

7-15-2026

Exploring ChatGPT's Role in Developing Productive Language Skills: A Systematic Review

Milana Abbasova

Khazar University, Baku, Azerbaijan

Eljan Salimli

Khazar University, Baku, Azerbaijan, eljan.salimli@khazar.org

Follow this and additional works at: <https://kjhss.khazar.org/journal>

How to Cite This Article

Abbasova, Milana and Salimli, Eljan (2026) "Exploring ChatGPT's Role in Developing Productive Language Skills: A Systematic Review," *Khazar Journal of Humanities and Social Sciences*: Vol. 29: Iss. 2, Article 8. DOI: <https://doi.org/10.5782/2223-2621.1597>

This Original Study is brought to you for free and open access by Khazar Journal of Humanities and Social Sciences. It has been accepted for inclusion in Khazar Journal of Humanities and Social Sciences by an authorized editor of Khazar Journal of Humanities and Social Sciences.

ORIGINAL STUDY

Exploring ChatGPT's Role in Developing Productive Language Skills: A Systematic Review

Milana Abbasova, Eljan Salimli *

Khazar University, Baku, Azerbaijan

ABSTRACT

Although the interactive features of ChatGPT offer substantial benefits for language education, its integration presents complex pedagogical and ethical challenges. This systematic review critically evaluates the capabilities and limitations of ChatGPT in enhancing expressive language capabilities, specifically speaking and writing. Adhering to PRISMA guidelines, a rigorous search of the ERIC database and a forward snowballing technique via Google Scholar yielded an initial pool of 1,124 records, culminating in the final selection of 22 empirical studies for in-depth analysis. Grounded in Vygotsky's Sociocultural Theory and the SAMR integration model, the findings conceptualize ChatGPT as a virtual "More Knowledgeable Other" that provides dynamic cognitive scaffolding. The synthesis indicates that ChatGPT substantially improves writing proficiency by delivering immediate, personalized feedback, assisting with structural organization, and alleviating academic stress. In conversational contexts, the tool provides a safe, non-judgmental platform that lowers the affective filter and encourages autonomous speaking practice. However, critical engagement with the literature reveals profound limitations. While the AI excels at the structural mechanics of language production, it currently falters at the socio-pragmatic realities of human connection, lacking emotional depth in conversational dialogues. Furthermore, the review identifies critical risks, including AI hallucinations, the degradation of authentic academic voice, and a paradox wherein increased learner autonomy inadvertently leads to AI dependency and "cognitive outsourcing." Ultimately, this paper argues for a paradigm shift in language pedagogy, emphasizing the urgent need to integrate critical AI literacy into curricula so that learners utilize generative tools as temporary scaffolds rather than permanent cognitive crutches.

Keywords: Generative AI, Language learning, Writing, Speaking, Cognitive scaffolding, AI literacy

Received 6 July 2025; revised 28 March 2026; accepted 15 May 2026.
Available online 15 July 2026

* Corresponding author.
E-mail address: eljan.salimli@khazar.org (E. Salimli).

<https://doi.org/10.5782/2223-2621.1597>

2223-2621/© 2026 Published by Khazar Journal of Humanities and Social Sciences. This is an open access article under the CC BY 4.0 Licence (<https://creativecommons.org/licenses/by/4.0/>).

Introduction

Interest in the application of Artificial Intelligence (AI) in education has grown over the last forty years. Facilitated by large language models (LLMs), individuals can effortlessly generate content, receive personalized instruction, and engage in interactive learning (Albadarin et al., 2024). A prominent example of how education has been profoundly transformed by LLMs and AI is evident in the way Generative Pre-trained Transformer (GPT) models have enhanced computational comprehension and language processing. These models can perform a multitude of tasks related to language instruction with high accuracy, such as generating texts, assisting with writing, and formulating questions (Kasneci et al., 2023). One of the most significant developments in this domain is ChatGPT, developed by OpenAI. Launched in November 2022 and expanding rapidly, ChatGPT garnered remarkable attention. By January 2023, it had amassed over 100 million active users (Williams, 2023), establishing a historical record as one of the fastest-growing AI tools.

In the context of language teaching and learning, ChatGPT has proven to be highly effective. Utilizing GPT technology, it constructs interactive and personalized learning experiences, which differ substantially from traditional pedagogical approaches (Teng, 2024a). The extensive training data underlying ChatGPT enables it to support meaningful and accurate applications across various educational purposes (Ali et al., 2024). These capabilities underscore the role of AI in transforming education and enhancing instructional methodologies.

However, the intersection of AI and education prompts significant inquiries regarding the rationale behind its utilization in educational settings. It is essential to consider how AI is implemented and integrated into educational environments, as well as the various levels at which it functions – whether on an individual, group, or global scale (Imran & Almusharraf, 2023). Moreover, debates persist concerning the application of ChatGPT within educational contexts. Prominent issues encompass risks associated with privacy and ethics, the proliferation of misinformation or inaccurate content, and the potential for biased responses. Furthermore, concerns regarding the lack of creativity and originality in AI-generated responses have also emerged (Sok & Heng, 2024).

Given the transformative potential of AI and the concerns regarding the use of ChatGPT, the objective of this paper is to investigate whether ChatGPT can enhance productive language skills, namely, speaking and writing. Additionally, this study examines whether the tool encounters limitations while performing these functions. By investigating how this AI tool assists language learners, this review aims to provide valuable insights for both educators and students. These insights align with the objectives of this article and form the basis for the following research questions:

- 1) How effective is ChatGPT in enhancing speaking and writing skills?
- 2) Does ChatGPT face any limitations in enhancing these skills?

Literature review

ChatGPT represents a pivotal AI resource within the field of education. Students frequently utilize ChatGPT as a learning assistant for various tasks, such as writing, checking grammar, and personalizing learning experiences. Educators, however, rely on it for lesson planning and the development of teaching materials (Albadarin et al., 2024). Abdullayeva and Musayeva (2023) illustrated the adaptability of ChatGPT as an educational resource and highlighted the necessity for additional empirical support to verify its long-term efficacy.

Numerous researchers across various countries have demonstrated substantial interest in exploring the application of ChatGPT in education and language learning. For instance, [Song and Song \(2023\)](#) investigated the influence of AI-assisted instruction on Chinese English as a Foreign Language (EFL) students. Their findings indicated notable enhancements in writing skills, motivation, and proficiency concerning organization, coherence, grammar, and vocabulary. Although these advancements are promising, concerns regarding contextual accuracy and an overreliance on AI tools persist. Consequently, [Song and Song \(2023\)](#) emphasized the necessity for the adaptation and development of AI tools to maintain their transformative potential in writing instruction.

More recently, [Mahapatra \(2024\)](#) examined the influence of ChatGPT as a formative feedback tool on undergraduate English as a Second Language (ESL) students' academic writing skills. The study revealed substantial gains in writing clarity, organization, and grammatical accuracy. The author also highlighted ChatGPT's potential as an innovative tool in large-class settings; however, its effectiveness may vary across different contexts, particularly in Global South regions. This variation raises crucial questions about the scalability of such technologies, as they may not universally benefit all educational environments. In a similar vein, [Teng \(2024b\)](#) adopted a mixed-methods approach to evaluate the experiences of EFL learners utilizing ChatGPT during a semester-long writing course in Macau. The findings indicated notable advances in learners' motivation, self-efficacy, engagement, and collaborative tendencies.

[Ghafouri et al. \(2024\)](#) conducted a critical investigation into the implementation of ChatGPT within a structured writing instruction program in Iran. Their research focused on the ChatGPT-based Writing Instruction Protocol (CGWIP), which encompassed distinct phases of planning, instruction, and assessment. By engaging both educators and students in collaborative writing endeavors, the 10-week program markedly boosted teacher self-efficacy and enhanced learners' writing abilities. The researchers demonstrated that ChatGPT could assist teachers in brainstorming, revising, providing feedback, and assessing students' written work.

The role of ChatGPT in delivering feedback on argumentative essays authored by Chinese EFL learners was explored by [Guo and Wang \(2024\)](#). [Allen and Mizumoto \(2024\)](#) conducted comparable research regarding Japanese EFL learners' preferences for utilizing ChatGPT versus peer writing groups for editing and proofreading academic assignments. The study revealed a marked preference for ChatGPT, as learners perceived it to enhance the clarity and cohesion of their writing. While these findings underscored ChatGPT's effectiveness in improving specific aspects of writing, they also highlighted the complementary value of peer interactions. Thus, this dual approach aligns with a growing consensus on the necessity of balancing AI tools with traditional collaborative methods to foster comprehensive writing development.

The implications of these studies extend beyond mere enhancements in expressive language capabilities, as they raise fundamental questions about the future of education in an increasingly digital landscape. Although a substantial body of literature is available regarding the role of ChatGPT in language education, to the best of the authors' knowledge, the present paper constitutes the first systematic review of existing studies examining the extent to which ChatGPT facilitates the development of active language output.

Theoretical framework

To systematically investigate the role of ChatGPT in developing oral and written competencies, this review is grounded in two foundational frameworks: [Vygotsky's \(1978\)](#)

Sociocultural Theory and the SAMR (Substitution, Augmentation, Modification, and Re-definition) technology integration model (Puentedura, 2009, 2020).

Vygotsky's (1978) Sociocultural Theory highlights the significance of social interaction and scaffolding in the cognitive development of language learners. Central to this paradigm is the Zone of Proximal Development (ZPD), which represents the distance between what individuals can achieve independently and what they can accomplish with guided support. In the context of AI-assisted language learning, ChatGPT assumes the role of a virtual "More Knowledgeable Other" (MKO). By delivering instantaneous feedback, generating structured outlines, and simulating interactive dialogues, the tool provides the necessary linguistic scaffolding to help learners navigate their ZPD. This personalized support system mitigates writing and speaking anxiety, allowing students to reach higher levels of communicative competence before gradually transitioning toward greater learner autonomy.

Furthermore, to conceptualize how this AI tool is implemented within educational environments, this study draws upon the SAMR model developed by Puentedura (2009, 2020). The SAMR framework categorizes technology integration into two distinct phases: enhancement (Substitution and Augmentation) and transformation (Modification and Re-definition). While ChatGPT can be utilized merely to substitute traditional pedagogical methods, such as functioning as a basic spelling and grammar checker, its profound impact on language education resides in the transformation phase. At the Modification and Redefinition levels, ChatGPT facilitates instructional tasks that were previously inconceivable in traditional classrooms. These include providing on-demand, non-judgmental conversational environments for speaking practice and establishing highly individualized, real-time feedback loops during the iterative writing process.

Methodology

This section describes the systematic procedure followed for reviewing the contribution of AI-assisted language learning and ChatGPT to the development of speaking and writing skills. This study followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) to conduct the systematic review process and make the study transparent. As illustrated in these guidelines, these principles involve distinct phases, including identification, screening, eligibility, and inclusion.

Literature search and identification

The analysis of the relevant studies related to the use of ChatGPT in improving productive language skills was conducted by following a strict search strategy. Two major databases, namely ERIC and Google Scholar, were chosen for their broad coverage of numerous research articles. Keywords for this systematic search were defined using a Boolean search strategy. The search was conducted using the expression ("*ChatGPT*" OR "*AI*") AND ("*improving*" OR "*enhancing*" OR "*developing*") AND ("*speaking*" OR "*writing*" OR "*productive language skills*" OR "*foreign language skills*" OR "*L2 skills*").

This approach was applied to the ERIC database to ensure a consistent and comprehensive retrieval of relevant studies focusing on the role of ChatGPT in fostering active communication abilities. Google Scholar, on the other hand, was searched through a forward snowballing technique to systematically screen the extensive pool of papers citing Ali et al. (2023). This specific article was selected as the seed paper because it represents one of the pioneering empirical investigations into ChatGPT's initial motivational impact on L2 skills in early 2023. Serving as a foundational chronological anchor, tracking its rapid

accumulation of citations allowed us to systematically trace the evolution of subsequent research focusing explicitly on how generative AI impacts expressive language acquisition.

Screening and exclusion criteria

Following the initial retrieval from the databases, the preliminary sample comprised 1,124 papers – 379 from the primary ERIC search and 745 identified via Google Scholar citation tracking. During the initial identification process, 8 duplicate publications appearing in both databases were identified and subsequently removed, yielding a total of 1,116 unique records for screening. Duplicates were systematically eliminated by cross-referencing the titles and authors of the papers to prevent the inclusion of identical articles.

During the primary screening phase (title and abstract screening), these 1,116 records were evaluated, leading to the rapid exclusion of 885 records that clearly violated the predefined criteria. First, papers were excluded due to receptive skill focus, meaning they investigated how generative AI impacts reading comprehension, listening skills, vocabulary memorization, or localized grammar mechanics rather than the active development of target second language productive language skills like holistic speaking and writing proficiency. Second, papers were excluded due to tool scope misalignment because they examined non-generative artificial intelligence, automatic writing evaluation software like Grammarly, traditional intelligent tutoring systems, or automated essay scoring tools rather than LLMs or conversational generative tools like ChatGPT. Third, papers were excluded due to population misalignment because they evaluated text generation within native first-language contexts, such as native English-speaking university students writing standard academic essays, completely lacking an EFL, ESL, or second language acquisition framework. Fourth, papers were excluded due to thematic misalignment because they addressed institutional administrative policies, plagiarism detection software accuracy, academic integrity guidelines, or generic philosophical frameworks regarding artificial intelligence governance in higher education without evaluating actual language learning outcomes or student pedagogical interactions.

The remaining 231 reports were then sought for retrieval to undergo a more detailed evaluation. However, 87 of these reports could not be retrieved due to text unavailability or restricted paywalls.

The remaining articles were evaluated against predefined inclusion and exclusion criteria to ensure the selection of relevant, formally published academic literature. These criteria are outlined in [Table 1](#). It is important to note that the exclusion of non-English articles was not a reflection of their quality, but rather a practical limitation to ensure accurate comprehension and reliable synthesis of the data by the researchers. Furthermore, to maintain a standard of academic rigor, the selection was restricted to formally published academic formats – specifically encompassing journal articles, book chapters, books, and conference proceedings. Informal or unauthenticated sources, such as blogs and preprints, were excluded.

Eligibility assessment

Of the 144 fully screened articles, 122 were excluded because they lacked relevance to the topic, did not focus on productive skills, examined alternative AI tools (such as Gliglish and Jenni), or failed to meet the language and timeline constraints. Following a rigorous examination of the full texts, a final selection of 22 articles was made to align with the research objectives. This entire process is visually illustrated in [Fig. 1](#).

Table 1. Inclusion and exclusion criteria.

Criteria	Eligibility	Exclusion
Duplication	One publication from any selected database was included.	Duplicate publications from both databases were excluded.
Timeline	Publications after November 2022 were included.	Publications before November 2022 were excluded.
Language	Articles published in English were included.	Non-English articles were excluded.
Literature Type	Journal articles, book chapters, books, and conference papers were included.	Blogs, preprints, editorials, and unauthenticated essays were excluded.
Availability	Full-text, open-access publications were included.	Studies where the full text was unavailable or behind a restricted paywall were excluded.
Quality	Works that have undergone a formal peer-review process were included.	Non-peer-reviewed works, including anecdotal evidence or opinion pieces, were excluded.
Relevance	Works focusing on productive L2/EFL skills (speaking/writing), using conversational generative AI, and evaluating actual pedagogical outcomes were included.	Works focusing on receptive skills, non-generative tools (e.g., Grammarly), native L1 contexts, or non-pedagogical topics (e.g., plagiarism detection, institutional policy) were excluded.

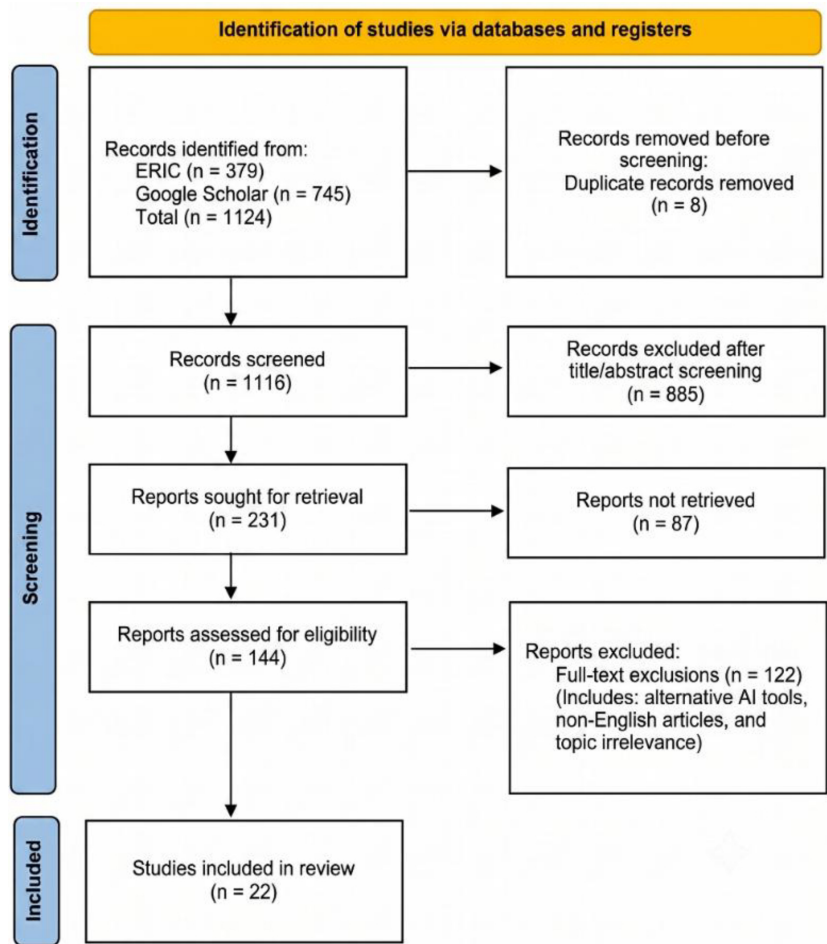


Fig. 1. PRISMA flow chart.

The 22 ultimately selected articles constituted a comprehensive dataset that thoroughly addressed the application of ChatGPT in enhancing speaking and writing proficiency. Relevant information was systematically extracted from these articles to address the formulated research questions.

Limitations of the study

Some limitations to this review of the literature on ChatGPT's role in improving speaking and writing skills need to be acknowledged. Firstly, this review paper has a scope limitation in that it reviewed and focused on only twenty-two selected articles, which may not fully reflect all the research done on this topic. Secondly, the context of a region is important in the application of ChatGPT into language learning. The outcomes of the reviewed research may be applicable in some regions and not in others, particularly in regions with little technological resources. Additionally, the focus on only English-written articles hardly allows broader generalizations to be made. Furthermore, it is acknowledged that some of the selected studies originate from journals with varying levels of indexing and impact. While a manual appraisal was conducted to ensure the methodological rigor of each selected paper, the inclusion of studies from emerging journals may influence the generalizability of the findings. Future studies could address these limitations by searching more reputable databases like Scopus and Web of Science, including more papers in their systematic reviews, and selecting non-English research articles for inclusion as well.

Results

The systematic review of the selected articles depicts the considerable influence of ChatGPT on the enhancement of writing and speaking skills. Nonetheless, it also identifies several key limitations. [Table 2](#) summarizes the principal findings regarding ChatGPT's potential for developing learners' speaking and writing abilities alongside its functional limitations.

ChatGPT and speaking

As [Table 2](#) depicts, ChatGPT's conversational interface was recognized as an invaluable instrument for fostering a safe and supportive environment for language practice. [Muniandy and Selvanathan's \(2024\)](#) study emphasized its capacity to facilitate seamless interaction. Building on this, recent studies illustrate that ChatGPT enables learners to enhance fluency, pronunciation, and overall communicative abilities through supplementary practice and immediate, structured feedback ([Salsabil et al., 2025](#); [Zulianingrum & Manggolo, 2025](#)). By acting as a non-judgmental speaking tutor, the tool substantially lowers speaking anxiety, allowing students to practice and refine their vocabulary and grammar without the fear of making mistakes ([Celik et al., 2025](#); [Lestari et al., 2026](#); [Salsabil et al., 2025](#)).

Furthermore, students reported increased motivation and confidence as ChatGPT provided flexible, low-anxiety opportunities for speaking practice beyond traditional classroom limitations ([Celik et al., 2025](#); [Lestari et al., 2026](#)). Nonetheless, the findings indicate that speaking skills benefit less from ChatGPT compared to writing skills ([Al-Obaydi et al., 2025](#); [Chávez & Naranjo, 2024](#)). Although tools like ChatGPT can simulate dialogues and provide vocabulary and grammar corrections, they fall short in replicating the nuanced feedback that human interaction offers.

Table 2. Synthesis of reviewed literature regarding ChatGPT in speaking and writing instruction.

Article Reviewed	ChatGPT in Speaking Skill Improvement	ChatGPT in Writing Skill Improvement	Limitations of ChatGPT
Al-Obaydi et al. (2025)	Not addressed.	Moderately improves written communication via content analysis (p. 16).	Ineffective at improving oral speaking skills (p. 16).
Balcı (2024)	Facilitates simultaneous reading, listening, and speaking practice via generated dialogues (p. 75).	Positively impacts EFL writing (p. 75).	Ineffective for informal writing. Offers limited actual speaking practice (p. 77).
Barrot (2023)	Not addressed.	Enhances writing habits and quality when integrated systematically across all writing phases (p. 5).	Lacks emotional depth and a unique authorial voice. Struggles with higher-order thinking, complex contexts, accuracy verification, and nuanced topics (pp. 4-5).
Celik et al. (2025)	Improves grammar, pronunciation, and vocabulary range. Boosts confidence and reduces language anxiety (p. 12).	Not addressed.	Responses lack human emotion, making interactions feel unnatural (p. 13).
Chávez & Naranjo (2024)	Not addressed.	Highly beneficial for language studies, specifically Spanish writing (p. 484).	Free version lacks enough speaking practice; advanced voice features require a paid subscription (pp. 488, 491).
Datskiv et al. (2024)	Not addressed.	Develops academic writing and critical thinking for future language teachers (pp. 133-134).	Produces predictable, unoriginal content. Prone to “hallucinations” requiring mandatory fact-checking (p. 132).
Fikri (2024)	Not addressed.	Effective for error detection and feedback, depending heavily on teacher planning (pp. 133, 137).	Initial learning curve for prompt engineering. Requires reliable internet and devices (p. 132).
Franjić and Božinovski (2025)	Not addressed.	Enhances ESL grammar and vocabulary across all proficiency levels (p. 71).	High cognitive demand. Risks over-reliance and increases writing time when unassisted (p. 71).
Hoang (2025)	Not addressed.	Improves task achievement, coherence, motivation, and engagement (p. 20).	Not addressed.
Jamshed et al. (2024)	Not addressed.	Mobile use effectively corrects basic ESL grammar (e.g., tenses, plurals) (p. 32).	Less effective for complex grammar rules. Requires human guidance for optimal use (p. 31).
Khojasteh et al. (2025)	Not addressed.	Enhances vocabulary acquisition via immediate, contextualized feedback (p. 15).	Not addressed.
Lestari et al. (2026)	Provides a non-judgmental, low-stress environment for practice. Improves vocabulary and sentence organization. Increases motivation (p. 18972).	Not addressed.	Lacks emotional connection. Repetitive conversations unsuitable for deep or prolonged discussions (p. 18972).

(Continued)

Table 2. Continued.

Article Reviewed	ChatGPT in Speaking Skill Improvement	ChatGPT in Writing Skill Improvement	Limitations of ChatGPT
Mohamed et al. (2025)	Not addressed.	Improves rule-based ESP (English for Specific Purposes) skills (basic grammar, formatting, organization) via self-directed learning (p. 15).	Inadequate for advanced skills like coherence, clarity, and complex grammar (p. 15).
Muniandy and Selvanathan (2024)	Conversational nature naturally improves ESL speaking skills (p. 15).	Not addressed.	Server capacity limits consistent access. Output accuracy and reliability concerns (p. 15).
Nam et al. (2026)	Not addressed.	Improves content and idea development. Leads to significant gains across all writing subcomponents, particularly grammar and content (p. 23).	Risks over-reliance, which may impair critical thinking, and is prone to misinformation or “hallucinations” (p. 6).
Özçelik and Ekşi (2024)	Not addressed.	Effectively improves formal writing styles and engages students (p. 14).	Struggles with informal/neutral styles. Susceptible to technical issues (p. 14).
Polakova and Ivenz (2024)	Not addressed.	Enhances proficiency in conciseness, grammar, inclusion of key information, and the use of passive voice (p. 6).	Occasional provision of wrong information requiring mandatory fact-checking. Potential for over-reliance leading to learning “laziness” (pp. 2-3)
Sain et al. (2025)	Not addressed.	Valuable for brainstorming, organizing thoughts, and translation (pp. 47-48).	Prone to providing incorrect information. Risks diminishing critical thinking and creativity. Lacks deep domain expertise (pp. 48-49).
Salsabil et al. (2025)	Enhances fluency, grammar, and pronunciation over traditional methods. Non-judgmental feedback lowers anxiety (p. 47).	Not addressed.	Extended use causes “cognitive fatigue,” especially with complex or unfamiliar topics (p. 47).
Xiao et al. (2025)	Not addressed.	Enables anytime practice with personalized feedback. Increases student motivation and autonomy (p. 10).	Risks academic dishonesty and reduced teacher interaction. Lacks cultural context. Struggles with complex grammar and creative evaluation (p. 11).
Zhao et al. (2024)	Not addressed.	Lowers stress during academic writing tasks (p. 6).	Requires heavy fact-checking. Raises fears regarding plagiarism detection and skill dependency (p. 6).
Zulianingrum & Manggolo (2025)	Offers real-time, organized feedback for independent practice and self-correction (p. 82).	Not addressed.	Not addressed.

ChatGPT and writing

The findings of the reviewed articles consistently illustrated that ChatGPT positively impacts learners' writing proficiency. Multiple studies emphasized its effectiveness in offering real-time feedback, analyzing errors, and improving grammar accuracy. For instance, [Fikri \(2024\)](#) and [Al-Obaydi et al. \(2025\)](#) noted that ChatGPT facilitated error analysis and feedback provision, while [Jamshed et al. \(2024\)](#) and [Franjić and Božinovski \(2025\)](#) highlighted its capacity to enhance grammatical accuracy, particularly in challenging areas such as tense usage. This is further corroborated by [Xiao et al. \(2025\)](#), [Hoang \(2025\)](#), [Khojasteh et al. \(2025\)](#), and [Polakova and Ivenz \(2024\)](#), who pointed out that ChatGPT provides students with real-time feedback on grammar, vocabulary, sentence structure, and organization, while offering flexible language input outside traditional classroom settings.

The tool also serves as a crucial asset during the initial and structural stages of writing. [Sain et al. \(2025\)](#) and [Nam et al. \(2026\)](#) emphasized its vital role in brainstorming, generating ideas, and organizing thoughts into a coherent structure. Additionally, ChatGPT's utility in supporting formal writing tasks was acknowledged, with [Özçelik and Ekşi \(2024\)](#) reporting that students actively benefited from its suggestions and corrections to improve the formal aspects of their work.

Furthermore, pedagogical integration is emerging as a crucial factor for using ChatGPT effectively, because this integration can significantly influence the overall learning outcomes. [Barrot \(2023\)](#) emphasized the significance of intentional and systematic integration during the writing process. This finding was further supported by [Fikri \(2024\)](#), who examined the incorporation of ChatGPT into lesson plans.

Beyond instructional design, the psychological and motivational impacts of the tool are noteworthy. Learner perceptions indicated that ChatGPT's user-friendliness and capacity to provide immediate, actionable feedback play a crucial role in its effectiveness in teaching writing. Consequently, students reported higher motivation and engagement due to the autonomy and personalized feedback offered by the AI ([Nam et al., 2026](#); [Xiao et al., 2025](#)), and the use of ChatGPT was often justified for reducing their writing-related stress ([Zhao et al., 2024](#)). Finally, ChatGPT's robust language translation capabilities make it a highly versatile tool for cross-lingual communication ([Sain et al., 2025](#)), rounding out its comprehensive support for language learners.

Limitations of ChatGPT

Despite the above-mentioned advantages, the limitations of ChatGPT were noted throughout the reviewed studies. One of the most significant drawbacks is its lack of originality and depth in content generation. This was particularly emphasized by [Datskiv et al. \(2024\)](#), who cautioned that the chatbot often produces predictable output with limited academic value. Similarly, [Xiao et al. \(2025\)](#) and [Mohamed et al. \(2025\)](#) observed that ChatGPT struggles with deep meaning, creativity, and coherence, while [Sain et al. \(2025\)](#) argued that its failure to understand the intricacies of specific academic fields compromises input quality. Furthermore, the generation of inaccurate information raises concerns regarding reliability and necessitates rigorous fact-checking ([Nam et al., 2026](#); [Polakova and Ivenz, 2024](#); [Sain et al., 2025](#); [Zhao et al., 2024](#)). Closely tied to this are issues of academic integrity and over-reliance; excessive dependence on the tool can hinder the development of critical thinking, discourage engagement with human instructors, and ultimately undermine students' independent writing development ([Franjić and Božinovski, 2025](#); [Nam et al., 2026](#); [Polakova and Ivenz, 2024](#); [Sain et al., 2025](#); [Xiao et al., 2025](#); [Zhao et al., 2024](#)).

Additionally, linguistic and pragmatic limitations hinder the tool's instructional versatility. Challenges in teaching informal and neutral registers were reported by [Özçelik and Ekşi \(2024\)](#), a finding expanded upon by [Xiao et al. \(2025\)](#), who noted ChatGPT's struggle to provide culturally and contextually sensitive feedback or consistently handle complex language rules and idiomatic expressions. Moreover, [Jamshed et al. \(2024\)](#) found that while ChatGPT is effective for basic grammar correction, traditional methods remain superior for mastering more complex grammatical structures. When used for speaking and conversational practice, the AI's lack of emotional engagement and natural human-like responsiveness can lead to repetitive interactions ([Celik et al., 2025](#); [Lestari et al., 2026](#)). Prolonged use in these conversational contexts can even induce cognitive fatigue and confusion among learners ([Salsabil et al., 2025](#)). Finally, technical and accessibility issues present persistent barriers, especially in contexts where reliable internet access or sufficient server capacity is unavailable, as highlighted by [Muniandy and Selvanathan \(2024\)](#) and [Fikri \(2024\)](#).

Discussion

The systematic analysis of the existing literature highlights ChatGPT's considerable capacity to enhance speaking and writing skills. Rather than merely cataloging these effects, this review synthesizes the findings through the theoretical lenses of [Vygotsky's \(1978\)](#) Sociocultural Theory and the SAMR model ([Puentedura, 2009, 2020](#)). Viewed as a highly advanced language learning assistant, ChatGPT helps ease language barriers for both learners and educators. However, a critical engagement with the literature reveals profound complexities, paradoxes, and ethical challenges that must be addressed to fully understand its long-term pedagogical implications.

Redefining productive skills via AI scaffolding

When evaluated through [Vygotsky's \(1978\)](#) framework, ChatGPT functions as an accessible, virtual MKO that provides essential scaffolding for productive language tasks. In writing instruction, the tool goes beyond mere substitution (as outlined in the SAMR model) by actively transforming the learning process through the Modification and Re-definition stages. Its ability to assist in brainstorming, structuring ideas, and delivering real-time, personalized feedback allows students to operate securely within their ZPD. By purposefully incorporating this dynamic scaffolding into lesson planning, educators observe improved grammatical accuracy (especially in complex areas like tense and plurality) upheld positive writing practices, and greatly alleviated writing-related stress.

Similarly, ChatGPT holds substantial promise in developing speaking skills, though an analysis of the literature reveals notable contradictions when compared to its efficacy in writing. As a supplementary speaking tutor, its conversational interface establishes a safe, non-judgmental platform, enabling learners to practice beyond traditional classroom limitations. In SAMR terms, this redefines speaking practice by providing on-demand, interactive dialogues that lower the affective filter. The instantaneous feedback serves as a scaffold that increases learner confidence and helps refine fluency and pronunciation. However, the literature indicates that while ChatGPT excels at the structural and mechanical elements of written text, its performance in advancing dynamic speaking skills is inconsistent, largely constrained by its inability to replicate authentic socio-emotional nuances.

The paradox of autonomy and dependency

A critical tension emerges across the reviewed studies regarding the impact of generative AI on learner independence. On the one hand, the flexibility of ChatGPT fosters learner autonomy; by providing outside-the-classroom access and on-demand feedback, it significantly boosts learners' motivation and empowers them to take charge of their linguistic exposure.

However, this increased autonomy paradoxically introduces a profound risk of AI dependency. According to [Vygotsky's \(1978\)](#) theory, scaffolding is intended to be temporary, gradually removed as the learner internalizes the target skill. Yet, prolonged reliance on ChatGPT for generating ideas, structuring essays, and correcting errors can lead to cognitive outsourcing. Instead of engaging in the productive struggle necessary for genuine language acquisition, students may become overly dependent on the virtual MKO. This dynamic threatens to hinder the development of their independent critical thinking and problem-solving skills. Consequently, the long-term pedagogical implication is that educators can no longer focus solely on language instruction; they must explicitly teach AI literacy, training students to critically self-regulate their use of AI as a temporary supportive scaffold rather than a permanent cognitive crutch.

Ethical challenges and pedagogical limitations

Despite its transformative potential, the integration of ChatGPT into language learning presents complex limitations and ethical concerns that dominate current educational debates. Chief among these is the phenomenon of AI hallucinations. Within a ZPD framework, an MKO must provide reliable guidance; when ChatGPT generates plausible but factually inaccurate information, it compromises the input quality and fundamentally misguides the learner. This necessitates rigorous fact-checking and places a new, often untaught, cognitive burden on students.

Furthermore, ethical concerns regarding academic integrity extend far beyond traditional definitions of plagiarism. Because ChatGPT relies on predictive language models rather than genuine comprehension, it frequently exhibits lower-order thinking and a lack of true creativity. The tool struggles with deep meaning, coherence, and the generation of highly original ideas. If students rely heavily on these outputs, there is a serious ethical concern regarding the degradation of authentic academic voice and original thought.

Additionally, linguistic and socio-cultural limitations restrict ChatGPT's pedagogical scope. The tool is primarily optimized for formal communication and frequently fails to provide culturally and contextually sensitive feedback for informal registers. In conversational settings, students report a distinct lack of emotional engagement and human-like responsiveness, which can induce cognitive fatigue and make extended interactions feel unnatural and repetitive. Ultimately, while AI excels at the mechanics of language production, it currently falters at the pragmatics of human connection. Finally, issues of digital equity must be acknowledged; restricted access due to limited server capacity, paywalls, or unstable internet connections disproportionately disadvantages economically vulnerable learners, threatening to widen the educational digital divide.

Conclusion and future directions

The objective of this systematic review was to evaluate the efficacy and limitations of ChatGPT in enhancing oral and written language skills. The synthesis of the twenty-two

selected articles suggests that ChatGPT functions as a highly capable virtual assistant that appears to improve grammatical accuracy, writing organization, and speaking confidence. By providing a safe, non-judgmental environment for practice, the tool may effectively lower the affective filter and encourage learner autonomy. However, the findings also reveal critical limitations, including the potential generation of inaccurate information, a perceived lack of socio-emotional depth in conversational settings, and the persistent risk of fostering cognitive dependency among learners.

By anchoring the analysis in [Vygotsky's \(1978\)](#) Sociocultural Theory and the SAMR model, this paper posits that generative AI may not simply substitute traditional teaching methods; rather, it appears to redefine the boundaries of learner autonomy and cognitive scaffolding. We highlight a fundamental complexity: although generative models display high proficiency in the structural elements of language, they remain fundamentally limited in replicating authentic, socio-emotional human discourse. Consequently, this review argues that the future of EFL research might benefit from pivoting from measuring short-term linguistic gains to evaluating the possible long-term cognitive and ethical impacts of human-AI interaction.

To address the paradox of AI dependency and the ethical challenges identified in this review, a paradigm shift in language pedagogy may be required. It seems educators can no longer afford to focus exclusively on linguistic input; rather, they may need to integrate critical AI literacy as a core component of the language curriculum. We propose that teachers could explicitly train students not just to use AI, but to critically evaluate its outputs. This might include teaching learners how to identify hallucinations, recognize cultural or linguistic biases in AI responses, and understand the difference between AI-assisted brainstorming and cognitive outsourcing. Educators could consider utilizing ChatGPT for the Substitution and Augmentation phases of learning (e.g., grammar checking, initial brainstorming) while potentially mandating AI-free communicative tasks to assess true socio-pragmatic competence and authentic human-to-human interaction. Furthermore, institutions might benefit from moving away from purely punitive plagiarism policies, instead exploring collaborative frameworks that teach students how to ethically co-create with AI while maintaining their authentic academic voice.

While the current body of literature provides a strong foundation, significant gaps remain. Future scholarship would arguably benefit from transitioning away from short-term, descriptive case studies to prioritize longitudinal research, which could better assess the potential long-term effects of AI dependency on critical thinking and original language production. Furthermore, empirical investigations seem necessary to explore how cognitive fatigue might manifest during prolonged human-AI conversational practice. Finally, as technical access and server limitations were frequently cited as barriers, future research ought to critically examine the impact of the digital divide to ensure that the integration of premium AI tools does not inadvertently exacerbate educational inequities among diverse socio-economic learner populations.

Responsible use statement for Artificial Intelligence tools

Gemini 3.1 Pro was employed during the preparation of this manuscript to assist with language editing, grammatical correction, and the enhancement of the paper's scholarly tone. The authors have thoroughly reviewed and validated all AI-generated suggestions and assume full responsibility for the accuracy and originality of the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Data availability

No datasets were generated or analyzed during the current study.

Author contribution

E.S.: Data curation, investigation, writing – original draft (abstract, introduction, literature review, results). M.A.: Methodology, formal analysis, writing – original draft (methodology, discussion, conclusion), writing – review & editing. Both authors read and approved the final manuscript.

References

- Abdullayeva, M., & Musayeva, Z. M. (2023, April). The impact of Chat GPT on student's writing skills: An exploration of AI-assisted writing tools. In *International Conference of Education, Research and Innovation* (Vol. 1, No. 4, pp. 61–66).
- Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, 3(1). <https://doi.org/10.1007/s44217-024-00138-2>
- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), 643. <https://doi.org/10.3390/educsci14060643>
- 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>
- Allen, T., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese EFL learners' preferences for editing and proofreading strategies. *REL C Journal*. <https://doi.org/10.1177/003368822412625>
- Al-Obaydi, L. H., Pikhart, M., & Hossain, M. K. (2025). ChatGPT and the development of core language skills: An exploratory study of EFL college students. *Contemporary Educational Technology*, 17(4), ep591. <https://doi.org/10.30935/cedtech/17242>
- Balci, 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). <https://doi.org/10.5281/zenodo.12544675>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, 8(1), 102418. <https://doi.org/10.29140/ajal.v8n1.102418>
- Chávez, Y., & Naranjo, C. (2024). ChatGPT in Spanish college classes: Is it beneficial for listening, speaking, reading and writing skills? *International Journal of Education Humanities and Social Science*, 7(6), 482–510.
- Datskiv, O., Zadorozhna, I., & Shon, O. (2024). Developing future teachers' academic writing and critical thinking skills using ChatGPT. *International Journal of Emerging Technologies in Learning (IJET)*, 19(07), 126–136. <https://doi.org/10.3991/ijet.v19i07.49935>
- Fikri, H. (2024). The use of ChatGPT in teaching writing. *International Journal of Language, Education, and Literature*, 2(1), 129–139. <https://doi.org/10.5281/zenodo.13864708>
- Frantić, M., & Božinovski, B. (2025). Exploring ChatGPT's impact on ESL writing proficiency: An experimental approach. *Teaching English With Technology*, 25(2). <https://doi.org/10.56297/vaca6841/zddd4208/luzg7476>
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. <https://doi.org/10.1177/13621688241239764>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>

- Hoang, N. H. (2025). ChatGPT-assisted language learning: Effects on Vietnamese English majors' writing skills and motivation. *The JALT CALL Journal*, 21(2), 102429. <https://doi.org/10.29140/jaltcall.v21n2.102429>
- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4), ep464. <https://doi.org/10.30935/cedtech/13605>
- Jamshed, M., Manjur Ahmed, A. S. M., Sarfaraj, M., & Warda, W. U. (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).
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khojasteh, L., Soori, A., & Javed, F. (2025). Comparing teacher E-feedback, AI feedback, and hybrid feedback in enhancing EFL writing skills. *Technology in Language Teaching & Learning*, 7(3), 102626. <https://doi.org/10.29140/tltl.v7n3.102626>
- Lestari, V. A., Hilman, A., Purba, A. B., Mutawally, A. N., Pratama, R. T., Khasani, I., Ludovikus, Alfiansyah, D. M., Febriani, D., Natalia, I., Indriani, S., Grahana, C. I., & Putra, J. A. (2026). Exploring students' voices on the use of Chatgpt as a speaking partner in english practice. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(3), 18967–18974. <https://doi.org/10.31004/jerkin.v4i3.5046>
- 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>
- Mohamed, A. M., Deen, A. A. M. M. A. E., Abukhait, R. O., Dahleb, F., Khan, Y. B., & Jmaiel, H. A. (2025). ChatGPT's impact on ESP writing proficiency and learner autonomy: An experimental study. *Technology in Language Teaching & Learning*, 7(3), 102964. <https://doi.org/10.29140/tltl.v7n3.102964>
- Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a Public University, Malaysia. *Teaching Public Administration*. <https://doi.org/10.1177/01447394241230152>
- Nam, Y., Lee, J. M., & Park, H. (2026). Improving L2 writing through Artificial intelligence: Comparing AIDT and ChatGPT feedback. *English Teaching*, 81(1), 3–36. <https://doi.org/10.15858/engtea.81.1.202603.3>
- Özçelik, N. P., & Ekşi, G. Y. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1). <https://doi.org/10.1186/s40561-024-00296-8>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S.,... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- 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>
- Puentedura, R. R. (2009). As we may teach: educational technology, from theory into practice. Apple.
- Puentedura, R. R. (2020, January 25). SAMR - A research perspective. http://hippasus.com/rpweblog/archives/2020/01/SAMR_AResearchPerspective.pdf
- Sain, Z. H., Serban, R., Abdullah, N. B., & Thelma, C. C. (2025). Benefits and drawbacks of leveraging ChatGPT to enhance writing skills in secondary Education. *At-tadzkir Islamic Education Journal*, 4(1), 40–52. <https://doi.org/10.59373/attadzkir.v4i1.79>
- Salsabil, A. D., Rakhmawati, L. A., & Febtwenesty. (2025). From silent learners to confident speakers: The effect of AI voice chat with ChatGPT on EFL speaking skills. *Journal of English Education*, 3(1), 38–50. <https://doi.org/10.61994/jee.v3i1.1137>
- Sok, S., & Heng, K. (2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Journal of Digital Educational Technology*, 4(1), ep2401. <https://doi.org/10.30935/jdet/14027>
- 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, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Teng, M. F. (2024a). 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. (2024b). “ChatGPT is the companion, not enemies”: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Williams, T. (2023). Turnitin announces AI detector with 97 per cent accuracy. *timeshighereducation.com*. <https://www.timeshighereducation.com/news/turnitin-announces-ai-detector-97-cent-accuracy>
- Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in english as a foreign language education: A scoping review. *ECNU Review of Education*, 9(1). <https://doi.org/10.1177/20965311241310881>
- Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-02904-x>
- Zulianingrum, A., & Manggolo, N. S. K. H. (2025). A human–AI interaction model for improving speaking skills in EFL learning using ChatGPT. *Nusantara Journal of Artificial Intelligence and Information Systems*, 1(2), 77–84. <https://doi.org/10.47776/nuai.v1i2.1752>