

Mapping the Impact of ChatGPT on EFL Learners' Writing Skills and Motivation: A Scoping Review

Nesia Oktania¹, Hendi Pratama², Intan Permata Hapsari³
Universitas Negeri Semarang¹²³

nesiaoktania@students.unnes.ac.id, hendipratama@mail.unnes.ac.id, intan74@mail.unnes.ac.id

Abstract

This scoping review synthesizes empirical evidence on the impact of ChatGPT on English as a Foreign Language (EFL) learners' writing skills and motivation across 24 studies published between 2023 and 2025. Using Arksey and O'Malley's scoping review framework, this review examines how ChatGPT-assisted instruction affects writing proficiency outcomes and learning motivation across diverse educational contexts. Findings reveal consistent improvements in writing quality dimensions including grammar, organization, coherence and cohesion, and vocabulary among EFL learners utilizing ChatGPT. Research demonstrates significant positive effects on learner motivation, particularly through enhanced engagement, reduced writing anxiety, and increased intrinsic motivation. However, critical gaps persist regarding long-term sustainability of learning gains, optimal pedagogical implementation strategies, and mechanisms underlying differential effectiveness across learner populations. This review identifies that ChatGPT's effectiveness is substantially mediated by instructional design quality, with structured prompts and teacher-guided implementation producing superior outcomes to unguided exploration. Recommendations emphasize cautious adoption within comprehensive pedagogical frameworks that position ChatGPT as a supplementary tool complementing rather than replacing human instruction.

Keywords: ChatGPT, EFL learners, writing skills, learning motivation, generative AI, language education, scoping review

1. Introduction

In the rapidly evolving field of artificial intelligence (AI), the release of Chat Generative Pre-Trained Transformer (ChatGPT) by OpenAI on November 30, 2022, marked a significant breakthrough (OpenAI, 2022). ChatGPT represents the integration of natural language processing, large language models, and generative AI, showcasing how these technologies can work together to analyze complex linguistic

patterns and produce coherent, human-like text. It is capable of simulating interactive communication and delivering contextually appropriate responses across a wide range of topics (Ray, 2023); (Lund et al., 2023).

With these amazing features, ChatGPT seems to be the most popular platform for learners around the world. As of August 2025, ChatGPT led the global AI chatbot market with approximately 46.6 billion visits, significantly outpacing competitors like

Gemini with 720 million visits and DeepSeek 319 million visits (Fischer, 2025). Similarly, some studies also point out that it is the most recognized and frequently used AI tool by students for multiple purposes in many countries (Al Mashagbeh et al., 2025; Xu, 2025; Deschenes & McMahon, 2024). No wonder that majority of the students in Indonesia are more familiar with this platform (Sarwanti et al., 2024; Martha et al., 2025).

Despite the growing popularity, there is a significant debate on the use of ChatGPT in EFL education. Supporters argue that ChatGPT enhances language proficiency in speaking, listening, reading, grammar, vocabulary, and particularly writing; it boosts motivation, learner autonomy, and engagement (Balci, 2024). Students also report positive effects on their social and emotional learning (SEL) because ChatGPT provides non-judgmental feedback and helps reduce anxiety (Bin-Hady et al., 2024). On the other hand, critics raise concerns about academic integrity, overreliance on AI, potential creativity loss, and errors in accuracy. For instance, students may rely too heavily on AI output without critically engaging with it (Setyaningsih et al., 2025). Teachers also express reservations about ethical issues and about the correctness of content generated by ChatGPT (Balci, 2024). This debate highlights both the promise and possible pitfalls of integrating ChatGPT into EFL settings.

Considering both the growing significance of ChatGPT in EFL education and the ongoing debate over its benefits and risks, studies examining ChatGPT's educational

applications has outpaced systematic synthesis of empirical evidence. This creates substantial uncertainty regarding the nature, magnitude, and sustainability of reported beneficial effects. Previous reviews on ChatGPT typically draw on a limited set of databases and focus either on its role in general EFL education (Alsaedi, 2024; Balci, 2024; Lo et al., 2024; Slamet & Basthomi, 2025) or only on writing skill (Teng, 2024; Le et al., 2025; Mali, 2025; Núñez & De Larreta-Azelain, 2025; Teng, 2024; Truong, 2024; Xiao et al., 2025), leaving some aspects such as motivation underexplored. Writing skills and motivation reflects their prominence in both EFL curriculum objectives and in the extant research literature on ChatGPT's educational applications. Writing proficiency development represents a foundational objective across EFL programs globally while motivation constitutes a critical determinant of language learning success. This scoping review responds to this evidence synthesis need by systematically mapping the research landscape examining ChatGPT's impact on EFL writing skills and learner motivation, identifying methodological patterns, synthesizing key findings, and delineating critical research gaps constraining confident generalization and implementation guidance.

2. Methods

This review employed Arksey and O'Malley's (2005) five-stage scoping review methodology, which involves: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies for inclusion, (4) charting the data, and (5) collating, summarizing, and reporting results. The

scoping review framework was selected as particularly appropriate for mapping the breadth of emerging research on ChatGPT in language education since, according to Tricco et al. (2016), it accommodates heterogeneous research designs and methodologies while identifies thematic patterns and knowledge gaps. Scoping reviews differ from systematic reviews in their emphasis on breadth of coverage rather than depth of quality assessment, making them well-suited to emerging research domains characterized by substantial methodological heterogeneity and rapid knowledge generation (Arksey & O'Malley, 2005).

2.1. Identifying Research Questions

This scoping review was guided by three primary research questions:

1. What is the nature and magnitude of ChatGPT's impact on EFL learners' writing skills and motivation?
2. How do instructional design, implementation approach, and contextual factors moderate ChatGPT's effectiveness for writing skill and motivation outcomes?
3. What critical research gaps exist? This question identifies documented limitations and concerns, and delineates areas where empirical evidence is sparse or absent.

2.2. Identifying Relevant Literature

The review used a broad search strategy to capture emergent evidence and identify research gaps. Searches

were run across different databases such as ERIC (Education Resources Information Center), Scopus, Web of Science, Google Scholar, ScienceDirect (Elsevier), Taylor & Francis Online, SpringerLink, ProQuest, ResearchGate for the period Nov 2022–Oct 2025 using combinations of keywords: “ChatGPT”, “large language model”, “generative AI”, “chatbot”, “EFL/ESL”, “writing”, “essay”, and “motivation”. Since this study seeks to map the breadth of empirical evidence on the use of ChatGPT in developing EFL learners' writing skills and motivation, peer-reviewed empirical studies were prioritized for synthesis. Review studies were retained for background reading and later used in the discussion section to help contextualize findings and compare this review's results with existing syntheses in the field. In addition, only studies published in English between 2022 and 2025 were considered. This timeline was chosen in relation to the emergence of ChatGPT.

2.3. Study Selection

Studies were included if they: (1) examined ChatGPT as a primary focus of investigation, (2) investigated writing skills, writing-related outcomes, or learner motivation (e.g., writing motivation, learning motivation, technology-use motivation) as primary dependent variables, (3) included English as a Foreign Language (EFL) or English as a Second Language (ESL) learners

as primary research participants, (4) reported empirical data from primary research including quantitative outcome measures, qualitative interview data, or mixed-methods approaches, and (5) were published in peer-reviewed journals or completed dissertations between 2023 and 2025. Studies were excluded if they: (1) examined Large language Model other than ChatGPT as the primary focus of investigation, (2) focused on language skills or affective variables (e.g., attitudes, anxiety) without substantive engagement with writing outcomes or motivation, (3) examined ChatGPT applications in non-language educational domains, (4) consisted solely of theoretical commentary or opinion pieces without empirical investigation, , or (5) appeared as conference abstracts without accompanying full-text publication enabling comprehensive data extraction. The inclusion and exclusion criteria are presented in following table

Table 1. Inclusion and Exclusion Criteria

Category	Criteria
Inclusion	1. ChatGPT is the primary focus of investigation 2. Study examines writing skills, writing-related outcomes, and/or

Category	Criteria
Exclusion	1. learner motivation 3. Participants are EFL/ESL learners 4. Empirical research design (quantitative, qualitative, or mixed methods) 5. Published in peer-reviewed journals or completed dissertations (2022–2025). 1. Large Language Model other than ChatGPT is the primary focus of investigation. 2. Study examined other language skills or affective variables without substantive engagement with writing

Category	Criteria
	or motivation
3.	ChatGPT studied in non-language educational domains
4.	Non-empirical commentary or opinion pieces
5.	Conference abstracts without full text.

A total of 1,247 studies were identified from Initial Database Search and 428 of them were removed for duplication. After title and abstract screening, another 736 studies were excluded based on eligibility criteria. After full-text eligibility assessment, 59 studies were excluded, yielding 24 studies which were finally included in the present review.

2.4. Charting the Data

Information from the 25 included studies was charted according to author, year, country, sample size, methodology, study focus, theoretical framework, main findings, and limitations. These data are synthesized and presented in Table 2.

2.5. Collating, Summarizing, and Reporting the Results.

Based on the charted data, the distribution of the studies can be analyzed by geography, learner groups, intervention range and method. This step is crucial to reveal main areas of interest, and consequently where the significant gaps were.

3. Results

Table 2. Data Charting

No	Author (Year) & Country	Sample Size (N)	Method & Focus Area	Intervention	No Author (Year) & Country Framework	Sample Size (N) Main findings	Method & Focus Area Limitation	Intervention
1.	Song & Song, (2023) China	50 University students; intermediate to upper- intermediate proficiency	Mixed methods; Writing skills & motivation	Use ChatGPT feedback in writing (12 weeks)	5. Polakova & Ivenz (2024) Social constructivism Czech Republic	110 university students; varying level of proficiency Improved writing skills and motivation	Quasi- experimental mixed methods; Contextual accuracy concerns, over- reliance on AI writing skill use; small sample size; short duration	Write summary of a text, get ChatGPT feedback, revise summary (3 sessions)
2.	Mahapatra (2024) India	134 (78 EG; 56 CG) University students	Mixed methods; Writing skills	Use ChatGPT for self and feedback assessments in writing 3 genres (1 month)	6. Seo (2024) Formative feedback theory, dialogic feedback, & AWE South Korea	44 university students, (23 female, 21 male), B1- CEFR Significant positive writing improvement; strong student acceptance	Mixed methods, short duration exploratory; Conducted in a single institution; short duration; no long-term retention measure	Use ChatGPT to assist narrative writing (15 weeks)
3.	Alzahrani & Alotaibi (2024) Saudi Arabia	27 High school students; intermediate level of proficiency	Quasi- experimental; Writing skills	Use ChatGPT feedback to write (8 weeks)	7. Mun (2024) Writing Evaluation South Korea	43 (23 EG, 20 CG) university students; high- intermediate level Improved organization, coherence, vocabulary	Mixed methods; Writing skill Grammar accuracy decreased; limited sample size; short duration	Use ChatGPT to proofread and revise writing (7 weeks)
4.	Karataş et al. (2024) Turkey	13 university students (10 female, 3 male); A2-B2 CEFR	Qualitative case study; Writing skills, Motivation & engagement	Integrate ChatGPT-in language-learning activities (4 weeks)	8. Alwasil & Al- Khafar (2025) Saudi Arabia Personalized learning principles; Self-Regulated Learning theory; Cognitive Load Theory (CLT); Engagement Theory	32 female university students; B1- B2 CEFR Positive influence on writing grammar vocabulary, motivation	a one-group pretest- posttest design; Writing productivity & proficiency Randomized experimental (4 experiment group); Homogenous sample; short duration Motivation & metacognitive laziness	Use ChatGPT for writing support in producing essays (9 weeks)
					9. Pan et al. (2025) China	117 university students (70% female) Improved writing motivation		Use ChatGPT feedback for planning, drafting, revising (one day)

598

No	Author (Year) & Country	Sample Size (N)	Method & Focus Area	Intervention	No	Author (Year) & Country	Sample Size (N)	Method & Focus Area	Intervention
19.	Werdiningsih et al. (2024) Indonesia	female); above B1 16 master’s degree students (intermediate-advanced)	informal, neutral Qualitative case study; Writing process	-	21.	Thao et al. (2025) Vietnam	40 university students (21 male, 19 female); low (grammar, vocab, essay structure, idea generation) level of proficiency	Mixed-methods: survey + interviews; Metacognitive in writing	Integrate ChatGPT for writing feedback (15 weeks) Sample size limitations; generalizability issues
20.	Hawanti & Zubayduloevna (2023) Indonesia	73 (36 EG; 37 CG) bachelor students in Purwokerto	Quasi-experimental: pre/post & control group; Writing anxiety	Use ChatGPT to support writing activities	22.	Ly & Linh (2025) Taiwan	40 university students (21 male, 19 female); low (grammar, vocab, essay structure, idea generation) level of proficiency	Overuse decrease critical thinking and increases overreliance; pedagogical change	
21.	Thao et al. (2023) Vietnam	20 EFL students from two universities (10 male; 10 female)	Qualitative interviews, thematic analysis; Writing	-	23.	Huang & Mizumoto (2024) Japan	327 (163 EG; 164 CG) undergraduate students; A2-B2 CEFR	Long term effect; diverse sample;	
22.	Ly & Linh (2025) Taiwan	310 undergraduate students	Quantitative survey; Academic writing	-					
23.	Huang & Mizumoto (2024) Japan	327 (163 EG; 164 CG) undergraduate students; A2-B2 CEFR	Experimental; Writing self-efficacy and motivation	EG: group writing, use ChatGPT with special designed prompts to revise essay outline and paragraph (not to generate essay)					

Based on the table above, several critical patterns can be analyzed in the current state of ChatGPT research for language learning. First, substantial geographic concentration exists in East and South Asia (58.3%), suggesting either genuine regional interest or potential publication biases. Second, the overwhelming focus on university-level learners (79.2%) creates notable research gaps regarding younger learners and postgraduate populations. Third, the dominance of writing-related research (66.7%) indicates potential blind spots regarding speaking, listening, and reading skills. Fourth, the predominance of small samples and relatively short interventions indicates the research field remains exploratory. While mixed-method designs represent 41.7% of studies, reflecting commendable efforts to capture both quantitative and qualitative dimensions, the lack of standardized assessment instruments across studies limits comparative analysis and synthesis. Fifth, writing skills show consistent positive outcomes (92.9%), yet grammar and mechanics demonstrate more mixed results, suggesting ChatGPT's effectiveness is not uniform across linguistic dimensions. Finally, consistent identification of over-reliance risks and limited attention to metacognitive development represent important areas where future research should examine how to structure ChatGPT use to promote learner autonomy and critical engagement rather than mere dependency on AI-generated feedback.

3.1. Research Question 1: The Nature and Magnitude of ChatGPT's

Impact on EFL Learners' Writing Skills and Motivation

Writing Skills Enhancement Across Multiple Dimensions

The scoping review identified substantial improvements in EFL learners' writing proficiency across diverse contexts and proficiency levels. In a high school experimental study, participants demonstrated significant gains in task achievement, with mean scores increasing from 3.0 to 5.2 (Alzahrani & Alotaibi, 2024). Coherence and cohesion improved from 3.2 to 5.0, while lexical resource increased from 3.0 to 4.4, yielding an overall writing skill improvement from 12.3 to 17.4 on the assessment (Alzahrani & Alotaibi, 2024). Notably, ChatGPT's evaluative function demonstrated over 90% agreement with assessments provided by expert English teachers, validating the quality of writing improvements observed (Alzahrani & Alotaibi, 2024).

University-level research similarly corroborated these findings across larger samples. A mixed-methods intervention study with 134 university students (78 experimental, 56 control) implementing ChatGPT for self and feedback assessments across three writing genres over one month showed significant positive writing improvement with strong student acceptance. More specifically, one study involving university ESL students found improvements from pre-test scores of 12.58 to post-test scores

of 19.22, with these gains maintained at 19.18 in delayed post-testing, indicating durability of improvements over time (Mahapatra, 2024). Concurrent research with 110 EFL university students documented quantitative improvements in conciseness, grammar, inclusion of key information, and appropriate use of passive voice (Polakova & Ivenz, 2024).

Component-specific analysis revealed that ChatGPT particularly strengthened grammatical accuracy and vocabulary development. Academic vocabulary, grammar structure, and idea development all showed positive influences (Ly & Linh, 2025). In one case study involving 33 Japanese EFL learners, students expressed strong preferences for AI-assisted editing over traditional peer review processes, appreciating the effective feedback that improved clarity and cohesion of their writing (Huang & Mizumoto, 2024). Students found ChatGPT particularly beneficial for formal register knowledge, with 11 undergraduate students in a pre-experimental design showing active engagement with the tool for writing tasks (Özçelik & Ekşi, 2024). However, one notable anomaly emerged: a study of high school students revealed that grammatical range and accuracy showed a mean score decrease from 3.0 to 2.9 despite overall writing improvement (Alzahrani & Alotaibi, 2024), suggesting potential complexities in how students apply ChatGPT feedback.

Motivation, Self-Efficacy, and Engagement Outcomes

The reviewed literature documented substantial improvements in multiple dimensions of learner motivation and psychological engagement. A quantitative study with 163 students using specially crafted prompts found that ChatGPT positively affected students' motivation related to Ideal L2 Self and L2 Learning Experience (Huang & Mizumoto, 2024). Critically, ChatGPT significantly enhanced writing self-efficacy, which correlated with all three motivational factors examined, though it showed no impact on Ought-to L2 Self motivation (Huang & Mizumoto, 2024). This distinction between intrinsic and extrinsic motivation pathways has important implications for understanding engagement mechanisms.

Interactive practice opportunities promoted heightened motivation and increased engagement in writing tasks through ChatGPT's conversational capabilities (Zhang et al., 2024). Students perceived ChatGPT as a friendly and supportive conversation partner, leading to heightened motivation and increased engagement in writing tasks, with learners feeling more at ease and less anxious when interacting with the tool (Zhang et al., 2024). A qualitative study of 22 Chinese university students documented enhanced writing proficiency and motivation as key outcomes, with students reporting

reduced writing anxiety (Zhang et al., 2024). The review identified that students specifically valued ChatGPT's personalized feedback, guidance, and convenience, with many reporting that ChatGPT assistance made the learning process more enjoyable and less intimidating (Seo, 2024).

Mediation analysis revealed important mechanisms through which ChatGPT operates. In a study of 178 secondary students in the Philippines, the influence of ChatGPT as an educational aid on students' academic performance was fully mediated by learning motivation (Caratiquit & Caratiquit, 2023). This substantial effect size ($f^2=0.6579$) suggested that ChatGPT as an academic support tool directly affected students' learning motivation through offering helpful tools, immediate feedback, and engaging features that encourage students' interest and participation (Caratiquit & Caratiquit, 2023). The implications of this mediation pathway indicate that ChatGPT's impact on performance operates through enhancing students' intrinsic drive to learn rather than through direct knowledge transfer.

3.2. Research Question 2: Instructional Design, Implementation Approach, and Contextual Factors as Moderators of Effectiveness

Intervention Duration and Pedagogical Framework Effects

The empirical studies revealed variation in intervention duration and

framework implementation affecting outcomes. A mixed-methods study spanning 12 weeks with 50 Chinese university students of intermediate to upper-intermediate proficiency using ChatGPT feedback demonstrated improved writing skills and motivation, though researchers identified contextual accuracy concerns, over-reliance risks, and sustainability concerns associated with AI use. In contrast, a more focused one-month quasi-experimental study with 134 university students showed significant positive writing improvement with strong student acceptance. These studies suggest that intermediate-length interventions (4-12 weeks) may provide optimal balance between skill consolidation and research feasibility, though evidence regarding optimal duration remains limited.

The pedagogical framework employed substantially moderated effectiveness. A qualitative case study with 12 EFL teachers and 48 learners implementing a three-phased writing instruction protocol comprising planning, instruction, and assessment phases significantly enhanced both teachers' self-efficacy and learners' writing performance (Teng, 2025). Another qualitative study with 16 Indonesian master's degree students demonstrated that ChatGPT improved writing across grammar, vocabulary, essay structure, and idea generation, supporting educational theories including Scaffolded Writing, Distributed Cognition, and Cognitive Flexibility

Theory. The evidence suggests that structured pedagogical protocols incorporating explicit instructional phases yield stronger implementation success than ad hoc tool integration.

Teacher Involvement and Student Strategy Development

The literature emphasizes that ChatGPT should supplement rather than replace educators, ensuring holistic language education with pedagogical synergy (Alzahrani & Alotaibi, 2024). The qualitative case study of Indonesian master's students revealed that effective ChatGPT use required students to adopt various strategies including critical evaluation, peer consultation, and incorporation of primary research sources, establishing ChatGPT as a versatile educational tool tailored to different learning styles (Werdiningsih et al., 2024). Students demonstrated sophisticated metacognitive awareness when using ChatGPT, crafting effective prompts, applying systematic thinking, and implementing recommendations (Teng, 2025). These findings suggest that learner agency and strategic tool use substantially moderate effectiveness outcomes.

Teacher enthusiasm and acceptance significantly influenced learner outcomes. Structural equation modeling demonstrated that generative AI acceptance and teachers' enthusiasm enhanced learners' well-being and self-efficacy (Esfandiari & Allaf-Akbary, 2025). However, the research also

noted that while ChatGPT-assisted instruction proved beneficial, some concerns emerged regarding potential over-reliance, with learners in 73 bachelor students (36 experimental, 37 control) showing reduced critical thinking abilities and increased technology dependence. This tension between accessibility benefits and autonomous learning development requires careful instructional design.

Feedback Quality, Personalization, and Comparative Effectiveness

The nature and quality of feedback provided substantially moderated outcomes. Personalized feedback from ChatGPT enhanced learners' motivation and confidence, contributing significantly to writing proficiency improvement through immediate, persistent availability (Alzahrani & Alotaibi, 2024). Prompt crafting specificity moderated results significantly: a study involving 163 students utilizing ChatGPT with specially crafted prompts for feedback found stronger results than those using generic approaches (Huang & Mizumoto, 2024). Instruction on appropriate ChatGPT application also affected outcomes, as learners understanding effective prompt crafting and systematic thinking implemented the tool more effectively (Teng, 2025).

However, comparative research revealed important limitations in ChatGPT feedback efficacy. A qualitative case study examining 110

EFL students found that while ChatGPT provides personalized feedback, teacher feedback remains more effective in improving overall writing quality, with ChatGPT potentially unable to account for errors stemming from cultural and L1-specific linguistic influences (Türkoğlu, 2025). This finding suggests a complementary rather than substitutive relationship, with ChatGPT optimally functioning as a supplement providing frequent corrective feedback while human teachers provide culturally aware, contextually sensitive guidance.

Register-Specific Limitations and Contextual Variation

The effectiveness of ChatGPT varied substantially by writing task characteristics. An 11-student pre-experimental study found ChatGPT beneficial for formal register knowledge but perceived it as unnecessary for informal writing, with effectiveness in teaching neutral register remaining questionable. This suggests that contextual writing demands substantially moderate the tool's utility, and learners recognize domain-specific limitations. Similarly, 20 EFL students from two Vietnamese universities reported enhanced writing skills, improved autonomy, and exposure to diverse language uses, yet identified over-reliance risks, accuracy concerns, and ethical issues. These findings indicate that contextual factors including writing genre, learner proficiency, and institutional support

systems substantially influence effectiveness.

3.3. Research Question 3: Critical Research Gaps

Methodological Limitations in Existing Research

The scoping review identified several critical methodological limitations constraining evidence quality and generalizability. Most studies employed small sample sizes, typically ranging from 11 to 134 participants, limiting statistical power and external validity. Specifically, the 11-student pre-experimental study of register learning and the 20-student qualitative interview study from two Vietnamese universities were explicitly noted as having small, homogenous samples that limited generalizability to broader EFL populations. Geographic concentration of research in East and Southeast Asian contexts (China, Indonesia, Vietnam) with minimal representation from African, Latin American, or South Asian regions substantially constrains transferability of findings across diverse linguistic and cultural contexts.

Short intervention durations represent another substantial evidence gap. Reviewed studies lasted from 2 weeks to 12 weeks, insufficient for examining long-term retention, skill sustainability, and autonomous writing ability development. No studies comprehensively examined whether improvements persist beyond intervention periods or whether

learners internalize writing strategies for independent use. The 2-week register study and 1-month genre-focused study were explicitly identified as requiring longitudinal follow-up and longer-term retention measures. This absence of long-term sustainability data prevents confident assertions about whether ChatGPT-assisted improvements represent durable skill acquisition or temporary performance enhancement contingent on tool availability.

Conceptual and Theoretical Underspecification

Research on ChatGPT's impact on EFL writing remains atheoretical or loosely theorized in most cases. While some studies invoked Social Constructivism, Scaffolded Writing approaches, Distributed Cognition, or the Technology Acceptance Model, few developed explicit theoretical frameworks predicting when, how, and through what mechanisms ChatGPT enhances writing and motivation. The mechanisms through which ChatGPT enhances motivation remain incompletely understood, with mediation analysis limited to one study examining how learning motivation mediates ChatGPT's effects on academic performance (Caratiquit & Caratiquit, 2023). This conceptual gap prevents understanding of moderating conditions and boundary conditions for effectiveness.

Majority of the studies in this review focused on students' use of AI tools in

writing, including idea generation, organization and structure, and spelling-grammar in writing (Esfandiari & Allaf-Akbary, 2025). However, research examining interactional metadiscourse in argumentative writing—an important dimension of academic writing proficiency—remains substantially underdeveloped. This content gap leaves unresolved questions about ChatGPT's utility for enhancing more nuanced and coherent written communication beyond surface-level mechanical corrections.

Ethical, Contextual Validity, and Implementation Concerns

Contextual accuracy and cultural appropriateness emerged as documented limitations. Research participants explicitly noted concerns regarding ChatGPT's accuracy in providing culturally sensitive feedback and recognizing non-native speaker perspectives. A comparative study examining teacher versus ChatGPT feedback identified that ChatGPT could not account for errors made due to cultural and L1 aspects of students, with feedback quality appearing culturally mediated (Türkoğlu, 2025). This raises unresolved questions about ChatGPT's performance with diverse learner populations across different linguistic and cultural backgrounds, and whether single-tool efficacy studies conducted in relatively homogenous contexts generalize to multicultural educational settings.

The sustainability and ethical dimensions of ChatGPT use remain insufficiently examined. Multiple reviewed studies identified over-reliance as a concern, with learners potentially dependent on technology for task completion, yet few studies systematically investigated strategies for preventing dependency or cultivating critical distance from the tool. Academic integrity issues related to plagiarism and undetected use of AI-generated content for assessment purposes were identified as concerns but rarely addressed through rigorous investigation. One study of 73 students noted that overuse of ChatGPT may decrease critical thinking abilities and increase over-reliance, yet implementation strategies for managing these risks remained underdeveloped

Outcome Measurement, Validity, and Cross-Skill Neglect

Writing skill operationalization and measurement varied substantially across reviewed studies, limiting meaningful comparison. While some studies employed standardized IELTS frameworks, others used locally developed rubrics with unclear reliability and validity, constraining evidence synthesis. The unexpected finding that grammatical range and accuracy decreased from 3.0 to 2.9 despite overall improvement raises fundamental questions about whether current assessment instruments adequately capture writing complexity

or whether ChatGPT assistance inadvertently facilitates surface-level corrections over deep structural improvement (Alzaharani & Alotaibi, 2024).

Language skill domains beyond writing remain severely under-researched. While writing received substantial empirical attention with 24 studies identified in the reviewed literature, speaking, listening, and reading showed minimal examination (Esfandiari & Allaf-Akbary, 2025). One study examining 80 teachers and students found neutral attitudes toward ChatGPT's effect on developing listening and speaking skills (Ali et al., 2023). This substantial imbalance creates an incomplete picture of ChatGPT's potential across language proficiency domains and prevents understanding of whether writing-specific benefits transfer to other skill areas or whether ChatGPT's utility is narrowly circumscribed to written expression.

Implementation Scalability and Real-World Application Gaps

The practical scalability and implementation feasibility of ChatGPT integration at scale remain entirely unexplored. Virtually all reviewed studies occurred in controlled research settings with explicit research protocols and instructor oversight, yet real-world classroom implementation by teachers without research training or institutional support systems remains unexamined. The resource

requirements, technical infrastructure needs, and training burdens necessary for widespread institutional adoption have not been systematically documented. This implementation-practice gap substantially constrains the applicability of research findings to authentic educational contexts, particularly in under-resourced institutions where ChatGPT might prove most transformative.

4. Discussion

The accumulated evidence from 24 empirical studies reviewed through this scoping review establishes that ChatGPT produces measurable and substantial improvements in EFL learners' writing proficiency across multiple dimensions including task achievement, coherence and cohesion, lexical resource, and grammatical accuracy (Alzahrani & Alotaibi, 2024). Writing scores improved from 12.3 to 17.4 in high school contexts and from 12.58 to 19.22 in university contexts, representing gains that exceed typical improvement trajectories in traditional writing instruction (Mahapatra, 2024). Similarly, the evidence demonstrates significant enhancements in learner motivation, writing self-efficacy, and engagement, particularly regarding intrinsically motivated dimensions such as Ideal L2 Self and L2 Learning Experience (Huang & Mizumoto, 2024). These findings collectively validate ChatGPT's considerable promise for enhancing EFL writing instruction.

However, the magnitude and durability of these improvements depend substantially on implementation factors, with effectiveness substantially moderated by intervention duration, pedagogical integration quality, feedback personalization, structured teacher involvement, and learner characteristics. The comparison between quasi-experimental control conditions reveals important context effects: experimental groups improved from 62.20 to 78.25 while control groups improved from 58.55 to 65.00 (Mun, 2024), indicating that while ChatGPT produces benefits above traditional instruction, the absolute magnitude may be smaller than headline improvements suggest. ChatGPT functions most effectively as a pedagogical supplement providing immediate, personalized feedback while human teachers maintain responsibility for culturally aware guidance, academic integrity oversight, and deep learning facilitation (Kamis, 2024). This complementary relationship differs fundamentally from conceptions of AI as a substitute for human instruction.

The finding that learning motivation fully mediates ChatGPT's effects on academic performance (Caratiquit & Caratiquit, 2023) suggests that ChatGPT's primary impact operates through psychological engagement and intrinsic motivation enhancement rather than through direct knowledge transfer. This distinction carries important implications for

sustainability, as research demonstrates that intrinsically motivated learning typically produces more durable behavioral change than compliance-based motivation. The sustained motivation benefits documented in post-test and delayed post-test measurements suggest some durability, though extended longitudinal follow-up remains absent from the research base. Notably, ChatGPT's null effect on Ought-to L2 Self motivation (Huang & Mizumoto, 2024) indicates that the tool engages personal identity investment rather than external obligation, potentially enhancing motivation stability.

The comparative ineffectiveness of ChatGPT for teaching neutral and informal registers (Özçelik & Ekşi, 2024) alongside its strong performance in formal register domains suggests that writing task demands substantially moderate tool utility. Similarly, the finding that teacher feedback addressing L1-specific errors outperforms generic ChatGPT feedback (Türkoğlu, 2025) indicates that culturally situated knowledge cannot be easily automated, requiring human judgment and intercultural understanding. This recognition that ChatGPT functions best as a supplement rather than substitute has important pedagogical implications for course design, suggesting that educators should strategically allocate ChatGPT to tasks where standardized feedback suffices (grammar correction, vocabulary expansion) while

maintaining human oversight for higher-order concerns including cultural appropriateness, L1 transfer management, and creative expression.

5. Conclusion

This scoping review synthesizing 24 empirical studies establishes that ChatGPT produces measurable improvements in EFL learners' writing proficiency, with writing scores improving from 12.3 to 17.4 in high school contexts and from 12.58 to 19.22 in university contexts (Alzahrani & Alotaibi, 2024; Mahapatra, 2024). The evidence demonstrates significant enhancements in writing self-efficacy and intrinsic motivation, with learning motivation fully mediating ChatGPT's effects on academic performance (Huang & Mizumoto, 2024; Caratiquit & Caratiquit, 2023). However, substantial methodological limitations—including small sample sizes, short intervention durations, geographic concentration in East/Southeast Asia, and absence of longitudinal follow-up—constrain generalizability and long-term durability claims, preventing confident assertions about implementation in diverse contexts without careful consideration of register-specific limitations and cultural validity concerns.

ChatGPT functions most effectively as a pedagogical supplement rather than substitute for human instruction, with effectiveness substantially moderated by intervention duration, pedagogical

framework quality, and teacher involvement (Alzahrani & Alotaibi, 2024). Future research must prioritize longitudinal designs examining skill retention beyond immediate post-testing, employ larger diverse samples across varied geographic contexts, and develop explicit theoretical frameworks predicting when ChatGPT enhances writing and motivation. Practitioners should recognize that effectiveness requires structured instructional design with comprehensive teacher training, while policymakers should require context-specific empirical evidence before mandating large-scale adoption. The field must move beyond experimental pilots to systematic investigation of implementation at scale across diverse contexts to realize ChatGPT's potential as a tool that enhances rather than replaces human pedagogical expertise.

References

- Al Mashagbeh, M., Alsharqawi, M., Tudevdagva, U., & Khasawneh, H. J. (2025). Student engagement with artificial intelligence tools in academia: a survey of Jordanian universities. *Frontiers in Education*, 10:1550147, 1. <https://doi.org/https://doi.org/10.3389/feduc.2025.1550147>
- Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation: *Journal of English Studies in Arabia Felix*, 2(1), 41–49. <https://doi.org/10.56540/jesaf.v2i1.51>
- Alsaedi, N. (2024). ChatGPT and EFL/ESL Writing: A Systematic Review of Advantages and Challenges. *English Language Teaching*, 17(5), 41. <https://doi.org/10.5539/elt.v17n5p41>
- 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>
- Alzahrani, F. K. J., & Alotaibi, H. H. (2024). The Impact of Artificial Intelligence on Enhancing EFL Writing Skills among High School Students. *Journal of Educational and Human Sciences*, 34. <https://doi.org/10.33193/jeahs.34.2024.482>
- Arksey, H., & O'Malley, H. (2005). Theory & Practice This is an electronic version of an article published in Arksey. *The International Journal of Social Research Methodology*, 8(1), 19–32. <http://journalsonline.tandf.co.uk/OpenURLlinktothearticle:http://www.journalsonline.tandf.co.uk/openurl.asp?genre=article&eissn=1464-5300&volume=8&issue=1&spage=19>
- Balcı, Ö. (2024). The Role of ChatGPT in English as a Foreign Language (EFL) Learning and Teaching: A Systematic Review. *International Journal of Current Educational Studies*, 3(1), 66–82. <https://doi.org/10.46328/ijces.107>
- Bin-Hady, W. R. A., Ali, J. K. M., & Alhumari, M. A. (2024). The effect of ChatGPT on EFL students' social and emotional learning. *Journal of Research*

- in Innovative Teaching and Learning*, 17(2), 243–255.
<https://doi.org/10.1108/JRIT-02-2024-0036>
- Caratiquit, K. D., & Caratiquit, L. J. C. (2023). ChatGPT as an academic support tool on the academic performance among students: The mediating role of learning motivation. *Journal of Social, Humanity, and Education*, 4(1), 21–33.
<https://doi.org/10.35912/jshe.v4i1.1558>
- Deschenes, A., & McMahon, M. (2024). A Survey on Student Use of Generative AI Chatbots for Academic Research. *Evidence Based Library and Information Practice*, 19(2), 2–22.
<https://doi.org/10.18438/eblip30512>
- Esfandiari, R., & Allaf-Akbary, O. (2025). Assessing the Impact of Microsoft Copilot and ChatGPT on EFL Learners' Interactional Metadiscourse in Argumentative Writing. *Innoeduca. International Journal of Technology and Educational Innovation*, 11(1), 47–73.
<https://doi.org/10.24310/ijtei.111.2025.20896>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gasevic, D. (2025). Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance. *British Journal of Educational Technology*, 56(2), 489–530.
- Fischer, S. (2025, September 6). *ChatGPT is still by far the most popular AI chatbot*. . Axios.
- Hawanti, S., & Zubaydullovna, K. M. (2023). AI chatbot-based learning: alleviating students' anxiety in english writing classroom. *Bulletin of Social Informatics Theory and Application*, 7(2), 182–192.
<https://doi.org/10.31763/businta.v7i2.659>
- 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>
- Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29(15), 19343–19366.
<https://doi.org/10.1007/s10639-024-12574-6>
- Le, T. T. P., Quach, T. T. N., Pham, T. X. T., & Nguyen, D. L. (2025). A Systematic Review of How ChatGPT is Perceived and Utilized in EFL Writing Classes: Implications for Vietnam. *AsiaCALL Online Journal*, 16(1), 295–311.
<https://doi.org/10.54855/acoj.2516115>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. yung. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies. In *Smart Learning Environments* (Vol. 11, Issue 1). Springer.

- <https://doi.org/10.1186/s40561-024-00342-5>
- Lund, B. D., Wang, T., Reddy Mannuru, N., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a New Academic Reality: AI-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing. *ChatGPT and a New Academic Reality: AI-Written Research Papers and the Ethics OfJournal of the Association for Information Science and Technology*, 74, 570–581.
<https://ssrn.com/abstract=4389887>
- Ly, L. T. K., & Linh, T. H. (2025). THE EFFECTS OF CHATGPT ON ACADEMIC WRITING PROFICIENCY AMONG ENGLISH-MAJORED UNDERGRADUATES AT A UNIVERSITY. *VNU Journal of Foreign Studies*, 41(3), 92–109.
<https://doi.org/10.63023/2525-2445/jfs.ulis.5514>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, 11(1).
<https://doi.org/10.1186/s40561-024-00295-9>
- Mali, Y. C. G. (2025). Exploring the Use of ChatGPT in EFL/ ESL Writing Classrooms: A Systematic Literature Review. In *Journal of Language and Education* (Vol. 11, Issue 2, pp. 137–156). National Research University Higher School of Economics (HSE University).
<https://doi.org/10.17323/jle.2025.21793>
- Martha, A. S. D., Widowati, S., & Rahayu, D. P. (2025). ASSESSING STUDENT READINESS AND PERCEPTIONS OF CHATGPT IN LEARNING: A CASE STUDY IN INDONESIAN HIGHER EDUCATION. *Journal of Information Technology Education: Research*, 24.
<https://doi.org/10.28945/5599>
- Mun, C. (2024). EFL Learners' English Writing Feedback and Their Perception of Using ChatGPT. *STEM Journal*, 25(2), 26–39.
<https://doi.org/10.16875/stem.2024.25.2.26>
- Núñez, R. P., & De Larreta-Azelain, M. D. C. (2025). The impact of using ChatGPT for feedback in EFL writing: a systematic review. *Revista Internacional de Lenguas Extranjeras / International Journal of Foreign Languages*, 23, 95–119.
<https://doi.org/10.17345/rile23.4221>
- OpenAI. (2022). *ChatGPT (Nov 30 version)* (November 30, 2022).
<https://chat.openai.com/>
- Özçelik, P. N., & Ekşi, G. Y. (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>
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1).

- <https://doi.org/10.1080/2331186X.2024.2410101>
- Qu, K., & Wu, X. (2024). ChatGPT as a CALL tool in language education: A study of hedonic motivation adoption models in English learning environments. *Education and Information Technologies*, 29(15), 19471–19503.
<https://doi.org/10.1007/s10639-024-12598-y>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154.
<https://doi.org/10.1016/j.iotcps.2023.04.003>
- Sarwanti, S., Sariasih, Y., Rahmatika, L., Islam, M. M., & Riantina, E. M. (2024). Are They Literate on ChatGPT? University Language Students' Perceptions, Benefits and Challenges in Higher Education Learning. *Online Learning Journal*, 28(3), 105–130.
<https://doi.org/10.24059/olj.v28i3.4599>
- Seo, J. Y. (2024). Exploring the Educational Potential of ChatGPT: AI-Assisted Narrative Writing for EFL College Students. *Language Teaching Research Quarterly*, 43.
<https://doi.org/10.32038/ltrq.2024.43.01>
- Setyaningsih, E., Zainnuri, H., Wahyuni, D. S., & Hariyanti, Y. (2025). EFL Students' Use, Perceptions, and Reliance on ChatGPT for Editing and Proofreading: A Technology Acceptance Model Perspective. *Journal of Languages and Language Teaching*, 13(3), 1367.
<https://doi.org/10.33394/jollt.v13i3.13484>
- Slamet, J., & Basthomi, Y. (2025). Examining the Challenges and Opportunities of ChatGPT in EFL Education: A Systematic Literature Review. *Journal of University Teaching and Learning Practice*, 22(2).
<https://doi.org/10.53761/deezkh88>
- 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>
- Teng, M. F. (2024). A Systematic Review of ChatGPT for English as a Foreign Language Writing: Opportunities, Challenges, and Recommendations. *International Journal of TESOL Studies*, 6(3), 36–57.
<https://doi.org/10.58304/ijts.20240304>
- Teng, M. F. (2025). Metacognitive Awareness and EFL Learners' Perceptions and Experiences in Utilising ChatGPT for Writing Feedback. *European Journal of Education*, 60, e12811.
<https://doi.org/10.1111/ejed.12811>
- Thao, L. T., Hieu, H. H., & Thuy, P. T. (2023). EXPLORING THE IMPACTS OF CHATGPT IN EFL WRITING: STUDENT PERCEPTIONS OF OPPORTUNITIES AND CHALLENGES IN VIETNAMESE HIGHER EDUCATION. *Jurnal Ilmu*

- Keguruan* ISSN, 1, 107–124.
<https://afeksi.id/journal3/index.php/kognisi/index>
- Tricco, A. C., Soobiah, C., Antony, J., Cogo, E., Macdonald, H., Lillie, E., Tran, J., D’Souza, J., Hui, W., Perrier, L., Welch, V., Horsley, T., Straus, S. E., & Kastner, M. (2016). A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. *Journal of Clinical Epidemiology*, 73, 19–28.
<https://doi.org/10.1016/j.jclinepi.2015.08.030>
- Truong, T. V. (2024). Using ChatGPT in EFL writing classrooms: A literature review. *International Journal of Research Publication and Reviews*, 5(12), 2207–2211.
<https://doi.org/10.55248/gengpi.5.1224.3527>
- Türkoğlu, S. (2025). *AI VS HUMAN: A QUASI-EXPERIMENTAL STUDY ON CHATGPT AND TEACHER FEEDBACK FOR ENGLISH AS A FOREIGN LANGUAGE YOUNG LEARNERS’ WRITINGS*. Middle East Technical University.
- Werdiningsih, I., Marzuki, Indrawati, I., Rusdin, D., Ivone, F. M., Basthomi, Y., & Zulfahreza. (2024). Revolutionizing EFL writing: unveiling the strategic use of ChatGPT by Indonesian master’s students. *Cogent Education*, 11(1).
<https://doi.org/10.1080/2331186X.2024.2399431>
- Xiao, F., Zhu, S., & Wen, X. (2025). Exploring the Landscape of Generative AI (ChatGPT)-Powered Writing Instruction in English as a Foreign Language Education: A Scoping Review. In *ECNU Review of Education*. SAGE Publications Ltd.
<https://doi.org/10.1177/20965311241310881>
- Xu, Z. (2025). *Patterns and Purposes: A Cross-Journal Analysis of AI Tool Usage in Academic Writing*.
<http://arxiv.org/abs/2502.00632>
- Yıldız, T. A. (2023). The Impact of ChatGPT on Language Learners' Motivation. *Journal of Teacher Education and Lifelong Learning*, 5(2), 582–597.
- Zhang, J., Abbas, Z., Ali, T., Liu, Q., & Wang, Y. (2024). The Impact of ChatGPT on Students’ Writing Proficiency in Second Language Acquisition: Students’ Perception and Experiences: A Qualitative Analysis. *International Journal of Artificial Intelligence and Machine Learning*, 4(1), 1–9.
<https://doi.org/10.51483/ijaiml.4.1.2024.1-9>
- Zhou, L., & Li, J. (2023). The Impact of ChatGPT on Learning Motivation: A Study Based on Self-Determination Theory. *Education Science and Management*, 1(1), 19–29.
<https://doi.org/10.56578/esm010103>