness depends on structured integration into teaching practices, including active instructor guidance and students' digital literacy. This study lines its contributions in mapping the range of ChatGPT applications and clearly distinguishing between writing performance, learner-related, and process-related outcomes. Future research should focus on more rigorous and context-specific designs, along with the development of clear pedagogical frameworks to support the ethical and effective use of AI in writing education.

AI DISCLOSURE STATEMENT

During manuscript preparation, the authors used Generative AI, such as ChatGPT and Grammarly (Grammarly Inc.), for language editing, including improving grammar, sentence structure, readability, and the overall clarity of the manuscript. The AI tool was not used to generate scientific content, formulate interpretations, analyze data, synthesize evidence, or draw conclusions. All modifications and suggestions provided by GenAI were critically reviewed, evaluated, and selectively accepted by the authors. The authors maintain full

responsibility for the originality, accuracy, integrity, and final content of this manuscript, including all interpretations, analyses, and conclusions presented.

ACKNOWLEDGMENTS

The authors acknowledge the time and effort of the team members and their immediate superior for the support.

DATA AVAILABILITY

The data supporting the findings of this scoping review are derived from publicly available published studies. Information on the study selection process (PRISMA flow diagram), study characteristics, and extracted outcomes is included in the article. Additional extraction records are available from the corresponding author upon reasonable request.

FUNDING

This research does not receive external funding.

CONFLICT OF INTEREST

The author hereby declares that this research is free from conflicts of interest with any party.

REFERENCES

- Abdulkhaled, S. A. J., & Pangandaman, H. (2025). Computer Literacy and Utilization of Health Information Technology in Nurses Documentation and Plan of Care [Research]. *Journal of Client-Centered Nursing Care*, 11(2), 103-112. <https://doi.org/10.32598/jccnc.11.2.571.6>
- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. <https://doi.org/10.30935/cedtech/13152>
- Al-Abdullatif, A. M., & Alsubaie, M. A. (2024). ChatGPT in Learning: Assessing Students' Use Intentions through the Lens of Perceived Value and the Influence of AI Literacy. *Behavioral Sciences*, 14(9), 845. <https://www.mdpi.com/2076-328X/14/9/845>
- Alshalan, K., & Alyousef, H. (2024). A Corpus-Based Study of the Experiential Meaning in Business Students' Hand-Written and ChatGPT-Generated Argumentative Essays. *Middle East Research Journal of Linguistics and Literature*, 4(5), 93-103. <https://doi.org/10.36348/merjll.2024.v04i05.002>
- Amoush, K. H., & Alhosban, A. A. (2025). Mastering EFL Writing With ChatGPT: A Systematic Review of Benefits, Challenges, and Best Practices. *Theory and Practice in Language Studies*, 15(8), 2576-2583. <https://doi.org/10.17507/tpls.1508.14>

- Avello-Martinez, R., Gajderowicz, T., & Gomez-Rodriguez, V. (2024). Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. *RED. Revista de Educación a Distancia*, 24(8), 1-15. <https://doi.org/10.6018/red.604621>
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. <https://doi.org/10.5114/biol sport.2023.125623>
- Dong, Y. (2023). Revolutionizing Academic English Writing through AI-Powered Pedagogy: Practical Exploration of Teaching Process and Assessment. *Journal of Higher Education Research*, 4, 52. <https://doi.org/10.32629/jher.v4i2.1188>
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (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, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Harahap, D. S. (2024). Implementation of ChatGPT to Improve Students' Critical Thinking Abilities. *Indonesian Journal of Education and Social Humanities*, 1(2), 33-39. <https://doi.org/10.62945/ijesh.v1i2.58>
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hidayatullah, E., & Untari, R. (2024). Exploring Interactivity and Engagement: Improving Writing Skills with ChatGPT for Fun Learning. *Journal of Language, Literature and Teaching*, 5(3), 16-26.
- Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics*, 1, 102324. <https://doi.org/10.29140/dal.v1.102324>
- Imran, M., & Almusharraf, N. M. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15, 1-14. <https://doi.org/10.30935/cedtech/13605>
- Jamshed, M., Manjur, A., S, M., & Warda, W. (2024). The Impact of ChatGPT on English Language Learners' Writing Skills: An Assessment of AI Feedback on Mobile. *International Journal of Interactive Mobile Technologies*, 18(19), 18-36. <https://doi.org/10.3991/ijim.v18i19.50361>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30(1), 1265-1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Li, J., Zong, H., Wu, E., Wu, R., Peng, Z., Zhao, J., Yang, L., Xie, H., & Shen, B. (2024). Exploring the potential of artificial intelligence to enhance the writing of english academic papers by non-native english-speaking medical students

- the educational application of ChatGPT. *BMC Medical Education*, 24(1), 736. <https://doi.org/10.1186/s12909-024-05738-y>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Malik, M. A., Amjad, A. I., Aslam, S., & Fakhrou, A. (2024). Global insights: ChatGPT's influence on academic and research writing, creativity, and plagiarism policies. *Frontiers in Research Metrics and Analytics*, 9, 1486832. <https://doi.org/10.3389/frma.2024.1486832>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Pangandaman, H. K. (2024). Advantages and Challenges in Using AI (Artificial Intelligence) in Research: A Literature Review. *Scope Journal*, 14(2), 1709-1721. <https://doi.org/10.5281/zenodo.18653936>
- Pangandaman, H. K., Hayudini, M. A. A., Tandih, F. C. L., Jupakkal, M. A., Joe, N. M., Abdurasul, J. N. A., Najar, A. A., Tan, E. C., Mercado, C. S., & Amang, D. D. I. J. (2025). Exploring the Role of Large Language Models (LLMs) as an Academic Resource for Students: A Scoping Review. *Library Progress International*, 14(1), 237-243. <https://doi.org/10.5281/zenodo.18728619>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>
- Peungcharoenkun, T., & Waluyo, B. (2023). Implementing process-genre approach, feedback, and technology in L2 writing in higher education. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1), 34. <https://doi.org/10.1186/s40862-023-00211-7>
- Pratiwi, H., Suherman, Hasruddin, & Ridha, M. (2025). Between Shortcut and Ethics: Navigating the Use of Artificial Intelligence in Academic Writing Among Indonesian Doctoral Students. *European Journal of Education*, 60(2), e70083. <https://doi.org/10.1111/ejed.70083>
- Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1), 10. <https://doi.org/10.1186/s40561-024-00296-8>
- Shahsavar, Z., Kafipour, R., Khojasteh, L., & Pakdel, F. (2024). Is artificial intelligence for everyone? Analyzing the role of ChatGPT as a writing assistant for medical students. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1457744>
- Silva, T. P., Ocampo, T. S. C., Alencar-Palha, C., Oliveira-Santos, C., Takeshita, W. M., & Oliveira, M. L. (2023). ChatGPT: a tool for scientific writing or a

- threat to integrity? *British Journal of Radiology*, 96(1152), 20230430.
<https://doi.org/10.1259/bjr.20230430>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14.
<https://doi.org/10.3389/fpsyg.2023.1260843>
- Veras, M., Dyer, J. O., Shannon, H., Bogie, B. J. M., Ronney, M., Sekhon, H., Rutherford, D., Silva, P. G. B., & Kairy, D. (2024). A mixed methods crossover randomized controlled trial exploring the experiences, perceptions, and usability of artificial intelligence in health sciences education. *Digital Health*, 10, 20552076241298485.
<https://doi.org/10.1177/20552076241298485>
- Wang, C. (2025). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning*, 30(3), 1825-1846.
<https://doi.org/10.1007/s10758-024-09744-3>
- Yang, G., & Xiaochen, W. (2022). Towards Understanding Teacher Mentoring, Learner WCF Beliefs, and Learner Revision Practices Through Peer Review Feedback: A Sociocultural Perspective. *Journal of Language and Education*, 8(4). <https://doi.org/10.17323/jle.2022.15962>
- Yang, M., Jiang, S., Li, B., Herman, K., Luo, T., Moots, S. C., & Lovett, N. (2025). Analysing nontraditional students' ChatGPT interaction, engagement, self-efficacy and performance: A mixed-methods approach. *British Journal of Educational Technology*, 56(5), 1973-2000.
<https://doi.org/https://doi.org/10.1111/bjet.13588>
- Yanthi, E. Y., Elmustian, & Charlina. (2024). The Effect of ChatGPT-based Constructivist Learning Model on Observation Report Text Writing Skills. *Journal of Language and Literature Teaching*, 12(4), 13.
<https://doi.org/10.33394/jollt.v12i4.12430>