Vol. 7, No. 1, 2026, pp. 1337-1352
DOI: https://doi.org/10.29210/993470

Contents lists available at Elora Center

Lentera Negeri

Journal homepage: https://lentera.eloracenter.org/lentera

Large language models for personalized feedback and communicative skills in english learning: a systematic review

Kristian Burhan*), M. Arie Desman
Universitas Negeri Padang, Indonesia

Article Info

Article history:
Received Aug 22th, 2026
Revised Aug 24th, 2026
Accepted Aug 26th, 2026

Keyword:
Generative artificial intelligence,
Large language models,
English language learning,
Automated feedback,
Personalized learning

ABSTRACT

Artificial intelligence is increasingly embedded in English language learning, but evidence on generative AI and large language models (LLMs) remains fragmented across skills, learner groups, and tool designs. This systematic review synthesized evidence on personalized learning, automated feedback, speaking, writing, vocabulary, assessment, English for specific purposes (ESP), and physical education. Reporting followed PRISMA 2020. Scopus, PubMed, and ScienceDirect returned 403 records; 14 duplicates were removed, 389 were screened, 54 reports were sought, 48 full texts were assessed, and 17 studies were included. The corpus spanned experimental, quasi-experimental, comparative, mixed-methods, qualitative, survey, review, and system-development designs. Where reported, participant samples included 40 undergraduates, 79 graduate students, and 327 primary pupils, while several design and qualitative studies did not report a single numeric sample. Narrative/thematic synthesis identified four themes: feedback and writing support; speaking and personalized tutoring; changing learner-teacher roles; and affective and ethical conditions. The evidence is recent and heterogeneous, limiting quantitative pooling. FICO was used as an internal appraisal rubric; reviewer-level logs were not retained, so Cohen's kappa and aggregate FICO scores were not reconstructed retrospectively. Future research should prioritize controlled longitudinal designs, validated outcome measures, transparent reporting, data privacy, and under-represented ESP and sport-science populations.

Corresponding Author:
Kristian Burhan,
kristianburhan@unp.ac.id

Introduction

Artificial intelligence now shapes contemporary education in ways that are hard to ignore, and the spread of accessible generative systems has only accelerated its adoption across language classrooms worldwide (Al-khresheh, 2024; Bin-Hady et al., 2024; Xu & Wang, 2024). Generative artificial intelligence (GenAI) and large language models (LLMs) can produce fluent text, converse in natural language, and deliver targeted feedback at scale, which is what makes them so attractive as tools for teaching and learning. These systems are moving into mainstream practice quickly. Educators and researchers therefore face an urgent question: not just whether these tools work, but where, for whom, and under what conditions they add value (Ekizer, 2025; Wang & Liu, 2026).

English holds a distinctive place in this transformation. It functions at once as a global lingua franca, a gatekeeper for academic and professional mobility, and a core school subject in most education systems (Yu

et al., 2025; Aung & Nikolov, 2025). Learners invest heavily in English for high-stakes examinations, higher education, and career advancement. Yet many still struggle with productive skills such as speaking and writing, and with sustaining motivation over time (Zhang et al., 2025; Richardson & Bowen, 2023). GenAI speaks directly to that problem, offering individualized, on-demand support in exactly these effortful domains.

A growing body of research now documents AI applications in English learning, from intelligent tutoring and digital game-based tasks to corpus-driven and multimodal designs (Nitisakunwut et al., 2022; Rahmanu & Molnár, 2024; Huang, M., 2025). The reported effects on achievement are generally positive but uneven, and the way interventions are designed and evaluated varies a great deal (Ekizer, 2025; Arthi & Gandhimathi, 2025). For practitioners, that heterogeneity is a real problem. Turning isolated findings into coherent instructional decisions is no small task.

Recent advances have shifted attention away from earlier machine-learning applications and toward generative systems that can hold dialogue, scaffold composition, and adapt to learner input in real time (Rashid et al., 2026; Wang, Y., et al., 2025). Studies increasingly treat ChatGPT and comparable LLMs as writing partners, conversation coaches, or sources of formative feedback. They also ask how learners and teachers perceive and adopt these tools (Wang & Liu, 2026; Wang, Q., et al., 2025). Consolidated, review-level evidence hasn't kept pace with that shift.

The first gap concerns fragmentation across sub-domains. Individual studies typically address a single skill, tool, or population, and they rarely situate their findings within an integrated map of the field (Huang, C., et al., 2025; Le et al., 2025). Teacher-facing and learner-facing strands of research often proceed in parallel rather than in conversation. The field therefore lacks a synthesis that connects application domains, pedagogical roles, and reported outcomes (Imran et al., 2024; Jin, 2024).

A second gap is methodological and affective. Many investigations are small, cross-sectional, and design-oriented, and outcome measures differ so widely that comparability and cumulative insight stay limited (Ding & Xue, 2025; Roy, 2024). At the same time, learner-affective variables such as anxiety, enjoyment, and self-efficacy appear to shape whether AI support actually translates into learning. Yet these mechanisms are theorized unevenly across the literature (Yan et al., 2025; Susoy, 2026).

Existing syntheses already indicate that AI-supported language learning can improve some learning outcomes, but they also show substantial variation in intervention type, learner population, and outcome measurement. Rather than assume a common effect size across unlike studies, the present review addresses a different gap: it integrates GenAI/LLM applications across feedback, writing, speaking, personalized tutoring, assessment, ESP, and teacher-facing use while examining learner/teacher roles, affective conditions, and methodological limitations. This cross-domain framing complements prior domain-specific syntheses.

These gaps carry particular urgency for English-for-specific-purposes (ESP) and professional contexts, including physical education and sport sciences. Communicative competence in discipline-specific English is essential there, yet mainstream tools under-serve it (Bessadok & Hersi, 2025; Pham & Abdullah, 2025). A timely, PRISMA-guided synthesis is needed to consolidate what is known, expose what is missing, and guide both practice and future inquiry across general and specialized English settings (Abbasi-Sosfadi et al., 2025; Pan et al., 2025; Baskara & Mukarto, 2023; Qassrawi & Al Karasneh, 2025).

Objective and methodological scope. This systematic literature review maps GenAI and LLM-based tools used in English learning, synthesizes reported learning, feedback, performance, and perception outcomes, and evaluates the methodological maturity of the evidence. The review uses PRISMA 2020 reporting, three databases (Scopus, PubMed, and ScienceDirect), an English-language screening frame covering 2015–2025, PICOS eligibility criteria, and narrative/thematic synthesis. The included corpus spans experimental and quasi-experimental comparisons, mixed-methods, qualitative, survey, review, and system-development studies.

Accordingly, this review pursues three research questions. The first maps the landscape of application so that educators can see, at a glance, which English-learning domains GenAI has actually addressed and with what reported effects (Zhen, 2025; Wang, Q., et al., 2025).

RQ1: In which domains of English language learning have generative AI and LLM-based tools been applied, and what effects on learning, feedback, and performance have been reported?

The second question turns from what to how. It examines the pedagogical designs and the learner and teacher roles that characterize GenAI integration, along with how these stakeholders perceive and adopt the technology (Chen et al., 2026; Khoso et al., 2025).

RQ2: What pedagogical designs and learner, teacher roles characterize the integration of GenAI into English learning, and how do learners and teachers perceive these tools?

The third question appraises the maturity of the evidence itself and charts a forward agenda. It delivers something prior single-domain reviews have not: an integrated, cross-domain, methodologically critical account of GenAI in English learning (Xu et al., 2025; Dai et al., 2026; Sun & Wu, 2026).

RQ3: What methodological patterns, limitations, and research gaps characterize the current evidence base, and what agenda should guide future work?

Method

Research Design and Framework

An SLR design, a systematic literature review, was adopted for this study. It was the sensible choice because it supports transparent, reproducible identification, appraisal, and synthesis of a defined evidence base (Tranfield et al., 2003; Kitchenham, 2007). Reporting followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement and its explanation and elaboration guidance (Page et al., 2021a, 2021b), which extends the earlier PRISMA formulation (Moher et al., 2009; Liberati et al., 2009). All of the review protocol components were set in advance: research questions, search strategy, eligibility criteria, and synthesis approach.

Search Strategy

Titles, abstracts, and keywords were searched with a structured Boolean strategy. Terms for generative AI and language technologies were combined with terms for English learning and instruction:

("generative AI" OR "large language model" OR "ChatGPT" OR "chatbot" OR "artificial intelligence") AND ("English language learning" OR "English language teaching" OR "EFL" OR "ESL" OR "second language") AND ("feedback" OR "speaking" OR "writing" OR "vocabulary" OR "assessment")

Each database has its own query syntax, so the search had to be adapted accordingly. Truncation and field codes such as TITLE-ABS-KEY in Scopus were used to capture morphological variants. Only English-language documents were included.

Database and Information Sources

We drew from three complementary databases. Scopus covers peer-reviewed literature across disciplines, PubMed indexes health, education, and applied behavioural sciences, and ScienceDirect offers full-text journals across social sciences and technology. The searches ran in 2025, and after exporting the records for screening, the three sources together returned 403 records.

Eligibility Criteria

A PICOS framework (Table 1) underpinned the eligibility structure. The explicit inclusion and exclusion criteria, shown in Table 2, grew directly out of it.

Table 1. PICOS framework guiding eligibility.

Element	Definition applied in this review
Population (P)	English language learners and teachers across school, university, and professional/ESP settings (including physical education and sport).
Intervention (I)	Generative AI and AI-assisted tools (ChatGPT and other LLMs, chatbots, AI coaches, ASR-based systems, automated feedback and assessment tools).
Comparison (C)	Traditional instruction, non-AI digital tools, human feedback, or alternative conditions, where reported.
Outcome (O)	Learning outcomes, feedback quality, speaking/writing performance, vocabulary, engagement, autonomy, and learner/teacher perceptions.
Study design (S)	Empirical studies (experimental, quasi-experimental, mixed-methods, qualitative, survey), technology reviews, and design/development studies.

Table 2. Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
Language	English only	Non-English
Document type	Journal article, review, or full conference paper	Proceedings front matter, editorial, retracted item, or abstract only
Publication period	2015–2025	Before 2015
Focus	Directly addresses AI/GenAI for English language learning or teaching	Unrelated NLP tasks (e.g., sentiment analysis in other languages) or non-English-learning topics
Outcome relevance	Reports a learning, feedback, performance, or perception outcome	No measurable English-learning outcome
Accessibility	Full text retrievable	Full text unavailable

Study Selection Process
Selection followed PRISMA 2020. After duplicate removal, records were screened by title and abstract and potentially eligible reports were assessed in full text against the PICOS criteria. The archived record documents consensus resolution after re-examination, but it does not retain independent reviewer-level screening logs or extraction sheets; therefore, Cohen's kappa cannot be calculated retrospectively without inventing data. This limitation is reported explicitly rather than assigning a post-hoc agreement coefficient.

Quality Assessment, FICO Framework
For appraisal of the included reports we used the FICO framework as a transparent internal rubric covering Focus, Information, Context, and Outcome. It was used to judge methodological adequacy and traceability of the evidence, but it is not a validated risk-of-bias instrument. The archived file does not retain item-level FICO ratings, so aggregate quality scores cannot be reconstructed responsibly. For future updates, a validated design-specific appraisal tool should be added rather than presenting FICO as a standardized risk-of-bias measure.

Data Extraction Procedure
Extraction ran through a structured form. For each study, it captured authors and year, country or context, study design, sample, the AI tool or intervention, the targeted English-learning domain, the outcome measures, and the principal findings. We tabulated everything so cross-study comparison and thematic synthesis had a solid base.

Bibliometric and Thematic Mapping
To characterize the corpus, we carried out descriptive mapping. Publication years, study and document types, thematic foci, and the frequency of salient author keywords across included studies were all covered. These descriptive analyses gave the qualitative synthesis its bearings, and they appear as figures in the Results.

Data Analysis and Synthesis
Findings were synthesized thematically, following Thomas & Harden (2008). Codes drawn from the study findings were grouped into descriptive themes, then organized so the research questions could be answered. Themes came about inductively. We checked them against the source studies and refined them iteratively, making sure they stayed grounded in the evidence.

Reporting and Documentation
Reporting followed the PRISMA 2020 checklist for systematic reviews (Page et al., 2021a). The selection process is documented in the PRISMA 2020 flow diagram (Figure 1). Every included study is reported with complete bibliographic detail.

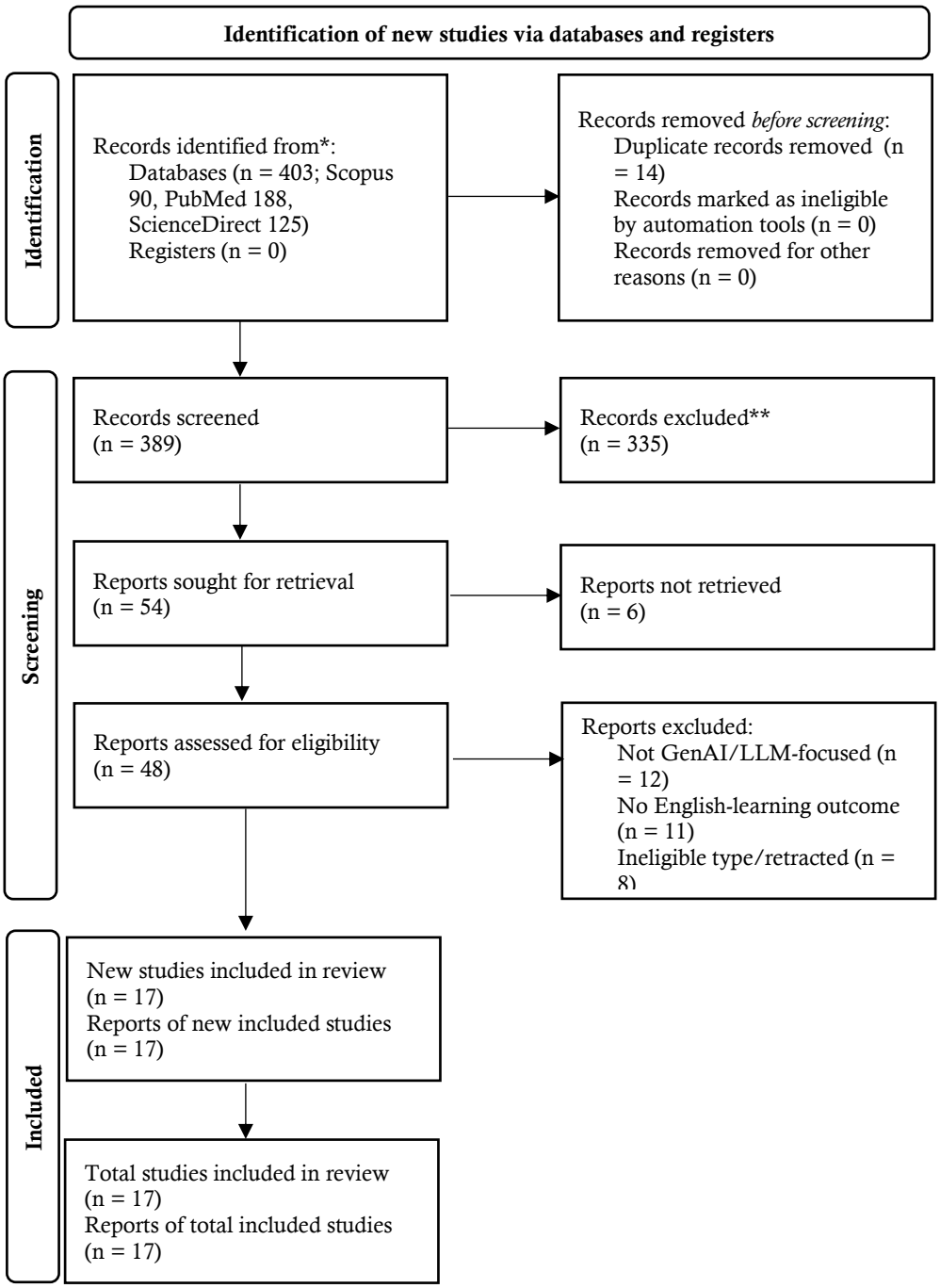

Figure 1. PRISMA 2020 flow diagram of the study selection process (Page et al., 2021a).

Results and Discussions

Study Selection Results

Scopus, PubMed, and ScienceDirect returned 403 records in the structured search. After 14 duplicates came out, 389 went to title and abstract screening. Of those, 335 were rejected. They were plainly out of scope, non-English, or ineligible by document type. Full text was sought for 54 reports. Six could not be retrieved. That left 48 reports for a full eligibility check, and 31 were then excluded. Twelve did not focus on generative AI or LLMs, eleven reported no English-learning outcome, and eight were retracted or of an ineligible type. Seventeen studies cleared every criterion and made it into the synthesis. These figures match the PRISMA flow diagram in Figure 1.

Descriptive Characteristics

Table 3 summarizes the 17 included studies, and Table 4 presents their classification across design, thematic focus, tool, and outcome. This is a recent corpus. Publication activity is heavily concentrated in 2024 and 2025, which together supplied the large majority of included studies (Figure 2). Thematically, three areas dominate: automated feedback and writing support, speaking and oral communication, and personalized learning and intelligent tutoring. Vocabulary and autonomous learning, teacher practice and professional development, and assessment appear next (Figure 3). Author keywords cluster around generative AI and LLMs, feedback, EFL/ESL, ChatGPT, and speaking. Those terms mirror the field's central concerns (Figure 4).

Table 3. Summary of included studies (n = 17).

Author(s) & year	Full title	Method	Country/context	Key findings
Chen & Warden (2019)	Application of Artificial Intelligence to the Small Open Online English Abstract Writing Course	Applied quasi-experimental (SMOOC + AI grading; 79 graduate students)	Not reported	An AI-assisted writing, grading, and feedback system embedded in a small MOOC helped EFL graduate students identify recurring errors and strengthen English abstract writing beyond class-size and time limits.
Mei, Qi, Huang & Huang (2024)	Speeko: An Artificial Intelligence-Assisted Personal Public Speaking Coach	Technology review	Not reported	Reviews the affordances of the Speeko app and argues that AI-generated feedback on oral performance can support EFL learners' public-speaking development.
Wang, Pang, Wallace, Wang & Chen (2024)	Learners' Perceived AI Presences in AI-Supported Language Learning: A Study of AI as a Humanized Agent from Community of Inquiry	Quantitative (327 primary pupils; Community of Inquiry)	Not reported	Learners' perceived social, cognitive, and teaching presences of an AI coach were related to actual L2 outcomes and attitudes, supporting AI as a humanized agent.
Wiboolyasarin et al. (2024)	Designing Chatbots in Language Classrooms: An Empirical Investigation from User Learning Experience	Survey-based empirical (pre-service and in-service teachers, educators)	Thailand	Identified design factors shaping language-teaching chatbots and mapped teacher attitudes to inform interactive, learner-centred chatbot design.
Park et al. (2024)	AI English Conversation Coaching Platform in the Metaverse: Focused on Korean Users	System design and development	South Korea	Proposed a GPT-driven metaverse platform offering scenario-based conversation practice to reduce exam-oriented limitations and first-language interference.
Patturose et al. (2024)	Ingeleze: Advanced AI Solution for English Mastery	System design and development	Not reported	Presented an AI English-tutor app integrating speech recognition, NLP, and chatbot components for personalized, interactive speaking and writing practice.
Niranon & Triyason (2024)	The Efficiency of ChatGPT Vs Google Against Self-Learning of Undergraduate Students	Comparative experimental (40 English and physical-education students)	Not reported	Compared ChatGPT and Google Search for autonomous self-learning among English and physical-education

Author(s) & year	Full title	Method	Country/context	Key findings
				undergraduates, examining relative effectiveness.
Tong (2024)	Assessing English Learning Needs in Sports Physical Education and the Application of Machine Learning Algorithms	Mixed-methods needs analysis with machine learning	China	Identified proficiency gaps (notably listening and speaking) and difficulty with sport-specific terminology among first-year sport physical-education students.
Liu, Huang, Zou & Zhu (2025)	Reconfigure the Classroom: An Activity-Theoretic Analysis of an AI Conversational Agent in Vocational Technical Communication	Mixed-methods pre-/post-test with observation (12 weeks)	Not reported	An AI conversational agent acted as a mediating artefact that reconfigured the vocational ESP classroom and supported technical-communication skill development.
Riser (2025)	Generative Artificial Intelligence (AI) in the High School English Classroom and Its Effect on Students' Identity and Teachers' Practice	Qualitative discourse analysis (videos, interviews, journals)	Not reported	Learners showed greater autonomy when collaboratively evaluating AI text; teachers acted as ethical facilitators balancing AI use, student voice, and integrity.
Remache & Remache (2025)	Arab EFL Students' Engagement with Prompt-Engineered AI Writing Feedback	Mixed-methods (feedback engagement)	Arab region	Prompt-engineered AI approximated rubric-based instructor feedback, and learners engaged with it behaviourally, cognitively, and affectively, easing feedback workload at scale.
Meniado (2025)	Utilizing ChatGPT for Self-Directed Professional Development: Practices of Southeast Asian Language Teachers	Qualitative (teachers from 10 countries)	Southeast Asia	English teachers used ChatGPT to address professional-development barriers such as limited time, workload, and resources, supporting self-directed development.
Smith (2025)	Enhancing ESP Student Critical Thinking Skills and Vocabulary Acquisition Through a GenAI-Based Project	Design and implementation (five-stage project)	Not reported	A GenAI-based multimodal project with sports-science ESP students developed both critical-thinking skills and vocabulary while foregrounding human agency.
Polasa et al. (2025)	Chai-Calibrated Hybrid Assessment for IELTS Speaking with Human-Referenced Validation	System design and evaluation (ASR framework)	Thailand	A dual-agent ASR framework combined pronunciation, prosody, and timing cues to deliver criterion-level IELTS-speaking feedback validated against human references.
Qinyuan (2025)	Design of a University English-Blended Teaching Quality Assessment Model Based on WFKP Clustering Algorithm	Algorithmic model design	Not reported	A weighted clustering model improved data-driven assessment of blended university-English teaching quality through weighted feature selection.

Author(s) & year	Full title	Method	Country/context	Key findings
Karri, Sai & Singh (2025)	Perceptions of Competitive Exam Aspirants in Visakhapatnam on ChatGPT's Role in Vocabulary Acquisition and Autonomous Learning: An ELT Theoretical Perspective	Mixed-methods (learner perceptions)	India	Aspirants perceived ChatGPT as effective for vocabulary development and autonomous learning, aligning its functions with established ELT theories.
Chen, Veng, Wilson, Li & Fang (2025)	CoachGPT: A Scaffolding-Based Academic Writing Assistant	System design and development	Not reported	An LLM-based scaffolding writing assistant supported second-language academic writers beyond rule-based tools by offering contextual, guided feedback.

Table 4. Classification of included studies by design, theme, tool, and outcome.

Author(s) & year	Research design	Thematic focus	Technology/tool	Outcome
Chen & Warden (2019)	Quasi-experimental	Automated feedback & writing	AI writing/grading system (QRP) in SMOOC	Error identification; improved abstract writing
Mei et al. (2024)	Technology review	Speaking & oral communication	Speeko (AI speaking coach)	Feedback affordances for public speaking
Wang et al. (2024)	Quantitative survey	Personalized learning & tutoring	AI coach as humanized agent	Perceived AI presences linked to L2 outcomes
Wiboolyasarin et al. (2024)	Survey-based empirical	Personalized learning & tutoring	Language-teaching chatbots	Design factors; teacher attitudes
Park et al. (2024)	System design	Speaking & oral communication	GPT-based metaverse platform	Scenario-based conversation practice
Patturose et al. (2024)	System design	Personalized learning & tutoring	AI tutor app (ASR + NLP + chatbot)	Personalized speaking/writing practice
Niranon & Triyason (2024)	Comparative experiment	Vocabulary & autonomous learning	ChatGPT vs Google Search	Relative effectiveness for self-learning
Tong (2024)	Mixed-methods + ML	Assessment & quality evaluation	ML modelling of needs	Proficiency gaps; sport-specific terminology
Liu et al. (2025)	Mixed-methods pre/post	ESP / professional communication	AI conversational agent	Reconfigured vocational ESP classroom
Riser (2025)	Qualitative discourse	Teacher practice & development	Generative AI in classroom	Learner autonomy; ethical facilitation
Remache & Remache (2025)	Mixed-methods	Automated feedback & writing	Prompt-engineered AI feedback	Rubric-aligned feedback; learner engagement
Meniado (2025)	Qualitative	Teacher practice & development	ChatGPT for teacher PD	Self-directed professional development
Smith (2025)	Design & implementation	Vocabulary & autonomous learning	GenAI-based multimodal project	Critical thinking; vocabulary (sports-science ESP)
Polasa et al. (2025)	System design & evaluation	Speaking & oral communication	Dual-agent ASR (CHAI)	Criterion-level IELTS-speaking feedback
Qinyuan (2025)	Algorithmic model	Assessment & quality evaluation	WFKP clustering algorithm	Blended-teaching quality assessment
Karri et al. (2025)	Mixed-methods	Vocabulary & autonomous learning	ChatGPT	Vocabulary acquisition; autonomous learning

Author(s) & year	Research design	Thematic focus	Technology/tool	Outcome
Chen et al. (2025)	System design	Automated feedback & writing	CoachGPT (LLM scaffolding)	Contextual academic-writing support

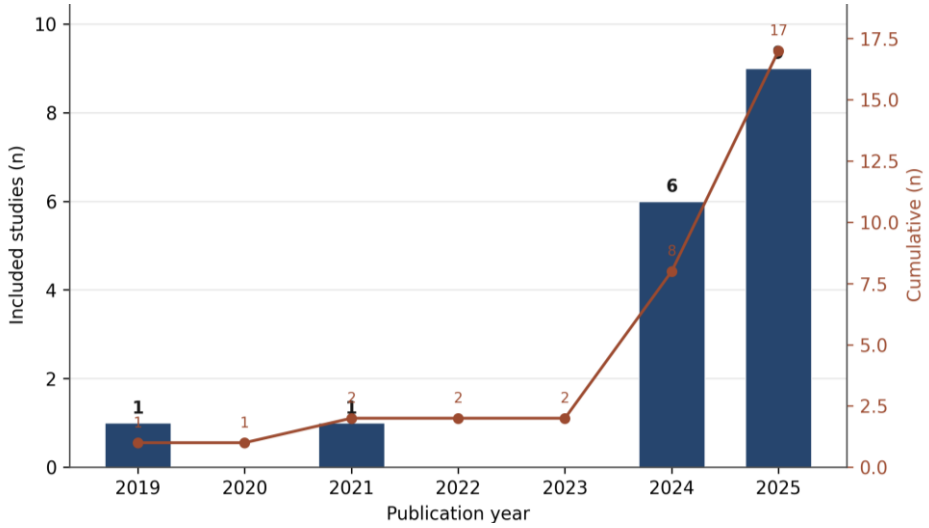

Figure 2. Temporal distribution of included studies by publication year, with cumulative count.

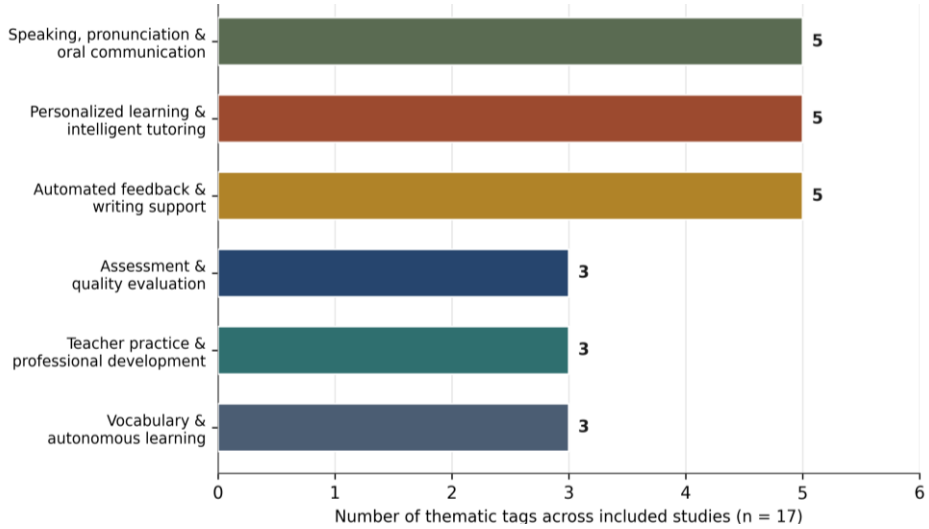

Figure 3. Distribution of thematic foci across included studies (studies may map to more than one theme).

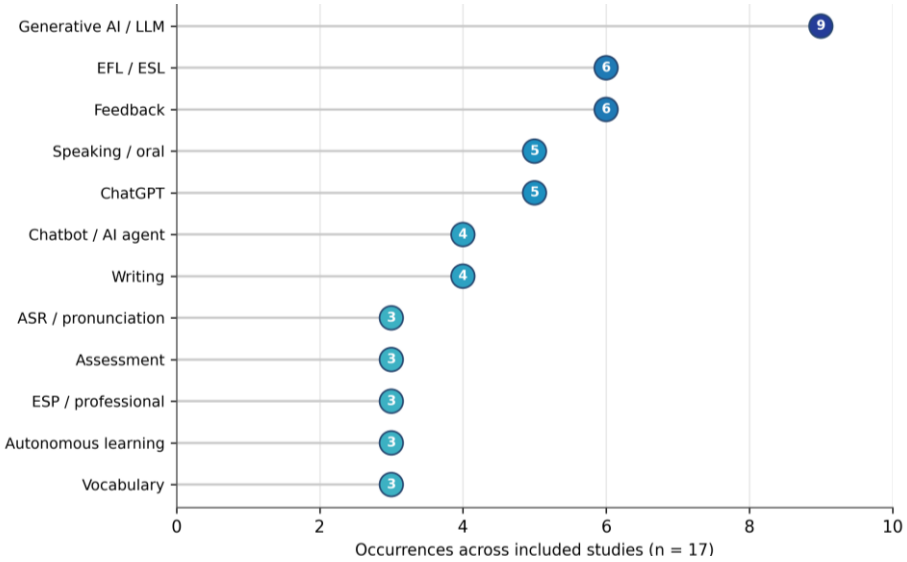

Figure 4. Frequency of salient author keywords across included studies.

Thematic Synthesis

Findings for RQ1: Application domains and reported effects

Read across the whole corpus, GenAI and AI-assisted tools keep landing on a recognizable set of English-learning domains, and automated feedback with writing support is the most prominent of them. The systems themselves span a wide arc: an early AI writing-and-grading environment embedded in a small MOOC (Chen & Warden, 2019), prompt-engineered feedback aligned to instructor rubrics (Remache & Remache, 2025), and LLM-based scaffolding assistants that push second-language academic writing well beyond rule-based grammar checking (Chen et al., 2025). What gets reported, over and over, is the identification of recurring errors, feedback delivered at scale, and improved composition. Put simply, generative feedback can relieve instructor workload and still remain pedagogically meaningful.

Speaking and oral communication form a second, equally substantial cluster. Consider the range of tools: an AI public-speaking coach (Mei et al., 2024), a GPT-driven metaverse conversation platform (Park et al., 2024), an integrated tutor app combining speech recognition with dialogue (Patturose et al., 2024), and an accent-aware ASR framework that delivers criterion-level IELTS-speaking feedback validated against human references (Polasa et al., 2025). The reported effects point to expanded opportunities for authentic practice and fine-grained feedback on pronunciation, fluency, and prosody. Conventional instruction finds those domains notoriously hard to support at scale.

Personalized learning and tutoring, vocabulary and autonomous learning, and assessment complete the map. Learners perceive AI coaches as humanized agents, and the social, cognitive, and teaching presences those coaches project correlate with real outcomes (Wang et al., 2024). Chatbots, on the other hand, hinge on deliberate design choices (Wiboolyasarin et al., 2024). ChatGPT came across as effective for vocabulary and self-directed study (Karri et al., 2025), and comparative work pitted it against conventional search for autonomous learning among English and physical-education students (Niranon & Triyason, 2024). Assessment-oriented studies applied machine learning to model learner needs (Tong, 2024) and to evaluate blended-teaching quality (Qinyuan, 2025). Take the evidence together, and coverage is broad but uneven. Feedback, speaking, and tutoring are the ones best represented.

Findings for RQ2: Pedagogical designs, roles, and perceptions

The included studies converge on a specific way of seeing GenAI. It is not a content-delivery mechanism. It is a mediating agent that reconfigures classroom activity. An activity-theoretic analysis, for instance, showed an AI conversational agent restructuring a vocational ESP course and altering the relationships among learners, teacher, and tools (Liu et al., 2025). In secondary English classrooms, learners exercised greater autonomy when they collaboratively evaluated AI-generated text, while teachers shifted into a new role, that of ethical facilitator safeguarding student voice and academic integrity (Riser, 2025). Broader accounts of teachers negotiating ChatGPT integration in language teaching point the same direction (Al-khresheh, 2024).

Teacher-facing designs extend into professional development, where English teachers across Southeast Asia used ChatGPT to overcome barriers of time, workload, and limited resources (Meniado, 2025). Learner

perceptions and acceptance turn out to be decisive. Motivation mediated behavioural engagement with AI (Wang, Q., et al., 2025), and extended technology-acceptance models captured the psychological determinants shaping GenAI adoption for foreign-language education (Wang & Liu, 2026). Graduate learners' attitudes toward AI-assisted academic writing drive the point home: uptake depends on perceived usefulness and trust (Wang, Y., et al., 2025).

Affective dynamics recur throughout the corpus. AI-driven personalization and emotion recognition shaped engagement by moderating anxiety (Ding & Xue, 2025), AI chatbots reduced speaking-assessment anxiety relative to human facilitation (Susoy, 2026), and personalized foreign-language learning interacted with pleasure, anxiety, and self-efficacy to influence learner well-being (Yan et al., 2025). Studies of Chinese-as-a-foreign-language learners show the same pattern, with generative AI providing cognitive scaffolding for oral script-writing within project-based tasks (Chen et al., 2026). The message is consistent: design and perception, not tool capability alone, decide whether GenAI support actually becomes learning.

Findings for RQ3: Methodological patterns, limitations, and gaps

Methodologically, the corpus is dominated by recent, small-scale, design-oriented work. Several reports are system-design or development studies (Park et al., 2024; Patturose et al., 2024; Chen et al., 2025), others are single-cohort mixed-methods or survey investigations (Remache & Remache, 2025; Karri et al., 2025). Only a minority employed pre-/post-test or comparative designs (Liu et al., 2025; Niranon & Triyason, 2024). Outcome measures are heterogeneous, running from perception scales to performance indices to system-evaluation metrics, and that heterogeneity limits direct comparison while precluding meta-analytic pooling.

Three gaps follow from this. First, longitudinal and adequately controlled evidence is scarce, so durability and causal claims remain tentative, a worry that broader AI-effectiveness syntheses have already voiced (Xu & Wang, 2024; Ekizer, 2025). Second, learner populations and disciplinary contexts are unevenly represented. ESP and professional English, physical education and sport included, are addressed by only a few studies (Smith, 2025; Tong, 2024), despite strong demand for discipline-specific communicative competence (Bessadok & Hersi, 2025). Third, ethical and data-privacy considerations are under-examined, even as learners voice concern about how AI language models handle their data (Xu et al., 2025) and dependency-related risks surface alongside the benefits (Khoso et al., 2025). Emerging frameworks for ethical GenAI integration in writing offer a starting point for addressing these concerns (Rashid et al., 2026).

Comparative and Critical Analysis

Comparing approaches across the corpus reveals a field in transition. The earlier, assessment-oriented studies leaned on machine-learning and clustering techniques (Tong, 2024; Qinyuan, 2025), while the most recent work foregrounds generative dialogue, scaffolding, and multimodal interaction (Chen et al., 2025; Smith, 2025). Design and development studies show technical feasibility and rich affordances. What they often stop short of is rigorous learning-outcome evaluation. Empirical studies with learners provide the outcome evidence, yet typically at small scale and over short horizons. That division of labour helps explain why enthusiasm about capability runs ahead of confidence about efficacy. Bridging it will require studies pairing well-engineered tools with robust, comparative outcome designs.

Discussion

Methodological limitations and boundary conditions

Methodological recommendations

Future studies should: (1) pre-register review protocols and preserve reviewer-level screening and extraction logs; (2) use validated, construct-specific outcome instruments and report reliability statistics; (3) conduct longitudinal or controlled quasi-experimental evaluations with transparent sample-size justification and subject-level validation where AI models are trained; (4) report intervention fidelity, missing data, and model performance using comparable metrics; and (5) test GenAI applications in under-represented ESP, physical-education, and sport-science populations while documenting access, privacy, and equity safeguards.

The review should be interpreted with several methodological boundaries in mind. First, the evidence base is heterogeneous: the 17 studies combine empirical comparisons, surveys, qualitative analyses, technology reviews, and system-development reports, so a common pooled effect size is not defensible. Second, several studies rely on self-reported perceptions or small single-cohort samples, which introduces measurement and generalizability concerns. Third, six full texts could not be retrieved, creating possible retrieval bias. Fourth, the review was limited to English-language records and three databases, so non-English and non-indexed studies may have been missed. Fifth, no prospective registration number or independent reviewer-level logs were retained in the archived record; consequently, Cohen’s kappa and aggregate inter-

rater agreement cannot be reconstructed retrospectively. These limitations are treated as evidence-quality constraints rather than grounds for overstating the effectiveness of GenAI.

Interpreted as a whole, the evidence indicates that GenAI is most useful in English learning where it supplies timely, individualized feedback and expanded practice in productive skills, writing and speaking above all. This reading extends prior reviews, which reported positive but heterogeneous AI effects on English achievement (Ekizer, 2025; Xu & Wang, 2024), by specifying the domains and mechanisms through which benefits arise. Theoretically, the recurrent framing of AI as a mediating, humanized agent connects these findings to activity theory and to community-of-inquiry perspectives, where social, cognitive, and teaching presences jointly shape learning (Wang et al., 2024; Liu et al., 2025).

Practically, the synthesis offers actionable guidance. Educators can deploy GenAI to scale formative feedback and conversational practice, but they should design deliberately, cultivate learners' evaluative and prompting skills, and preserve the teacher's role as ethical facilitator (Riser, 2025; Al-khresheh, 2024). For professional and ESP settings such as physical education and sport, discipline-specific tasks that build terminology and communicative confidence seem especially promising (Smith, 2025; Tong, 2024). These recommendations resonate with wider evidence on teacher readiness and technology integration in English instruction (Le et al., 2025; Imran et al., 2024).

The findings also expose contradictions. AI personalization can reduce anxiety and support well-being in some settings (Ding & Xue, 2025; Susoy, 2026), yet foster dependency and fear of missing out in others (Khoso et al., 2025). Enthusiasm about autonomy coexists with unresolved concerns about data privacy and integrity (Xu et al., 2025; Riser, 2025). Motivation and acceptance repeatedly moderate outcomes, so the same tool can help or hinder depending on learner disposition and context (Wang, Q., et al., 2025; Zhen, 2025; Wang & Liu, 2026). Situating these results against motivation- and affect-focused reviews clarifies a key point: GenAI operates through, rather than around, established learner variables (Aung & Nikolov, 2025; Derakhshan et al., 2022).

Relative to prior single-domain reviews and bibliometric analyses (Arthi & Gandhimathi, 2025; Rahmanu & Molnár, 2024; Huang, C., et al., 2025), this review contributes an integrated, cross-domain map coupled with a critical appraisal of methodological maturity. Three research gaps are salient: the scarcity of longitudinal and controlled designs; the under-representation of ESP, professional, and non-Western populations; and the thin treatment of ethics, privacy, and equity. Corpus-driven and blended-implementation methods point to feasible ways of strengthening evaluation in future designs (Huang, M., 2025; Pham & Abdullah, 2025; Nitisakunwut et al., 2022).

Three limitations of this review should be acknowledged. First, the search was confined to three databases and to English-language documents, so relevant work in other venues or languages may have been missed. Second, the qualitative synthesis reflects interpretive judgement and was not accompanied by meta-analysis, given outcome heterogeneity. Third, the included corpus is recent and fast-moving, so these conclusions represent a snapshot of a rapidly evolving field. The future research agenda follows directly: conduct longitudinal, adequately powered, and comparative studies; standardize and report outcome measures transparently; and embed ethical, privacy, and equity safeguards while extending inquiry to under-served learners and disciplinary contexts, sport and physical education included (Dai et al., 2026; Sun & Wu, 2026; Bin-Hady et al., 2024).

In direct response to the research questions: for RQ1, GenAI has been applied most extensively to automated feedback and writing, speaking and oral communication, and personalized tutoring, with reported gains in engagement, autonomy, and skill-specific performance; for RQ2, effective integration reconfigures learner and teacher roles and depends on deliberate design and on learner motivation, acceptance, and affect; and for RQ3, the evidence base is recent, small-scale, and methodologically heterogeneous, with pressing needs for longitudinal rigor, broader populations, and ethical safeguards.

Conclusions

Seventeen studies were synthesized using PRISMA 2020 to map GenAI and LLM applications in English language learning. RQ1 indicates that automated feedback and writing, speaking and oral communication, and personalized tutoring are the most recurrent application domains. RQ2 shows that GenAI integration reshapes learner and teacher roles and that engagement, motivation, anxiety, autonomy, and perceived usefulness influence outcomes. RQ3 indicates that the evidence base remains recent and methodologically

heterogeneous, with substantial use of system-design, survey, qualitative, and small comparative studies. Because the included studies do not share sufficiently comparable designs, outcomes, or effect measures, no quantitative meta-analysis was conducted. The review is also limited by English-language screening, three databases, six unretrieved reports, and the absence of retained reviewer-level logs and a prospective registration record. The evidence therefore supports cautious pedagogical use and feasibility rather than universal claims of effectiveness. Future research should prioritize controlled longitudinal designs, validated outcome measures, transparent reviewer procedures, psychometrically grounded evaluation, and stronger representation of ESP, physical-education, and sport-science contexts.

Acknowledgments

The Authors would like to thank Faculty of Sport Sciences, Universitas Negeri Padang for providing excellent facilities, as well as for being very supportive of this research from the beginning. Authors also thank the team members who had generously shared their brilliant ideas, engaging discussion, and academic supports during the study.

References

Abbasi-Sosfadi, S., Davoudi, M., Amirian, S. M. R., & Zareian, G. (2025). Development and validation of the critical thinking in academic reading scale for English language teaching students (CTARS-ELT). Thinking Skills and Creativity, 56, 101776. https://doi.org/10.1016/j.tsc.2025.101776

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. Computers and Education: Artificial Intelligence, 6, 100218. https://doi.org/10.1016/j.caeai.2024.100218

Arthi, M. P., & Gandhimathi, S. N. S. (2025). Research trends and network approach of critical thinking skills in English language teaching – A bibliometric analysis implementing R studio. Heliyon, 11, e42080. https://doi.org/10.1016/j.heliyon.2025.e42080

Aung, M. M., & Nikolov, M. (2025). Tertiary students' English language learning motivation: A systematic review. Acta Psychologica, 261, 105854. https://doi.org/10.1016/j.actpsy.2025.105854

Baskara, R., & Mukarto. (2023). Exploring the implications of ChatGPT for language learning in higher education. Indonesian Journal of English Language Teaching and Applied Linguistics, 7(2), 343–358. https://doi.org/10.21093/ijeltal.v7i2.1387

Bessadok, A., & Hersi, M. (2025). A structural equation model analysis of English for specific purposes students' attitudes regarding computer-assisted language learning: UTAUT2 model. Library Hi Tech, 43, 36. https://doi.org/10.1108/LHT-03-2023-0124

Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., & Ali, J. K. M. (2024). Exploring the dimensions of ChatGPT in English language learning: A global perspective. Library Hi Tech, 43, 1315. https://doi.org/10.1108/LHT-05-2023-0200

Chen, F., Veng, S., Wilson, J., Li, X., & Fang, H. (2025). CoachGPT: A scaffolding-based academic writing assistant. SIGIR 2025 – Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 4051–4055. https://doi.org/10.1145/3726302.3730143

Chen, J. F., & Warden, C. A. (2019). Application of artificial intelligence to the small open online English abstract writing course. Lecture Notes in Computer Science, 11937, 802–808. https://doi.org/10.1007/978-3-030-35343-8_83

Chen, X., Zhou, X., & Goh, Y. S. (2026). How Chinese as a foreign language learners use generative AI for oral script-writing: A qualitative perspective on cognitive scaffolding in project-based learning. Acta Psychologica, 262, 106071. https://doi.org/10.1016/j.actpsy.2025.106071

Dai, X., Wang, Y., Yu, J., Qiu, B., Wang, R., & Na, M. (2026). Motivation and engagement as pathways: How AI-augmented online assessment shapes English-speaking competency in vocational EFL classrooms. Frontiers in Psychology, 16, 1730953. https://doi.org/10.3389/fpsyg.2025.1730953

Derakhshan, A., Dewaele, J. M., & Azari Noughabi, M. (2022). Modeling the contribution of resilience, well-being, and L2 grit to foreign language teaching enjoyment among Iranian English language teachers. System, 109, 102890. https://doi.org/10.1016/j.system.2022.102890

Ding, Z., & Xue, W. (2025). Navigating anxiety in digital learning: How AI-driven personalization and emotion recognition shape EFL students' engagement. Acta Psychologica, 259, 105466. https://doi.org/10.1016/j.actpsy.2025.105466

Ekizer, F. N. (2025). Exploring the impact of artificial intelligence on English language teaching: A meta-analysis. Acta Psychologica, 260, 105649. https://doi.org/10.1016/j.actpsy.2025.105649
Huang, C., Xu, W., & Shen, K. (2025). Mapping the landscape of TPACK in English language learning. International Journal of Web-Based Learning and Teaching Technologies, 20. https://doi.org/10.4018/IJWLTT.373233
Huang, M. (2025). Fuzzy decision support system for English language teaching with corpus data. Egyptian Informatics Journal, 29, 100612. https://doi.org/10.1016/j.eij.2025.100612
Imran, M., Almusharraf, N., Sayed Abdellatif, M., & Ghaffar, A. (2024). Teachers' perspectives on effective English language teaching practices at the elementary level: A phenomenological study. Heliyon, 10, e29175. https://doi.org/10.1016/j.heliyon.2024.e29175
Jin, S. (2024). Optimizing English teaching: ARCS motivation model and task-based language teaching in university. Learning and Motivation, 87, 102028. https://doi.org/10.1016/j.lmot.2024.102028
Karri, S. K., Sai, B. S., & Singh, P. K. (2025). Perceptions of competitive exam aspirants in Visakhapatnam on ChatGPT's role in vocabulary acquisition and autonomous learning: An ELT theoretical perspective. Discover Education, 4(1), Article 445. https://doi.org/10.1007/s44217-025-00862-3
Khoso, A. K., Honggang, W., Younas, M., & Salam El-Dakhs, D. A. (2025). The dual forces of AI: How generative AI and perceived AI dependency influence fear of missing out (FoMO) and EFL students' vocabulary acquisition. Journal of Visualized Experiments, 222, e69637. https://doi.org/10.3791/69637
Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering (EBSE Technical Report No. EBSE-2007-01). Keele University and University of Durham.
Le, H. V., Huynh, D. T., & Nguyen, T. T. (2025). Technology integration in English language teaching: Assessing the digital readiness of Vietnamese high school teachers. International Journal of Information and Learning Technology, 42, 449. https://doi.org/10.1108/IJILT-03-2025-0085
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Medicine, 6(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100
Liu, W., Huang, X., Zou, Y., & Zhu, K. (2025). Reconfigure the classroom: An activity-theoretic analysis of an AI conversational agent in vocational technical communication. 8th International Conference on Educational Technology Management, ICETM 2025, 308–313. https://doi.org/10.1109/ICETM67477.2025.11413604
Mei, B., Qi, W., Huang, X., & Huang, S. (2024). Speeko: An artificial intelligence-assisted personal public speaking coach. RELC Journal, 55(2), 596–600. https://doi.org/10.1177/00336882221107955
Meniado, J. C. (2025). Utilizing ChatGPT for self-directed professional development: Practices of Southeast Asian language teachers. Language Teaching Research Quarterly, 53, 1–20. https://doi.org/10.32038/ltrq.2025.53.01
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
Niranon, P., & Triyason, T. (2024). The efficiency of ChatGPT vs Google against self-learning of undergraduate students. 8th International Conference on Information Technology 2024, InCIT 2024, 382–386. https://doi.org/10.1109/InCIT63192.2024.10810629
Nitisakunwut, P., Hwang, G. J., & Chanyoo, N. (2022). Pedagogical design perspective of digital game-based English language learning. International Journal of Online Pedagogy and Course Design, 12. https://doi.org/10.4018/IJOPCD.311437
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. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
Page, M. J., Moher, D., 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., … McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. BMJ, 372, n160. https://doi.org/10.1136/bmj.n160

Pan, R., Li, B., Sun, Y., & Zhou, Q. (2025). A discourse-historical analysis of research on digital empowerment in English language teaching in China. Social Sciences & Humanities Open, 12, 101920. https://doi.org/10.1016/j.ssaho.2025.101920

Park, J., Yang, S., Yoon, C., Si, J., Jung, Y., & Kim, S. (2024). AI English conversation coaching platform in the metaverse: Focused on Korean users. 18th IEEE International Conference on Application of Information and Communication Technologies, AICT 2024. https://doi.org/10.1109/AICT61888.2024.10740456

Patturose, J. G. B., Priscilla, R., Karthick, M. V., & Jayasurya, P. S. (2024). Ingeleze: Advanced AI solution for English mastery. 3rd International Conference on Advances in Computing, Communication and Materials, ICACCM 2024. https://doi.org/10.1109/ICACCM61117.2024.11058704

Pham, A. T., & Abdullah, M. R. T. L. (2025). Validating and prioritizing key indicators for blended MOOC implementation in English language learning using the fuzzy Delphi method. MethodsX, 15, 103668. https://doi.org/10.1016/j.mex.2025.103668

Polasa, P., Laoaree, S., Thanadunpremdet, T., Rodjananant, N., & Kritsuthikul, N. (2025). Chai-calibrated hybrid assessment for IELTS speaking with human-referenced validation. 2025 20th International Joint Symposium on Artificial Intelligence and Natural Language Processing, iSAI-NLP 2025. https://doi.org/10.1109/iSAI-NLP66160.2025.11320542

Qassrawi, R., & Al Karasneh, S. M. (2025). Redefinition of human-centric skills in language education in the AI-driven era. Studies in English Language and Education, 12(1), 1–19. https://doi.org/10.24815/siele.v12i1.43082

Qinyuan, H. (2025). Design of a university English-blended teaching quality assessment model based on WFKP clustering algorithm. International Journal of High Speed Electronics and Systems, 34(4), Article 2540287. https://doi.org/10.1142/S0129156425402876

Rahmanu, I. W. E. D., & Molnár, G. (2024). Multimodal immersion in English language learning in higher education: A systematic review. Heliyon, 10, e38357. https://doi.org/10.1016/j.heliyon.2024.e38357

Rashid, S., Malik, S., & Ghauri, F. (2026). The active learning–GenAI synergy framework: Ethical integration of generative AI in EFL/ESL writing in resource-constrained contexts. F1000Research, 14, 1089. https://doi.org/10.12688/f1000research.171435.2

Remache, A., & Remache, S. (2025). Arab EFL students' engagement with prompt-engineered AI writing feedback. 2025 3rd International Conference on Foundation and Large Language Models, FLLM 2025, 1010–1017. https://doi.org/10.1109/FLLM67465.2025.11391002

Richardson, D. J., & Bowen, N. E. J. A. (2023). Grade variations across EFL, CLIL, and EMI platforms: A longitudinal case study in an English language teaching department. System, 119, 103163. https://doi.org/10.1016/j.system.2023.103163

Riser, P. (2025). Generative artificial intelligence (AI) in the high school English classroom and its effect on students' identity and teachers' practice. English in Education, 59(2), 176–192. https://doi.org/10.1080/04250494.2025.2491361

Roy, A. (2024). Impact of digital storytelling on middle school students' attitudes toward English language learning. International Journal of Web-Based Learning and Teaching Technologies, 19. https://doi.org/10.4018/IJWLTT.352496

Smith, J. (2025). Enhancing ESP student critical thinking skills and vocabulary acquisition through a GenAI-based project. Teaching English with Technology, 25(2), 82–107. https://doi.org/10.56297/vaca6841/ZDDD4208/HNVJ7828

Sun, Y., & Wu, Y. (2026). The efficacy of artificial intelligence-powered scaffolding in individual acquisition efficiency of EFL in tertiary educational context. Frontiers in Psychology, 16, 1613285. https://doi.org/10.3389/fpsyg.2025.1613285

Susoy, Z. (2026). Reducing anxiety and enhancing performance: The impact of AI chatbots versus human facilitation on EFL speaking assessment outcomes. Frontiers in Psychology, 16, 1745942. https://doi.org/10.3389/fpsyg.2025.1745942

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8, Article 45. https://doi.org/10.1186/1471-2288-8-45

Tong, S. (2024). Assessing English learning needs in sports physical education and the application of machine learning algorithms. Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte, 24(97), 72–89. https://doi.org/10.15366/rimcafd2024.97.006

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 14(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Wang, H., & Liu, T. (2026). Psychological determinants of GenAI adoption for foreign language education: An extended UTAUT2 model and sentiment analysis approach. Frontiers in Psychology, 16, 1622926. https://doi.org/10.3389/fpsyg.2025.1622926

Wang, M., Liang, H. N., Liu, Y., Ji, C., & Yu, L. (2025). Tangible progress: Employing visual metaphors and physical interfaces in AI-based English language learning. Big Data Research, 42, 100570. https://doi.org/10.1016/j.bdr.2025.100570

Wang, Q., Amini, M., & Fu, Z. (2025). AI acceptance and Chinese EFL learners' behavioral engagement with mediating effects of motivation. Scientific Reports, 15, 25698. https://doi.org/10.1038/s41598-025-11305-2

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2024). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. Computer Assisted Language Learning, 37(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Wang, Y., Peng, Z., & Hou, J. (2025). A study on the attitudes of graduate students in foreign languages disciplines towards GAI-assisted academic paper writing. Acta Psychologica, 260, 105665. https://doi.org/10.1016/j.actpsy.2025.105665

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Boonyakitanont, P., & Jinowat, N. (2024). Designing chatbots in language classrooms: An empirical investigation from user learning experience. Smart Learning Environments, 11(1), Article 32. https://doi.org/10.1186/s40561-024-00319-4

Xu, T., & Wang, H. (2024). The effectiveness of artificial intelligence on English language learning achievement. System, 125, 103428. https://doi.org/10.1016/j.system.2024.103428

Xu, X., Liu, J., Zheng, R., Lei, V. N., & An, Q. (2025). Learners' perception of data privacy when using AI language models: Reflective diary analysis of undergraduates in China. Acta Psychologica, 259, 105491. https://doi.org/10.1016/j.actpsy.2025.105491

Yan, J., Wu, C., Tan, X., & Dai, M. (2025). The influence of AI-driven personalized foreign language learning on college students' mental health: A dynamic interaction among pleasure, anxiety, and self-efficacy. Frontiers in Public Health, 13, 1642608. https://doi.org/10.3389/fpubh.2025.1642608

Yu, X., Gazi, M. A. I., Faisal-E-Alam, M., Emon, M., Rahaman, M. A., Al Masud, A., & Senathirajah, A. R. S. (2025). Chinese students' English language learning intention and behavior: Mediating effect of attitude and moderating effect of institutional support. Acta Psychologica, 260, 105596. https://doi.org/10.1016/j.actpsy.2025.105596

Zhang, Q., Jin, H., & Teng, W. W. (2025). The effects of learning experience on college students' deep English learning: A study of the chain mediation effect of motivation and strategy. PLOS ONE, 20(6), e0325491. https://doi.org/10.1371/journal.pone.0325491

Zhen, X. (2025). Application of artificial intelligence in English learning: The interactive effects of teacher support, learning motivation, and autonomous learning ability. Acta Psychologica, 261, 105960. https://doi.org/10.1016/j.actpsy.2025.105960