AUREVIAJ JOURNAL OF SOCIAL SCIENCES AND HUMANITIES
https://aureviaj.com/ajssh/article/view/17

AUREVIAJ

Research Article | Open Access

Generative Artificial Intelligence in English Language Learning: A Scoping Review

Liam Pani

Master's in TESOL, Swinburne University of Technology Sarawak Campus, Malaysia

ABSTRACT

The rapid emergence of generative artificial intelligence (GenAI) has transformed the technological landscape of English language learning. This scoping review maps current evidence on GenAI applications, learning outcomes, pedagogical functions, challenges, and research gaps in English language education. The review followed the Arksey and O'Malley scoping-review framework and subsequent methodological refinements, with reporting guided by PRISMA-ScR and PRISMA 2020 principles. A structured search of publicly accessible scholarly records was conducted and cross-checked against recent systematic and scoping reviews of GenAI, ChatGPT, and English language education. Literature published from the emergence of contemporary public-facing GenAI through 22 August 2026 was considered, with primary empirical research prioritized in the synthesis. The evidence demonstrates a strong concentration on English writing, particularly academic writing, revision, feedback, and assessment. Other applications include speaking practice, vocabulary and grammar development, reading support, personalized and self-directed learning, and instructional activities. Overall, the literature reports positive effects or perceptions concerning writing performance, engagement, learner autonomy, feedback, and opportunities for language practice, although outcomes vary according to pedagogical design and human supervision. Major concerns include inaccurate output, overreliance, reduced learner agency, academic dishonesty, privacy, bias, and inequitable access. Research remains concentrated in higher education and East Asian EFL contexts, with limited evidence concerning listening, pronunciation, younger learners, low-resource contexts, longitudinal development, and independently assessed outcomes. The review concludes that GenAI is most appropriately conceptualized as a pedagogically mediated learning resource rather than a substitute for language teachers.

Cite as: Pani, L. (2026). Generative Artificial Intelligence in English Language Learning: A Scoping Review. *Aureviaj Journal of Social Sciences and Humanities. 1*(1), 17-38
https://aureviaj.com/ajssh/article/view/17

ARTICLE HISTORY

Received 18 Aug 2026
Accepted 23 Aug 2026

KEYWORDS

Generative artificial intelligence; ChatGPT; English language learning; EFL; Large language models; Artificial intelligence; Language education

INTRODUCTION

Artificial intelligence has been incorporated into language education for decades through automated writing evaluation, speech recognition, intelligent tutoring systems, machine translation, adaptive learning environments, and other forms of computer-assisted language learning. The emergence of generative artificial intelligence (GenAI), however, represents an important shift because contemporary large language models can generate novel linguistic

*CONTACT [1]Liam Pani pa0ni.liam@gamil.com Master's in TESOL, Swinburne University of Technology Sarawak Campus, Malaysia

content and participate in interactive exchanges with learners rather than merely evaluating or retrieving predefined information. The public release of ChatGPT on 30 November 2022 accelerated this transformation and resulted in an unprecedented expansion of research examining generative models in education and language learning (Lo et al., 2024).

The relevance of GenAI to English language learning derives from its capacity to perform several functions traditionally distributed among teachers, textbooks, language-learning applications, peers, and reference materials. A learner can request an explanation of a grammatical structure, generate examples of vocabulary in context, obtain feedback on a written paragraph, practice a simulated conversation, receive alternative formulations of an utterance, or engage in repeated language practice. Such interactions potentially increase access to linguistic input, output opportunities, feedback, and individualized support. Recent empirical research consequently describes GenAI as a potentially flexible learning partner capable of supporting multiple dimensions of English language development (Lo et al., 2024).

The rapid expansion of this field is evident from the growth of review literature. An early systematic review of the first year of ChatGPT-related language-education research identified 36 studies published between November 2022 and November 2023. The review found that research was dominated by higher education and empirical investigations, while identifying self-directed learning, content generation, and teacher support among the principal uses of ChatGPT (Lee et al., 2024).

A subsequent systematic review by Lo et al. (2024) identified 70 empirical studies concerning ChatGPT in English as a second or foreign language education during approximately the first 18 months following its release. The review found that ChatGPT provided increased learning opportunities, personalized learning, and teacher support, but also generated concerns regarding inaccurate information, privacy, and academic dishonesty. Importantly, writing dominated the literature, whereas reading, speaking, and listening were substantially less investigated.

The evidence base has continued to expand. A 2025 scoping review of empirical GenAI research in language education identified 43 studies published in Social Sciences Citation Index journals between 2022 and 2024. Most of the studies focused on English language teaching and learning, particularly among Chinese learners and teachers in higher education, with writing representing the most prominent language skill. The authors also identified methodological limitations involving insufficient reporting of AI versions and prompts and limited theoretical grounding (Wang et al., 2025).

A broader systematic review subsequently examined 144 peer-reviewed empirical studies published during 2023–2024 and reported an exponential expansion of empirical GenAI research in language learning and teaching. The review identified six major research areas: perceptions and attitudes; psychological and cognitive dimensions; teaching and learning strategies; language-skill development; writing and feedback; and implementation and integration. Higher education represented 86.7% of the research context, EFL represented 86.1%, and writing accounted for 42.4% of the research focus, demonstrating a substantial concentration of research in particular contexts and skills (Li et al., 2025).

Recent 2026 evidence confirms that the field remains heavily concentrated in particular contexts. Deng and Jamaludin (2026), for example, identified 51 articles from an initial pool of 284 studies in Scopus and Web of Science in a PRISMA-based review of GenAI in EFL instruction. Higher education, East Asia and the Middle East, and writing instruction were especially prominent. Similarly, Uygun and Demir (2026) synthesized empirical AI research in EFL/ESL education published between January 2023 and September 2025 using Web of Science and Scopus and reported that AI applications primarily supported personalized and adaptive learning, particularly writing, while technical, ethical, privacy, and teacher-preparation challenges remained significant.

The dominance of writing is understandable. Text-based GenAI systems are particularly capable of generating, revising, evaluating, and reformulating written language. Empirical studies have reported improvements in several dimensions of EFL writing following structured use of ChatGPT. For example, Polakova and Ivenz (2024), using a quasi-experimental mixed-methods design involving 110 university EFL learners, reported improvements in aspects of writing following ChatGPT feedback. Mahapatra (2024) similarly examined ChatGPT as a formative feedback tool for undergraduate ESL academic writing, while Song and Song (2023) reported positive effects of AI-assisted instruction on EFL writing and writing motivation.

Other research has examined the quality of GenAI feedback rather than its effect on learner performance. Li et al. (2024), for example, compared ChatGPT-3.5 and ChatGPT-4 with experienced EFL teachers in evaluating 30 university students' essays. ChatGPT-4 demonstrated higher scoring reliability than the participating teacher raters, while both ChatGPT versions provided relevant feedback across language, content, and organization. These findings suggest that GenAI may function as a supplementary assessment and feedback resource, although they should not be interpreted as evidence that automated evaluation can replace expert human judgment.

Writing is not the only domain receiving attention. Recent research has begun to examine GenAI-mediated speaking practice. A 2024 experimental study involving Chinese university EFL learners compared conventional classroom interaction with text/voice-based conversational GenAI and avatar-based conversational GenAI, examining willingness to communicate, foreign-language speaking anxiety, self-perceived communicative competence, and speaking performance. More recent research has investigated GenAI-mediated informal digital learning of English speaking among vocational college students and reported significant improvement in performance on the practiced speaking task, although intrinsic motivation and speaking self-efficacy did not differ significantly.

GenAI may also influence learner engagement. A 2026 study of 89 Chinese English learners practicing spoken English with a GenAI chatbot found that students receiving teacher guidance demonstrated higher behavioral, cognitive, and social engagement than students in a regular-use condition. This finding is particularly important because it suggests that the educational value of GenAI may depend not simply on access to the technology but on how pedagogical guidance structures learner interaction with it.

Another important domain is self-directed English learning. GenAI provides learners with a potentially available conversational partner outside formal classroom settings. A multi-

methods study involving a systematic review of 40 empirical studies, a survey of 344 EFL learners, and interviews with 19 learners found that interactivity, enjoyment, trust, and subjective norms influenced continued ChatGPT use, while learners employed the system for reading, writing, vocabulary, and grammar learning (2024). This evidence suggests that GenAI may contribute to informal and autonomous language learning, but it also raises questions concerning learner self-regulation and the ability to evaluate AI-generated information independently.

The relationship between GenAI and learner autonomy is therefore complex. On one hand, learners can access individualized explanations and repeated practice without waiting for teacher assistance. On the other hand, excessive reliance on generated answers may reduce the cognitive effort required to formulate language independently. Recent scholarship increasingly emphasizes this tension between technological assistance and learner agency. A 2026 synthesis of research on GenAI in language education identified learner overreliance, weakened agency, reliability problems, unequal access, and inadequate teacher preparation as important concerns alongside the pedagogical opportunities offered by GenAI.

Accuracy constitutes another fundamental issue. Large language models can generate grammatically well-formed but factually incorrect or contextually inappropriate responses. In language education, this problem is particularly consequential because learners may interpret fluent output as authoritative linguistic information. Lo et al. (2024) identified incorrect information as one of the principal drawbacks reported across the ESL/EFL literature. The problem becomes particularly important for pragmatics, culturally embedded expressions, specialized vocabulary, and language varieties that are less strongly represented in training data.

Academic integrity is another recurring concern. GenAI can produce essays, summaries, translations, discussion responses, and other forms of academic writing that learners may submit as their own. Reviews of the field repeatedly identify academic dishonesty as an important challenge. However, framing GenAI exclusively as a plagiarism problem may overlook its potential to transform assessment itself. Rather than attempting simply to prohibit AI use, educators may need assessment designs that emphasize process, oral defense, reflection, drafting histories, source evaluation, and individualized application of knowledge.

Ethical considerations extend beyond academic integrity. UNESCO (2023) emphasizes privacy, human agency, equity, transparency, cultural and linguistic diversity, and appropriate human oversight in educational applications of GenAI. These considerations are especially important in language education because English is a globally dominant language while many learners and educational institutions operate within contexts characterized by unequal technological resources. GenAI therefore has the potential simultaneously to expand access to English learning and to reproduce existing digital and linguistic inequalities.

The geographical distribution of research reinforces this concern. Recent reviews demonstrate a strong concentration of GenAI language-learning research in East Asian higher education contexts, particularly China. Consequently, evidence from South Asia, Africa, the Middle East, Latin America, low-resource educational environments, and non-university populations remains comparatively limited. The research concentration may also shape what is understood

as effective English learning because the pedagogical priorities, institutional structures, assessment practices, and English-learning motivations of these contexts differ substantially.

Theoretical development is another emerging concern. Wang et al. (2025) reported inadequate use of explicit theoretical frameworks in the empirical GenAI language-education literature. Research that merely asks whether learners "like" ChatGPT provides useful descriptive information but does not necessarily explain how AI-mediated interaction influences acquisition processes. Stronger theoretical integration with second-language acquisition, sociocultural theory, interactionist perspectives, self-regulated learning, learner autonomy, cognitive load theory, and technology-adoption models may provide more explanatory power.

The literature is also evolving from research on ChatGPT specifically toward research on GenAI as a broader technological category. Recent work includes multimodal conversational systems, voice-enabled chatbots, AI-supported feedback, and other large-language-model-based systems. The 2026 literature already demonstrates this expansion, including research on spoken-English engagement and GenAI-mediated informal learning. Therefore, a review restricted exclusively to ChatGPT risks becoming technologically obsolete as new systems and multimodal capabilities emerge.

A scoping review is appropriate for this rapidly developing field because the objective is to map the breadth and characteristics of existing evidence rather than estimate one pooled intervention effect. Arksey and O'Malley (2005) originally conceptualized scoping reviews as a means of examining the extent, range, and nature of research on an emerging topic. Subsequent methodological work has emphasized transparent research questions, systematic identification and selection of studies, data charting, and descriptive and thematic synthesis (Levac et al., 2010). PRISMA-ScR subsequently provided reporting guidance specifically for scoping reviews, while PRISMA 2020 established updated standards for transparent systematic-review reporting (Page et al., 2021; Tricco et al., 2018).

Although several systematic reviews now examine ChatGPT, AI, or GenAI in language education, the continuing acceleration of publication warrants an updated mapping specifically focused on English language learning. The present review therefore differs from reviews restricted to ChatGPT, teacher use, writing, or higher education. It considers GenAI applications across English language skills and learning domains and examines both educational opportunities and constraints while identifying under-researched populations, skills, geographical contexts, and methodological approaches.

Accordingly, this scoping review addresses six research questions:

1. What applications of generative artificial intelligence have been reported in English language learning?
2. Which English language skills and learning domains have received the greatest research attention?
3. What learning, cognitive, affective, and behavioral outcomes have been associated with GenAI-supported English language learning?
4. What pedagogical roles does GenAI perform in English language learning?
5. What challenges, limitations, and ethical concerns are reported in the literature?

6. What methodological, geographical, educational, technological, and theoretical gaps characterize the existing evidence base?

RESEARCH METHODOLOGY

Research Design

The study employed a scoping-review methodology. The choice of methodology was determined by the breadth and heterogeneity of the emerging GenAI literature. Unlike a conventional systematic review that typically addresses a relatively focused intervention-effect question, a scoping review is designed to map the characteristics, concepts, evidence types, and research gaps within a developing field (Arksey & O'Malley, 2005; Levac et al., 2010).

The review was structured according to the stages of the Arksey and O'Malley framework, refined through later scoping-review methodology, and reported using PRISMA-ScR principles. PRISMA 2020 was additionally used to guide transparent reporting of identification, screening, eligibility, and inclusion decisions (Page et al., 2021; Tricco et al., 2018).

The review was not registered in PROSPERO because PROSPERO primarily registers systematic reviews addressing health-related questions and is not generally the appropriate registration platform for an educational scoping review. The review should therefore be considered a PRISMA-ScR-informed evidence map rather than a registered systematic review.

Search Strategy

A structured literature search was conducted on 22 August 2026. The search was designed around three conceptual components:

1. generative artificial intelligence;
2. English language learning; and
3. language-learning applications or outcomes.

The principal search expression was:

("generative artificial intelligence" OR "generative AI" OR GenAI OR ChatGPT OR GPT OR "large language model" OR LLM OR Gemini OR Claude OR Copilot)

AND

("English language learning" OR "English language education" OR EFL OR ESL OR "English as a foreign language" OR "English as a second language" OR "second language learning")

AND

(writing OR speaking OR listening OR reading OR vocabulary OR grammar OR pronunciation OR feedback OR interaction OR motivation OR autonomy OR "critical thinking" OR assessment)

The search terminology was informed by recent database-based reviews. Lo et al. (2024), for example, searched seven databases using combinations of ChatGPT/GPT terms and

ESL/EFL/English-L2 terms. Deng and Jamaludin (2026) used Scopus and Web of Science with GenAI-related terms combined with EFL and foreign-language-teaching terms.

Information Sources

The search and verification process incorporated publicly accessible scholarly records from major academic publishers and indexing services, including records associated with Scopus, Web of Science, ERIC, ScienceDirect, SpringerLink, Taylor & Francis Online, SAGE Journals, Frontiers, Google Scholar-indexed scholarly records, and publisher-level bibliographic records.

Recent systematic reviews that had already conducted authenticated Scopus, Web of Science, ERIC, and related database searches were also used as secondary search-validation sources, particularly for identifying relevant primary studies that might not be surfaced reliably through general web indexing.

This triangulation was important because current reviews demonstrate that database coverage differs substantially across studies. For example, Lo et al. (2024) searched seven databases, including Scopus, Web of Science, ERIC, ACM Digital Library, IEEE Xplore, Academic Search Ultimate, and Education Source Ultimate. More recent reviews used Scopus and Web of Science, while other reviews incorporated ERIC and ScienceDirect.

Eligibility Criteria

Studies were eligible when they satisfied all of the following broad criteria:

Table 1: *Eligibility Criteria*

Criterion	Inclusion criteria	Exclusion criteria
Population	Learners of English as a foreign, second, or additional language	Studies unrelated to English learning
Technology	Generative AI, ChatGPT, GPT-based systems, LLMs, Gemini, Claude, Copilot or comparable GenAI	Conventional non-generative AI only
Educational purpose	English learning, English language development, EFL/ESL instruction	General educational AI without an English-learning component
Outcomes	Language proficiency, writing, speaking, reading, listening, vocabulary, grammar, pronunciation, engagement, motivation, autonomy, feedback, critical thinking, etc.	Purely computational/technical evaluations without language-learning relevance
Publication type	Peer-reviewed empirical studies; relevant reviews used for citation chaining and contextual verification	Editorials, news reports, commercial webpages and promotional content
Time	November 2022–22 August 2026 for contemporary GenAI evidence	Studies predating public GenAI unless theoretically relevant
Language	English-language publications	Publications for which relevance could not be reliably established

Criterion	Inclusion criteria	Exclusion criteria
Accessibility	Sufficient bibliographic and abstract/full-text information for evidence mapping	Records with insufficient information to establish eligibility

The November 2022 starting point was selected because ChatGPT's public release represented a major inflection point in publicly accessible GenAI research. Earlier AI-assisted language-learning research was used only where necessary to establish conceptual continuity.

Screening and Selection

The search process followed four conceptual stages:

Identification → duplicate removal → title/abstract screening → full-text eligibility assessment → final evidence mapping.

The search was supplemented through backward reference checking and forward citation searching of major reviews, particularly the reviews by Lee et al. (2024), Lo et al. (2024), Wang et al. (2025), Li et al. (2025), Deng and Jamaludin (2026), and Uygun and Demir (2026).

Importantly, the exact number of records retrieved from proprietary Scopus and Web of Science interfaces is not reported as if independently exported by the present study, because authenticated database access was unavailable. Instead, database-specific retrieval figures reported by existing PRISMA studies were used as validation benchmarks. For example, Deng and Jamaludin (2026) reported 284 initial records and 51 included articles from Scopus and Web of Science, while Lee et al. (2024) reported 523 records from its broader database and hand-search strategy before arriving at 36 included articles.

This approach avoids the methodological error of presenting invented PRISMA counts as primary database retrievals.

Data Synthesis

Because the included evidence involved quantitative, qualitative, mixed-methods, experimental, quasi-experimental, survey, interview, case-study, and observational designs, statistical meta-analysis was not the primary synthesis method. Instead, findings were synthesized through descriptive mapping and thematic analysis.

The evidence was organized according to the six research questions. Research concentration was assessed by comparing the relative representation of language skills, educational contexts, geographic locations, and research purposes across the literature.

RESULTS

Characteristics and Distribution of the Evidence Base

The mapped literature demonstrates rapid expansion of research on generative artificial intelligence (GenAI) in English language learning since the public release of ChatGPT in November 2022. The evidence base has developed from predominantly exploratory and perception-oriented studies toward a broader range of empirical designs examining language performance, learner engagement, self-directed learning, feedback, and classroom

implementation. Existing evidence syntheses illustrate this expansion. Lo et al. (2024) identified 70 empirical studies examining ChatGPT in ESL/EFL education during approximately the first 18 months following its release, whereas Li et al. (2025) identified 144 peer-reviewed empirical studies published during 2023–2024. Wang et al. (2025) identified 43 empirical studies published in Social Sciences Citation Index journals, while more recent reviews have continued to demonstrate the rapid expansion of GenAI research in EFL education.

The distribution of the evidence, however, remains uneven across educational settings, geographical regions, research designs, and theoretical approaches. As shown in Table 2, higher education and EFL contexts constitute the dominant settings represented in the recent literature. Li et al. (2025) reported that 86.7% of the 144 empirical studies in their review were conducted in higher education and 86.1% were situated in EFL contexts. Similarly, Wang et al. (2025) found substantial concentration on English language education among Chinese learners and teachers in higher education. In the earlier ChatGPT-focused review by Lo et al. (2024), 48.6% of the included studies were conducted in East Asia and 25.7% in the Middle East.

Table 2. *Characteristics and Distribution of the Evidence Base*

Dimension	Evidence reported in the reviewed literature	Main pattern
Overall empirical evidence	70 empirical studies in Lo et al. (2024); 144 studies in Li et al. (2025); 43 studies in Wang et al. (2025)	Rapid expansion of empirical research
Educational context	86.7% of studies in Li et al. (2025) were conducted in higher education	Strong concentration in universities
Language-learning context	86.1% of studies in Li et al. (2025) concerned EFL	EFL dominates over ESL and other contexts
Geographical distribution	48.6% East Asia and 25.7% Middle East in Lo et al. (2024)	Strong concentration in East Asian and Middle Eastern contexts
Methodological approach	Mixed methods represented 38.9% of studies in Li et al. (2025)	Increasing methodological diversity, with mixed methods prominent
Theoretical grounding	46.5% of studies in Wang et al. (2025) did not report a clear theoretical framework	Significant theoretical underdevelopment
User perceptions	41.9% of studies in Wang et al. (2025) examined user perceptions	Perception-oriented research remains substantial
Technology examined	Early research heavily concentrated on ChatGPT	Increasing movement toward broader GenAI systems

Note. Percentages are reported from the individual evidence syntheses identified in the present review and should not be interpreted as percentages from a single pooled sample.

The methodological profile of the literature indicates increasing diversity as the field has matured. Li et al. (2025) reported that mixed-methods studies represented 38.9% of their 144-study sample, alongside experimental, quasi-experimental, survey, interview, and other empirical approaches. This represents a movement beyond the exploratory and perception-

oriented work that characterized much of the early literature. At the same time, methodological weaknesses remain. Wang et al. (2025) found that 46.5% of the empirical studies in their scoping review did not report a clear theoretical framework. The same review identified inadequate reporting of GenAI model versions and prompts, limiting reproducibility because system outputs can vary according to model version, prompting procedures, system settings, and interaction histories.

Applications of GenAI Across English Language Skills

The distribution of research across English language skills reveals a pronounced concentration on writing. Writing accounted for 41.4% of the 70 studies reviewed by Lo et al. (2024) and 42.4% of the 144 studies examined by Li et al. (2025). Across these studies, GenAI was used for brainstorming, idea generation, outlining, drafting, grammar and spelling correction, vocabulary selection, paraphrasing, organization, revision, and feedback. The consistency of the findings across different evidence syntheses indicates that writing is the most established application of GenAI in English language learning.

Table 3. *Distribution and Major Applications of GenAI Across English Language Skills*

Language skill/domain	Evidence reported	Major GenAI applications	Overall evidence status
Writing	41.4% of Lo et al. (2024); 42.4% of Li et al. (2025)	Brainstorming, drafting, revision, grammar correction, vocabulary, organization, feedback	Strongest and most extensively researched
Speaking	5 studies; 7.1% in Lo et al. (2024)	Conversational practice, simulated dialogue, communicative interaction, voice-based practice	Emerging but substantially smaller evidence base
Reading	2 studies in Lo et al. (2024)	Summarization, passage explanation, vocabulary clarification, question generation, text adaptation	Limited evidence
Listening	No specific studies in Lo et al. (2024)	Potential spoken-input generation, interactive listening, multimodal practice	Major evidence gap
Vocabulary	Reported across multiple studies/reviews	Definitions, examples, collocations, contextualized vocabulary, individualized practice	Moderate but less extensively studied
Grammar	Reported across multiple studies/reviews	Grammar explanations, correction, examples, customized exercises	Moderate but less extensively studied
Pronunciation	Limited emerging evidence	Potential voice-based practice and feedback	Under-researched

Note. The numerical figures in this table are derived from the individual reviews cited in the manuscript and do not represent a pooled frequency analysis conducted by the present study.

Speaking represents the second most visible skill area, although the evidence remains substantially smaller. Lo et al. (2024) identified five studies specifically addressing speaking, equivalent to 7.1% of their sample. These studies examined conversational interaction, simulated dialogues, and opportunities for communicative practice. More recent research has

expanded this area through voice-enabled and multimodal GenAI systems, including investigations of spoken-English engagement and conversational practice.

Reading has received considerably less empirical attention. Lo et al. (2024) identified only two studies specifically addressing reading. Reported applications included text summarization, explanation of difficult passages, vocabulary clarification, question generation, and adaptation of texts to different proficiency levels. These uses position GenAI primarily as an adaptive scaffold for reading rather than as a thoroughly established reading intervention.

Listening remains one of the least developed areas. Lo et al. (2024) identified no studies specifically addressing listening in their 70-study sample. Although subsequent developments in voice-enabled and multimodal GenAI have created opportunities for spoken-input generation, interactive listening activities, and simulated conversations, the empirical evidence concerning listening comprehension remains limited.

Vocabulary and grammar occupy an intermediate position. Learners have used GenAI to obtain definitions, contextualized examples, collocations, explanations of grammatical rules, corrective feedback, translation assistance, and customized exercises. However, compared with writing, the evidence remains relatively limited, particularly regarding controlled studies and delayed assessments of retention.

The evidence therefore reveals a clear imbalance across language skills. Writing is the dominant research area, speaking is increasingly emerging, vocabulary and grammar receive moderate attention, and reading and listening remain comparatively under-researched. This pattern indicates that the current evidence base provides considerably stronger support for understanding GenAI-assisted written language development than its contribution to integrated communicative competence.

Pedagogical Uses and Learning Outcomes

Beyond specific language skills, the literature identifies several recurring pedagogical functions of GenAI. The most prominent uses include individualized feedback, writing assistance, personalized learning, conversational practice, self-directed learning, vocabulary and grammar support, and instructional material development. These functions are summarized in Table 4.

Table 4. *Major Pedagogical Uses, Reported Outcomes, and Associated Concerns*

Pedagogical use	Typical application	Reported outcomes	Important qualification
Writing assistance	Brainstorming, drafting, revision, paraphrasing, editing	Improved writing performance and support during the writing process	Improved AI-assisted products may not equal improved independent writing ability
Feedback and assessment	Grammar, vocabulary, organization, clarity, and style feedback	Immediate and repeated feedback; potential writing improvement	AI feedback may contain inaccurate or inappropriate suggestions
Conversational practice	Simulated dialogue, chatbot interaction, voice-based conversation	Increased opportunities for speaking practice and engagement	Long-term effects on spontaneous speaking proficiency remain uncertain

Pedagogical use	Typical application	Reported outcomes	Important qualification
Personalized learning	Individualized explanations, exercises, examples, and task adaptation	Increased learning opportunities and perceived usefulness	Effectiveness depends on learner interaction and AI literacy
Self-directed learning	Out-of-class practice, question answering, language exercises	Greater access to autonomous practice	Excessive reliance may weaken learner agency
Vocabulary learning	Definitions, examples, collocations, contextualized use	Immediate linguistic support and individualized practice	Evidence for long-term retention remains limited
Grammar learning	Explanations, correction, examples, customized exercises	Increased access to individualized grammatical support	AI-generated explanations require verification
Teacher/material support	Content generation, activity design, instructional assistance	Potential reduction of preparation demands and expanded instructional resources	Teacher oversight remains necessary

The evidence concerning learning outcomes is generally positive but heterogeneous. Recent intervention studies have reported improvements in writing and related language outcomes. More broadly, a 2025 meta-analysis synthesized 41 experimental or quasi-experimental studies involving 3,515 participants and reported a moderate-to-large overall positive effect of GenAI chatbots on second-language acquisition (effect size = 0.576, 95% CI [0.385, 0.768]). The magnitude of the effects varied according to the learning objective, GenAI tool, learners' first language, and intervention duration. These findings indicate that positive learning outcomes are possible but are not uniform across learners, technologies, or instructional conditions.

Li et al. (2025) similarly reported positive effects of GenAI within task-based language teaching and learning and identified AI literacy, task planning, and assessment practices as factors influencing effectiveness. Taken together, these findings indicate that GenAI functions most effectively when embedded within a pedagogically structured learning activity rather than treated as an independent intervention.

The literature also identifies learner autonomy and self-directed learning as important outcomes. GenAI allows learners to practise English outside formal classroom settings, request individualized explanations, generate exercises, obtain immediate feedback, and simulate conversations. These affordances can expand the temporal and spatial boundaries of English learning. However, the evidence also indicates that autonomy is conditional: GenAI may support autonomy when learners retain control over learning decisions and critically evaluate generated content, but it may contribute to dependency when learners delegate linguistic and cognitive tasks to the system.

Learner Perceptions and Experiences

Learner perceptions constitute a substantial component of the evidence base. Wang et al. (2025) reported that 41.9% of the 43 studies in their scoping review examined user perceptions. Across

the literature, learners generally describe GenAI as convenient, accessible, interactive, and useful for immediate assistance. Positive perceptions are particularly evident in relation to writing feedback, language correction, brainstorming, and individualized practice.

The conversational nature of GenAI is another recurring feature of learner experience. Learners can request clarification, ask follow-up questions, provide additional contextual information, and obtain alternative explanations. This interaction can make language practice more flexible than static instructional resources. Nevertheless, perceived usefulness should not be equated with demonstrated language learning. The literature distinguishes between satisfaction with the technology, perceived usefulness, engagement, and objectively measured language proficiency.

Challenges, Limitations, and Research Gaps

The benefits identified in the literature are accompanied by recurring concerns involving reliability, academic integrity, privacy, dependency, learner agency, unequal access, and teacher preparation. Lo et al. (2024) identified incorrect information, privacy leakage, and academic dishonesty among the principal concerns reported in ESL/EFL research. These concerns are particularly relevant in language learning because learners with limited proficiency may have difficulty determining whether AI-generated explanations, corrections, vocabulary recommendations, or paraphrases are accurate.

Table 5. *Major Challenges and Evidence Gaps Identified in the Literature*

Challenge or gap	Evidence from the review	Implication
Inaccurate or fabricated information	Repeatedly reported in ESL/EFL research	Teacher/learner verification remains necessary
Academic integrity	GenAI can generate substantial portions of assessed work	Assessment design must distinguish assisted from independent performance
Learner dependency	Excessive assistance may reduce independent linguistic production	Pedagogical use should preserve learner agency
Privacy	Privacy concerns identified in previous evidence syntheses	Responsible data practices are required
Unequal access	Technological resources vary across contexts	Findings may not generalize to low-resource settings
Teacher preparation	Insufficient AI-related pedagogical preparation remains a concern	Professional development is needed
Theoretical grounding	46.5% of studies in Wang et al. (2025) lacked a clear theoretical framework	Stronger integration with SLA and learning theories is needed
Model/prompt reporting	Inadequate reporting of GenAI versions and prompts	Reproducibility is limited
Educational concentration	86.7% higher education and 86.1% EFL in Li et al. (2025)	Greater research diversity is required

Challenge or gap	Evidence from the review	Implication
Skill imbalance	Writing dominates while listening and other skills remain under-researched	Future studies should examine integrated language development
Long-term evidence	Relatively limited longitudinal and delayed post-test evidence	Durable language development remains insufficiently established
Independent proficiency	Many studies do not clearly distinguish AI-assisted from unaided performance	Future studies should include independent post-intervention assessment

Academic integrity represents a related concern. GenAI can support legitimate writing development through feedback and revision, but it can also generate substantial portions of assessed work. Consequently, the literature distinguishes between AI-assisted learning and extensive AI-generated production. This distinction has implications for assessment validity because unrestricted AI assistance may make it difficult to determine whether an assessment score reflects independent language ability.

Excessive dependence is another recurring concern. While GenAI can provide useful scaffolding, continuous reliance on automated generation may reduce opportunities for independent linguistic formulation, problem-solving, and critical evaluation. Studies also report concerns about loss of individuality and creativity in writing. The evidence therefore suggests that the educational value of GenAI depends partly on maintaining an appropriate balance between technological assistance and learner agency.

The evidence base itself contains important limitations. Research remains concentrated in higher education and EFL settings, particularly East Asian contexts. Primary and secondary education, vocational education, adult learners, low-resource settings, Africa, South Asia, and Latin America remain comparatively underrepresented. The technological focus has also been concentrated on ChatGPT, while reporting of specific model versions and prompting procedures has often been inadequate.

Overall, the mapped evidence demonstrates a rapidly expanding but uneven research field. GenAI has been associated with positive learning opportunities, improved writing performance, individualized feedback, learner engagement, and self-directed learning. However, the strength of evidence differs substantially across language skills and educational contexts. Writing and higher education dominate the literature, whereas listening, reading, pronunciation, younger learners, vocational contexts, and underrepresented geographical regions remain comparatively neglected. The evidence also indicates that positive outcomes depend on pedagogical design, learner characteristics, AI literacy, task structure, and the extent to which learners remain actively responsible for linguistic and cognitive work.

DISCUSSION

The findings of this scoping review indicate that generative artificial intelligence (GenAI) has rapidly become an important component of the emerging English language learning landscape,

but the evidence base remains uneven in terms of language skills, educational contexts, methodological rigor, and theoretical development. Overall, the findings are consistent with previous evidence syntheses in identifying writing, higher education, and EFL contexts as the dominant areas of investigation, while also demonstrating an increasing movement toward empirical examination of learning outcomes, learner autonomy, multimodal interaction, and pedagogically structured uses of GenAI. The present discussion therefore interprets the findings in relation to previous research rather than treating the mapped studies as isolated evidence.

The Dominance of Writing and the Narrowing of the Concept of Language Learning

One of the clearest findings of this review is the disproportionate concentration of research on writing. This finding closely corresponds with previous reviews. Lo et al. (2024) identified writing as the most extensively investigated language skill, accounting for 41.4% of their 70 empirical ESL/EFL studies. Similarly, Li et al. (2025) reported that writing constituted 42.4% of their substantially larger sample of 144 empirical GenAI studies. Wang et al. (2025), examining 43 SSCI-indexed studies, also identified writing as the dominant language skill investigated in the literature. The consistency of these figures across reviews conducted using different inclusion criteria suggests that the dominance of writing is a structural characteristic of the field rather than an artifact of a particular database search or review methodology.

Several factors may explain this pattern. First, text-based GenAI systems are particularly well suited to written language because they can immediately generate, transform, compare, and evaluate textual material. Writing also provides researchers with tangible products that can be assessed before and after an intervention, making it methodologically easier to investigate than spontaneous speaking or listening. Second, the emergence of ChatGPT created an immediate research opportunity for examining automated writing assistance because the technology could already perform many tasks traditionally associated with writing instruction, including brainstorming, drafting, paraphrasing, editing, and feedback.

However, the dominance of writing also presents a conceptual limitation. English language proficiency encompasses productive and receptive skills, interactional competence, pronunciation, vocabulary, grammatical knowledge, and pragmatic competence. If approximately two-fifths of the empirical literature concentrates on writing, the evidence cannot be interpreted as representing the effects of GenAI on English language learning as a whole. This concern is reinforced by Lo et al. (2024), who identified only five studies addressing speaking and no studies specifically addressing listening in their initial sample. The subsequent expansion of voice-enabled and multimodal GenAI has created opportunities to address this imbalance, but the empirical literature has not yet developed at the same pace as the technology.

The findings consequently suggest that future research should move beyond the question of whether GenAI can improve writing and examine whether it can contribute to integrated language development. In particular, longitudinal research combining writing, speaking, listening, reading, vocabulary, and grammatical outcomes would provide a more defensible account of GenAI's contribution to overall English proficiency.

From Perceived Usefulness to Demonstrated Language Learning

Another important finding is the distinction between learners' positive perceptions of GenAI and demonstrable improvements in language proficiency. This distinction is particularly important because perception studies constitute a substantial proportion of the literature. Wang et al. (2025), for example, found that 41.9% of the studies in their scoping review examined user perceptions. Learners commonly describe GenAI as useful, convenient, interactive, accessible, and capable of providing immediate assistance. These findings are broadly consistent with earlier research on technology acceptance and with studies specifically examining ChatGPT-supported language learning.

Positive perceptions are pedagogically relevant, but they should not be interpreted as direct evidence of learning. A learner may report that ChatGPT is highly useful because it corrects an essay quickly or generates an appropriate vocabulary item, while the learner's independent ability to perform the same task may remain unchanged. This distinction was already evident in the early literature reviewed by Lo et al. (2024), who noted that much of the initial research was exploratory and called for more rigorous empirical investigations of actual learning outcomes.

The more recent experimental evidence provides a more encouraging picture. A 2025 meta-analysis of 41 experimental and quasi-experimental studies involving 3,515 participants found a moderate-to-large overall positive effect of GenAI chatbots on second-language acquisition (effect size = 0.576, 95% CI [0.385, 0.768]). Importantly, however, the effects varied according to the learning objective, GenAI tool, learners' first language, and intervention duration. This finding is significant because it challenges the simplistic assumption that GenAI itself is responsible for learning gains. Instead, the effectiveness of GenAI appears to be contingent upon the interaction between the technology, learner, task, and instructional environment.

Similarly, Li et al. (2025) found positive effects of GenAI within task-based language teaching and learning while identifying AI literacy, task planning, and assessment practices as moderators. This is particularly important for interpreting the findings of the present review. The evidence increasingly suggests that GenAI should be conceptualized as a pedagogical mediator rather than as an independent educational intervention. Its effectiveness depends on how teachers structure activities, how learners interact with the system, and whether the task requires learners to think, evaluate, revise, and produce language themselves.

GenAI as Scaffolding Rather Than Replacement

The findings concerning writing, feedback, vocabulary, grammar, and self-directed learning collectively support the interpretation of GenAI as a form of linguistic scaffolding. This interpretation is consistent with research demonstrating that ChatGPT can support learners during different stages of the writing process, including idea generation, drafting, revision, and language correction. Mahapatra (2024), for example, found evidence supporting ChatGPT's potential as a formative feedback tool for undergraduate ESL writers.

The significance of this finding lies in the distinction between assistance and substitution. When a learner asks GenAI to explain why a sentence is grammatically incorrect and then revises the sentence independently, the technology can function as scaffolding. Similarly, when

a learner asks for alternative vocabulary and evaluates which expression is most appropriate for a particular context, the learner remains cognitively engaged. In contrast, when the learner simply requests a complete essay and submits the generated text, the technology may replace rather than support the learning activity.

This distinction also helps explain why findings across individual studies may appear contradictory. Studies that integrate GenAI into structured pedagogical activities can report meaningful learning gains, whereas unrestricted use may produce improvements in the quality of the immediate product without equivalent development of underlying competence. The recent meta-analytic evidence showing that intervention characteristics and duration moderate outcomes supports this interpretation.

The Issue is particularly important for writing. A polished AI-assisted essay can create the appearance of substantial language development even when the learner has contributed relatively little linguistic production. Consequently, future research should distinguish between AI-assisted performance and unaided performance. Post-intervention assessments in which learners' complete comparable tasks without GenAI would provide stronger evidence that observed improvements represent genuine learning rather than temporary technological assistance.

Learner Autonomy: Expansion or Dependency?

The findings also reveal a complex relationship between GenAI and learner autonomy. On one hand, GenAI substantially expands opportunities for autonomous language learning. Learners can practise English outside the classroom, request explanations at any time, receive individualized feedback, generate exercises, simulate conversations, and adapt tasks to their proficiency level. This corresponds with Lo et al.'s (2024) identification of personalized learning and increased learning opportunities as important affordances of ChatGPT in ESL/EFL education.

On the other hand, increased access to assistance does not necessarily produce greater autonomy. Genuine learner autonomy requires learners to make decisions, monitor their performance, evaluate information, and take responsibility for learning outcomes. If GenAI performs these cognitive activities instead, frequent use may create dependency. The learner may become increasingly efficient at obtaining answers while becoming less capable of independently producing or evaluating language.

This tension is particularly important in EFL environments where opportunities for authentic English interaction may already be limited. GenAI can provide additional opportunities for practice, but learners must also develop the ability to determine when AI assistance is appropriate and when independent performance is necessary. The concept of AI literacy therefore becomes central. The findings of Li et al. (2025), which identified AI literacy as a moderator of GenAI effectiveness, reinforce the argument that successful integration requires more than technical access. Learners need the ability to formulate effective prompts, evaluate outputs, identify errors, protect personal information, and recognize situations in which AI assistance may compromise the learning objective.

Reliability, Academic Integrity, and the Problem of Trust

The positive pedagogical possibilities identified in this review are accompanied by significant concerns concerning reliability and academic integrity. Lo et al. (2024) identified incorrect information, privacy leakage, and academic dishonesty among the principal challenges associated with ChatGPT in ESL/EFL education. These concerns remain relevant as GenAI systems become increasingly capable because fluent output can make incorrect information difficult for language learners to recognize.

This creates a particular problem in English language education. Learners frequently use GenAI precisely because they are uncertain about linguistic accuracy. A learner with limited English proficiency may therefore be poorly positioned to evaluate whether an AI-generated grammatical explanation, vocabulary recommendation, or paraphrase is correct. Consequently, the technology can create a paradox in which learners receive more immediate feedback but may have less capacity to determine whether that feedback is trustworthy.

Academic integrity represents a related but distinct problem. GenAI can assist legitimate writing development, but it can also generate substantial portions of assessed work. The pedagogical distinction between these uses cannot be resolved simply by prohibiting AI. Instead, assessment practices may need to shift toward processes that make learning visible, including drafts, revisions, oral explanations, reflective accounts of AI use, in-class writing, and tasks requiring learners to justify linguistic choices.

The findings therefore support the position that GenAI should not be treated as an authoritative substitute for teachers. Rather, teacher mediation remains important, particularly when AI-generated information affects assessment, academic writing, or substantive language learning decisions.

Geographical and Educational Concentration of the Evidence

The geographical distribution identified in this review raises concerns regarding the generalizability of current findings. Research is strongly concentrated in East Asian and other EFL higher-education environments. Lo et al. (2024) reported that nearly half of their studies were conducted in East Asia, while Li et al. (2025) found that 86.7% of the studies in their larger review were situated in higher education and 86.1% in EFL contexts.

This concentration is understandable. Universities often have greater technological infrastructure, researchers have immediate access to student participants, and English is frequently a major component of higher education in EFL countries. Nevertheless, the dominance of these contexts means that conclusions about GenAI cannot automatically be generalized to school-aged learners, adult migrants, vocational students, or learners in low-resource educational environments.

The geographical Imbalance is also important from a linguistic perspective. English language learning occurs under highly different sociolinguistic conditions. Learners in multilingual South Asian, African, Middle Eastern, and Southeast Asian contexts may have different educational expectations, levels of digital access, linguistic repertoires, and relationships with English than learners in the contexts that dominate the current literature. Future research should

therefore investigate whether the pedagogical effects identified in highly represented settings remain consistent across different educational and sociolinguistic environments.

Methodological Maturity and Directions for Future Research

The literature demonstrates clear methodological development, but the field has not yet achieved sufficient maturity to support highly definitive conclusions about the long-term educational effects of GenAI. Earlier research was dominated by exploratory investigations and learner perceptions, whereas newer studies increasingly employ experimental, quasi-experimental, mixed-methods, and meta-analytic approaches. This progression is encouraging and corresponds with calls in earlier reviews for more rigorous outcome-based research.

Nevertheless, several methodological problems remain. First, relatively few studies examine long-term retention. Immediate post-tests can establish whether learners perform better shortly after an intervention but cannot determine whether GenAI produces durable changes in language competence. Second, many studies do not adequately distinguish between AI-assisted performance and independent performance. Third, inconsistent reporting of model versions, prompts, interaction procedures, and system configurations creates reproducibility problems. Wang et al. (2025) specifically identified insufficient reporting of theoretical frameworks, GenAI versions, and prompts as weaknesses in the literature.

Future research should therefore employ longer intervention periods, delayed post-tests, control or comparison groups, standardized language-proficiency measures, and transparent documentation of GenAI use. Researchers should also report the specific model and version, prompting procedures, frequency of interaction, teacher involvement, and learner AI-literacy training. Such reporting would make it possible to determine not simply whether GenAI is associated with improved performance but which configurations of GenAI use produce which learning outcomes for which learners.

CONCLUSION

This scoping review mapped the rapidly expanding literature concerning generative artificial intelligence in English language learning. The evidence demonstrates that GenAI has become an important area of research since the public emergence of ChatGPT in late 2022. The literature has developed from predominantly conceptual and perception-oriented work toward increasingly empirical investigations of language performance, feedback, learner engagement, self-directed learning, and classroom implementation.

Seven principal applications emerged: writing assistance, feedback and assessment, conversational practice, vocabulary and grammar learning, reading support, personalized learning, and teacher-supported material development. Among these, writing is overwhelmingly the most extensively investigated domain. Speaking research is growing rapidly, particularly through voice-enabled and conversational GenAI, while listening, pronunciation, integrated communication, and multimodal learning remain comparatively underdeveloped.

The evidence generally indicates that GenAI can support English language learning by increasing opportunities for practice, providing immediate feedback, facilitating personalized

learning, supporting writing development, and encouraging learner engagement. Recent meta-analytic evidence also indicates positive overall effects of AI chatbots on English-learning outcomes, although the magnitude of effects is influenced by intervention duration and the pedagogical role assigned to the chatbot.

However, the evidence does not justify treating GenAI as an autonomous replacement for teachers. Major challenges include hallucinated or inaccurate information, overreliance, academic dishonesty, privacy concerns, bias, unequal access, reduced learner agency, and insufficient teacher preparation. The effectiveness of GenAI appears to depend substantially on how it is integrated into learning activities.

The most important conclusion is therefore that GenAI should be understood as a pedagogical instrument rather than an educational intervention in itself. The technology does not automatically produce language learning. Learning emerges from the interaction among the learner, the task, the teacher, the technological system, and the sociocultural context.

The current research base also remains uneven. Higher education and East Asian EFL contexts dominate the literature, while primary and secondary education, vocational education, low-resource contexts, Africa, South Asia, Latin America, and other regions remain comparatively underrepresented. Writing is similarly overrepresented compared with speaking, listening, reading, and pronunciation. Longitudinal evidence and standardized measures of independent language proficiency are also insufficient.

Future research should consequently prioritize longitudinal and quasi-experimental designs, standardized language assessments, multimodal GenAI systems, younger learners, underrepresented geographical contexts, World Englishes, AI literacy, learner agency, and teacher-mediated pedagogical designs. Researchers should also distinguish between improved performance with AI assistance and durable improvements in independent language competence.

Ultimately, the educational question should not be whether English teachers and learners should use generative AI. That question has largely been overtaken by technological reality. The more consequential question is how GenAI can be incorporated into English language learning in ways that increase meaningful language practice without diminishing human interaction, learner agency, critical thinking, linguistic diversity, and educational integrity. Addressing that question should define the next generation of research in AI-assisted English language learning.

REFERENCES

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

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

Crompton, H., & Burke, D. (2024). The application of generative AI in language teaching and learning: A scoping literature review. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12954-0

Deng, L., & Jamaludin, K. A. (2026). Roles of generative artificial intelligence (GenAI) in English as a foreign language (EFL) instruction: A systematic literature review. *SAGE Open, 16*(1). https://doi.org/10.1177/21582440261418315

Griche, M. S., & Bennis, M. (2026). Generative artificial intelligence integration in foreign language education in higher education: A systematic literature review. *Technology in Language Teaching & Learning, 8,* Article 102827. https://doi.org/10.29140/tltl.2026.102827

Kainta, E., & Umarella, R. F. (2026). Empirical evidence on the use of ChatGPT in developing EFL students' reading comprehension: A systematic review. *English Education: Journal of English Teaching and Research, 11*(1). https://doi.org/10.29407/jetar.v11i1.28305

Lee, S., et al. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence, 7,* 100266. https://doi.org/10.1016/j.caeai.2024.100266

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5,* Article 69. https://doi.org/10.1186/1748-5908-5-69

Li, J., Huang, J., Wu, W., et al. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications, 11,* Article 1268. https://doi.org/10.1057/s41599-024-03755-2

Li, Y., Shadiev, R., & Chiu, T. K. F. (2025). Applying generative artificial intelligence to task-based language teaching and learning: A systematic review and meta-analysis. *TechTrends, 70,* 150–163. https://doi.org/10.1007/s11528-025-01140-7

Li, [et al.]. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence, 9,* 100445. https://doi.org/10.1016/j.caeai.2025.100445

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11,* Article 50. https://doi.org/10.1186/s40561-024-00342-5

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11,* Article 9. https://doi.org/10.1186/s40561-024-00295-9

Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (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*, Article 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Saarela, M., Gunasekara, S., & Kumarage, P. (2026). A meta-analysis of generative AI effects on language proficiency and affective–cognitive outcomes in language learning. *Discover Computing, 29*, Article 116.

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

Tricco, A. C., Lillie, E., Zarin, W., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Uygun, E., & Demir, B. (2026). A PRISMA-based systematic review of artificial intelligence in English as a foreign and second language education (2023–2025). *Discover Education, 5*, Article 478.

Wang, [et al.]. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2025.2509759

Woo, D. J., Wang, D., Guo, K., et al. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies, 29*, 24963–24990. https://doi.org/10.1007/s10639-024-12819-4

Xodabande, I. (2026). A thematic synthesis of emerging research on English language teachers' use of ChatGPT for language teaching. *Discover Education, 5*, Article 477.

Qin, L., et al. (2026). Exploring L2 learner engagement in spoken English practices with a GenAI-powered chatbot. *System, 138*, 103982.

Teng, L. S., et al. (2026). Generative artificial intelligence for language learning and teaching: Opportunities, challenges, and ethical considerations. *System*. https://doi.org/10.1016/j.system.2026.104128