International Journal of Language and Literary Studies
Volume 8, Issue 6, 2026
Homepage : http://ijlls.org/index.php/ijlls

Generative AI in Secondary Language Education: Pedagogy, Skills, Outcomes, and Teacher Mediation

Joenel Bulan Lacaden
Special Science Teacher III, Philippine Science High School – Cagayan Valley Campus in Nueva Vizcaya, Philippines
Corresponding Author Email: jblacaden@cvc.pshs.edu.ph

DOI: https://doi.org/10.36892/ijlls.v8i6.2855

APA Citation: Lacaden, J. (2026). Generative AI in Secondary Language Education: Pedagogy, Skills, Outcomes, and Teacher Mediation. *International Journal of Language and Literary Studies, 8*(6).409-425. https://doi.org/10.36892/ijlls.v8i6.2855

Received:
27/092026

Accepted:
05/11/2026

Keywords:
Generative Artificial Intelligence, Secondary Language Education, ChatGPT, Language Learning, Teacher Mediation

Abstract

Generative artificial intelligence (GenAI) is increasingly used in language education, yet evidence from secondary-school settings remains fragmented. This integrative review examines empirical research on GenAI in formal secondary language education published from January 2023 to September 2026, with attention to pedagogical applications, language skills and activities, learning outcomes, teacher mediation, implementation challenges, and gaps in the evidence. Guided by Whittemore and Knafl's integrative-review methodology and a PICo framework, the review drew on Google Scholar, ERIC, publisher and journal websites, and citation chasing. Twenty-two empirical studies were examined against the review criteria and available original publication records. Narrative and thematic synthesis were used to compare study contexts, methods, reported outcomes, and limitations. The studies describe uses of GenAI for writing support, feedback, conversation, tutoring, learning companionship, teacher support, and AI-literacy activities. Writing is the most represented language focus. Reported outcomes are mixed, ranging from gains in selected skills and learner perceptions to null or qualified findings. Overall, the evidence indicates that the educational contribution of GenAI depends on task design, teacher mediation, verification, AI literacy, and the conditions in which the technology is used.

1. Introduction

Generative artificial intelligence has rapidly changed the technological environment of language teaching and learning. Large language models and generative chatbots can produce explanations, examples, dialogue, feedback, summaries, revisions, and other language-related content. At the same time, their use raises questions about accuracy, overreliance, academic integrity, authorship, privacy, bias, unequal access, and changing teacher roles. These issues are especially important in secondary education, where learners work within prescribed curricula, teacher-supervised classrooms, age-related developmental needs, and assessment systems that differ from those of higher education.

The review is relevant to Sustainable Development Goal 4, particularly targets concerning effective learning, relevant skills, and educational equity. These links are implications rather than outcomes measured by the included studies. GenAI may create opportunities for language practice and digital competence, while also raising concerns about unequal access, privacy, bias, assessment, and teacher capacity. The review therefore treats equity and quality as implementation questions rather than assuming that GenAI automatically advances them.

2. Literature Review

GenAI in Language Education and Secondary-School Contexts

Broader reviews provide context for this secondary-school focus. Li et al. (2025) synthesized 144 empirical articles on generative AI in language learning and teaching published from 2023 to 2024 and found that higher education and English as a foreign language dominated the evidence, while writing received more attention than other language skills. Lee et al. (2026) likewise reviewed classroom-based GenAI research in second- and foreign-language learning and reported that the evidence was concentrated in higher education and often relied on self-reported learner data. A task-based review by Li et al. (2026) also found that K–12 settings represented only a small share of the studies it analyzed. More directly, Pérez-Montesdeoca et al. (2026) reviewed GenAI integration in secondary education and identified a fragmented and methodologically uneven evidence base. Lee (2026), in a broader review of AI-based instruction in K–12 education, emphasized that teacher judgment and classroom orchestration mediate the instructional use of AI-generated information. Taken together, these reviews support the need for a focused synthesis of empirical research in formal secondary language education that distinguishes student outcomes from teacher-focused evidence and examines GenAI roles in relation to instructional tasks and teacher mediation.

Philippine Context and Research Gap

The Philippine context makes this issue particularly relevant. Recent secondary-school studies have examined students' use of ChatGPT for writing and English-language skills and have begun to document teachers' knowledge, attitudes, and practices. Javier and Moorhouse (2024) examined productive and critical ChatGPT use among secondary English learners in the Philippines; Maghamil and Sieras (2024) investigated academic writing among Grade 12 students; Lariosa et al. (2025) examined ChatGPT use and English-language skills among 117 Grade 12 students; and Cabato (2025) studied secondary ESL teachers in Zamboanga City. These studies indicate that GenAI is already entering the Philippine secondary-language environment, while also raising questions about teacher preparation, responsible use, and the relationship between frequency of AI use and measurable language development.

Objective and Research Questions

The objective of this integrative review is to synthesize empirical research on the use of generative artificial intelligence in secondary language education from January 2023 to September 2026, with particular attention to pedagogical applications, language skills and instructional activities, learning outcomes, teacher-mediated practices, implementation challenges, and gaps in the evidence base.

1. What applications and pedagogical roles of generative artificial intelligence have been reported in secondary language education?
2. Which language skills and instructional activities have been investigated through generative-AI-supported secondary language education?
3. What cognitive, linguistic, affective, and behavioral learning outcomes have been reported?
4. What teacher-mediated pedagogical practices, implementation conditions, and challenges characterize the integration of generative artificial intelligence in secondary language education?
5. What methodological, geographical, linguistic, theoretical, and empirical gaps characterize the current evidence base?

3. Review Framework and Theoretical Positioning

Table 1. ***PICo framework and operationalization***

PICo element	Operationalization
Population	Secondary school students and/or secondary language teachers, including lower-secondary, upper-secondary, junior/high school, senior high school, and nationally equivalent settings.
Interest	Generative AI used for language teaching, learning, assessment, feedback, interaction, content generation, writing support, self-directed learning, or related instructional activities.
Context	Formal secondary language education in English, EFL/ESL, second/foreign/additional language, and other language-education settings.

Note. PICo = Population, Interest, Context. The framework defines the population and educational setting without restricting the review to a single language or GenAI platform.

3.1 Theoretical positioning

Because this study is an integrative review rather than a primary hypothesis-testing study, it does not impose a single grand theory on all included studies. Instead, it maps the theoretical perspectives explicitly reported in the evidence base. These include Activity Theory in research on teacher feedback practices (Yao et al., 2025), self-directed learning and autodidaxy in research on secondary learners' use of GenAI as a learning companion (Ma et al., 2026), and Dynamic Systems Theory in work examining writing complexity, accuracy, and fluency (Wang & Shi, 2026). The review uses these perspectives to understand different dimensions of GenAI integration rather than treating them as a single integrated causal framework. The conceptual synthesis therefore operates at the level of pedagogical positioning and teacher mediation while retaining theoretical diversity across studies. No hypotheses are proposed because the review synthesizes existing evidence rather than testing a primary causal model.

4. Methodology

4.1 Review design

This study employed an integrative review design guided by the five-stage methodology described by Whittemore and Knafl (2005): problem identification, literature search, data evaluation, data analysis, and presentation. This approach was appropriate because the evidence includes qualitative, quantitative, mixed-methods, quasi-experimental, survey, case-based, and practitioner studies. The review used predefined eligibility criteria, a multi-source literature search, structured data extraction, and narrative and thematic synthesis to compare findings across diverse empirical designs.

PRISMA 2020 was not adopted as the review methodology or as a protocol governing this study. PRISMA is a reporting guideline for systematic reviews, whereas the present study follows the integrative-review methodology of Whittemore and Knafl (2005). The review nevertheless describes its information sources, eligibility criteria, and documented selection stages to make the process as transparent as the retained records allow. Accordingly, the study does not claim full PRISMA 2020 compliance, and Figure 1 is not presented as a numerical PRISMA flow diagram.

4.2 Review framework and eligibility criteria

The review used a PICo framework. The Population comprised secondary-school students and/or secondary language teachers, including lower-secondary, upper-secondary, junior/high school, senior high school, and nationally equivalent settings. The Interest was generative AI used for language teaching, learning, assessment, feedback, interaction, content generation, writing support, self-directed learning, or related instructional activities. The Context was formal secondary language education in English, EFL/ESL, second/foreign/additional language, and other language-education settings. Student-focused and teacher-focused studies

were retained within the same educational context but were analyzed as distinct evidence populations: teacher-focused studies were used primarily to address pedagogical roles, teacher mediation, and implementation, and were not treated as evidence of student learning outcomes.

4.3 Information sources and search strategy

The literature was identified through Google Scholar and ERIC, supplemented by publisher and journal websites and backward and forward citation chasing. Publisher and journal records were used to check bibliographic details and consult available full texts. ChatGPT (OpenAI) was used as an auxiliary literature-discovery and search-concept development tool; suggestions were considered only after the underlying publications had been checked. The review period was January 2023 through September 2026. Search concepts were organized into three blocks: (A) GenAI terminology, (B) secondary-education terminology, and (C) language-education terminology.

4.3.1 Search strategy

The GenAI block included "generative artificial intelligence," "generative AI," GenAI, ChatGPT, Gemini, Bard, Copilot, Claude, "large language model*," and "generative chatbot*." The secondary-education block included "secondary education," "secondary school*," "high school*," "senior high school," "junior high school," "middle school," "lower secondary," and "upper secondary." The language-education block included "language education," "language learning," "language teaching," "language instruction," "language classroom," "language learner*," "foreign language," "second language," EFL, ESL, and "additional language." Skill-specific terms were not required in the core search because requiring writing, reading, speaking, or listening terms could exclude studies addressing integrated language activities. These terms describe the search concepts retained for the review; they should not be read as verbatim historical queries for every platform.

4.3.2 Search dates and search log

The evidence was assembled and checked through September 2026. Complete platform-level query logs, record exports, and retrieval totals were not preserved for every source, so the review reports the sources and search concepts that can be substantiated rather than reconstructing database counts. Citation chasing and publisher-level verification supplemented the initial search.

4.4 Search-access limitation and evidence-set construction

Scopus and Web of Science were not directly searched during final evidence assembly because institutional access was unavailable. The review therefore relied on Google Scholar, ERIC, publisher and journal websites, and citation chasing. Twenty-two empirical studies met the substantive eligibility criteria and were checked against available original articles and publication records. They constitute the evidence set analyzed in this review, not a definitive census of every eligible study published during the period.

4.5 Evidence identification and eligibility assessment

Candidate studies were assessed for relevance and then checked against the full-text eligibility criteria. The author evaluated records against the predefined population, context, GenAI, empirical-design, publication-type, and relevance criteria and made the final inclusion decisions. The 22 included studies were checked against available original articles and publication records. Screening and inclusion were conducted by one reviewer, without duplicate independent screening or inter-rater agreement. Reasons for exclusion are reported only where supported by retained records.

To make the selection pathway visible without implying that unavailable historical counts can be recovered, Figure 1 summarizes the documented sources and decision stages. The retained records do not contain complete platform-level retrieval totals, duplicate counts, or stage-specific exclusion counts. Accordingly, no numerical values are assigned to the identification, screening, or eligibility stages. The final inclusion figure of 22 refers to the

empirical studies verified and retained for synthesis. This diagram is a transparent account of the documented process, not a reconstruction of missing database statistics.

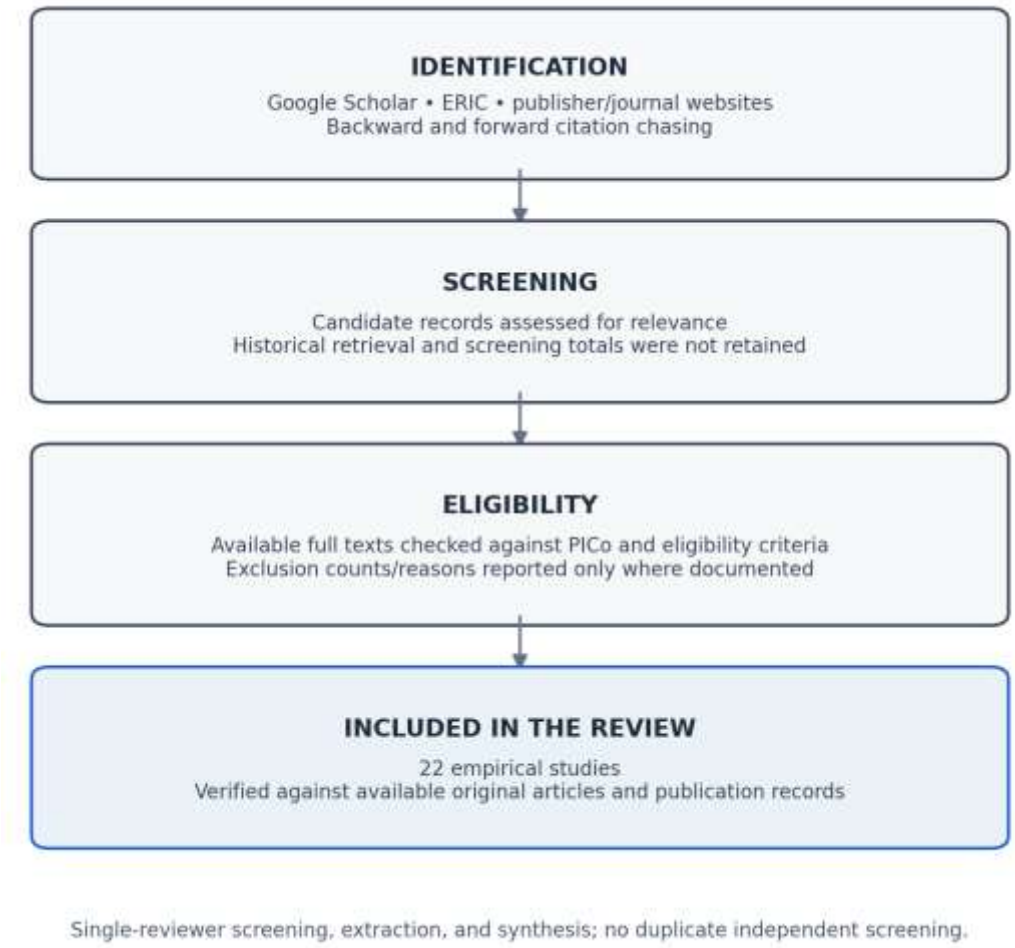


Figure 1. ***Documented study-identification, screening, eligibility, and inclusion pathway.***

Note. Historical retrieval totals, duplicate counts, and stage-specific exclusion counts were not retained and are therefore not estimated or presented. This figure is a documented study-selection pathway, not a numerical PRISMA 2020 flow diagram. PRISMA 2020 was not adopted as the review methodology, and the manuscript does not claim full PRISMA compliance. The review included 22 empirical studies.

4.6 Eligibility criteria

Studies were eligible when they (a) were published from January 2023 through September 2026; (b) were journal articles meeting the review's peer-reviewed publication criterion; (c) reported empirical quantitative, qualitative, mixed-methods, experimental, quasi-experimental, case, survey, interview, observational, or practitioner research; (d) involved secondary students and/or secondary language teachers; (e) were conducted in a formal secondary language-education context; (f) substantially investigated a generative AI system or generative-AI-supported activity; (g) provided evidence relevant to at least one research question; and (h) provided accessible full text sufficient for evidence assessment. Publication-type eligibility was considered from the bibliographic and publisher information available during evidence verification; database indexing status was not used as an eligibility criterion.

Studies were excluded when they involved primary or higher education only, adult education only, nonseparable mixed educational levels, non-language subjects without a separable language component, conventional AI without a generative component, incidental GenAI mentions, non-empirical publications, or insufficient evidence for the review questions. Reviews, editorials, commentaries, conference abstracts, and other non-journal publications were not treated as empirical studies in the included evidence set. Grammarly-only studies were

excluded because Grammarly does not meet the operational definition of GenAI used in this review.

4.7 Data extraction

A structured extraction form captured bibliographic information; country and educational setting; grade and participant characteristics; language; GenAI system and reported model or version; pedagogical role; language skill and instructional activity; linguistic, cognitive, affective, and behavioral outcomes; teacher-mediated practices; implementation challenges; research design; instruments; study limitations; and links to the five research questions. Extracted information was checked against the available original articles and publication records. Details that could not be established from the article text were recorded as not reported or Cannot Tell rather than inferred. Because screening, extraction, and synthesis were conducted by one reviewer, some selection and interpretation bias remains possible.

Pedagogical roles were coded as feedback provider, tutor/instructional assistant, conversation partner, content generator, writing assistant, assessment support, personalized-learning support, interaction facilitator, brainstorming support, learning companion, or AI-literacy/metacognitive support. Outcomes were classified as linguistic, cognitive, affective, or behavioral, with findings distinguished as positive, negative, null, mixed/conditional, descriptive, or not reported. Objective or observed outcomes were distinguished from self-reported outcomes where the primary study permitted this distinction. For synthesis, extracted findings were first mapped to the five research questions, then grouped into higher-order categories covering applications, language skills, outcomes, teacher-mediated practices and implementation conditions, and research gaps. Contradictory, null, and conditional findings were retained rather than collapsed into a single directional conclusion.

4.8 Evidence evaluation

In keeping with the integrative-review approach, studies were considered in relation to their research design, educational setting, participants, data collection and analysis, relevance to the review questions, and reported limitations. These features informed comparisons across studies and the strength of the conclusions drawn. No standardized numerical quality score was calculated.

This review synthesized previously published research and did not recruit or interact directly with human participants; institutional ethics approval was therefore not required for the review itself. Ethical issues relevant to GenAI-supported language education—including privacy, bias, authorship, academic integrity, and responsible use—were considered where they informed interpretation of the evidence.

4.9 Data synthesis

Narrative, thematic, and cross-dimensional synthesis followed the analysis and interpretation stages described by Whittemore and Knafl (2005). Extracted findings were organized by the five research questions and compared across populations, language skills, GenAI roles, designs, outcomes, teacher-mediated practices, and implementation conditions. The synthesis distinguished observed or measured outcomes from self-reported perceptions and treated teacher-focused evidence separately from student learning outcomes. Recurrent patterns were developed into higher-order themes, while contradictory, null, mixed, and conditional findings were retained. Counts of studies addressing a particular system or language skill describe this included evidence set; they are not estimates of prevalence or effectiveness. No pooled effect size was calculated because the studies differed substantially in interventions, comparison conditions, outcome measures, duration, and design.

5. Results and Discussion

5.1 Characteristics of the 22-study included evidence set

The final included evidence set comprises 22 empirical studies that met the review's substantive eligibility criteria. The studies span eight geographical contexts: China (5), the Philippines (4), Hong Kong (4), Türkiye (3), Morocco (2), Vietnam (2), Korea (1), and Malta (1). These counts describe the included evidence set rather than population estimates. The corpus includes secondary students and secondary language teachers and covers qualitative, quantitative, quasi-experimental, mixed-methods, case-based, and repeated-measures designs.

The corpus is strongly concentrated on English/EFL and writing. Writing is the primary coded language focus in 11 of 22 studies (50.0%). Other primary foci include integrated/productive language, reading, grammar/vocabulary, vocabulary, speaking, and writing/communicative competence. Four studies (18.2%) primarily involve secondary language teachers or teacher-mediated practice and should not be interpreted as student-outcome studies. ChatGPT is explicitly named in the GenAI-system field of 16 of 22 studies (72.7%); this is a system-label count, not evidence that ChatGPT is more effective than other GenAI systems.

Table 2. ***Characteristics of the 22-study included evidence set***

Study ID	Study	Context / participants	Primary focus	Design pattern
S01	Javier & Moorhouse (2024)	Philippines; 49 secondary English learners	Productive and critical ChatGPT use	Qualitative/practitioner
S02	Woo et al. (2024)	Hong Kong; 21 secondary EFL students	Writing; motivation; cognitive load; satisfaction	Mixed methods
S03	Zheng (2024)	China; 84 secondary EFL students	Reading anxiety and reading performance	Mixed-method quasi-experiment
S04	Shim & Kim (2024)	Korea; 30 high-school students	ChatGPT feedback for English writing	Experimental/quasi-experimental
S05	Maghamil & Sieras (2024)	Philippines; 54 Grade 12 students	Academic writing quality	Randomized pretest-posttest
S06	Pei et al. (2025)	Hong Kong; 99 Grade 10 students	Formal ChatGPT integration in EFL writing	Quasi-experiment
S07	Yao et al. (2025)	China; 13 secondary English teachers	AI-guided feedback on writing	Qualitative
S08	Żammit (2025)	Malta; 47 secondary teachers / 22 institutions	GenAI in Maltese language teaching	Qualitative
S09	Cabato (2025)	Philippines; 181 secondary ESL teachers	Knowledge, attitudes, practices	Quantitative
S10	Lariosa et al. (2025)	Philippines; 117 Grade 12 students	ChatGPT and English skills	Correlational
S11	Woo, Guo & Susanto (2025)	Hong Kong; 4 secondary case studies	ChatGPT writing task pathways	Multiple case study
S12	Farhane (2025)	Morocco; 87 secondary students	ChatGPT formative feedback and paragraph writing	Quasi-experimental / mixed
S13	Gok Acan & Liman Kaban (2026)	Türkiye; 53 secondary EFL students	Human vs. ChatGPT-4 writing feedback	Mixed methods
S14	Kök (2026)	Türkiye; Grade 12 EFL students	GenAI literacy and autonomous learning	Quasi-experimental / qualitative

Study ID	Study	Context / participants	Primary focus	Design pattern
S15	Ma et al. (2026)	Hong Kong; 278 students from 22 schools	GenAI as learning companion	Mixed methods
S16	Wang, Cui & Fang (2026)	China; 2 secondary EFL teachers	Teacher identity and GenAI	Case study
S17	Wang & Shi (2026)	China; Chinese high-school EFL learners	GenAI and writing complexity, accuracy, fluency	Intervention / repeated measures
S18	Mai (2026)	Vietnam; 42 Grade 11 students	ChatGPT-supported argumentative writing	Quasi-experimental mixed methods
S19	Li (2026)	China; 154 Grade 11 students	DeepSeek-supported writing and communicative competence	Quasi-experiment / repeated measures
S20	Ozan & Azap (2026)	Türkiye; 30 Grade 11 students	ChatGPT feedback on topic-specific writing	Mixed-methods quasi-experiment
S21	Pham & Pham (2026)	Vietnam; 208 Grade 9 students + 8 interviewees	ChatGPT vocabulary learning; motivation and attitudes	Mixed methods
S22	Farhane (2026)	Morocco; 22 second-year Baccalaureate students	ChatGPT-assisted speaking feedback	Qualitative focus groups

Note. These counts describe the 22 studies checked against available scholarly sources, publisher records, and citation trails. They are not database-wide screening counts.

5.2 RQ1: Applications and pedagogical roles

Several recurring pedagogical roles emerge across the evidence. GenAI functions as a feedback provider, particularly in writing; a conversational or tutoring partner; a writing and content-generation assistant; a learning companion for self-directed activity; a teacher-support resource; and an AI-literacy or metacognitive resource. The expanded corpus also illustrates role-specific variation: Mai (2026) examined supervised ChatGPT-supported argumentative writing; Li (2026) examined DeepSeek-supported continuation writing and communicative competence; Ozan and Azap (2026) examined ChatGPT feedback on topic-specific writing; Pham and Pham (2026) examined ChatGPT-supported vocabulary learning; and Farhane (2026) examined ChatGPT-assisted speaking feedback. These studies reinforce that GenAI is not a single instructional intervention but a set of affordances positioned differently within learning tasks.

The evidence suggests that these roles are not mutually exclusive. A single instructional sequence may position GenAI as a generator, feedback provider, conversation partner, and object of critical evaluation. The relevant analytical distinction is therefore between technological presence and pedagogical positioning. The corpus does not support treating frequency of a particular role as a prevalence estimate because the studies were not selected through a database-wide prevalence study.

To distinguish the presence of a GenAI tool from its pedagogical contribution, Table 3 compares the roles identified in the included studies with the instructional tasks they support, the teacher's mediating responsibility, and the type of outcome evidence reported. The table is an interpretive synthesis of the reviewed studies; it does not imply that every application produced a measurable improvement.

Table 3. ***Comparative analytical framework of GenAI applications, instructional tasks, teacher mediation, and outcome evidence***

GenAI role	Instructional task	Teacher involvement	Observed or reported outcome evidence
Content generator	Producing model texts, sample paragraphs, examples, or alternative formulations for analysis.	Sets the language objective; asks learners to critique, verify, transform, or use models as a basis for independent production.	Access to models and examples is documented; generation alone does not establish independent language development.
Feedback provider	Providing formative comments on grammar, vocabulary, organization, coherence, or task-specific writing.	Checks accuracy and relevance; compares AI feedback with criteria or human feedback; guides students in deciding which suggestions to adopt.	Writing quality and revision outcomes are mixed or conditional across studies; feedback availability is not itself proof of improvement.
Writing assistant	Supporting brainstorming, outlining, drafting, continuation writing, and revision.	Structures the writing process, preserves opportunities for independent drafting, and examines students' revision decisions and authorship.	Studies report selected writing and communicative-competence outcomes, but findings vary by task, measure, and intervention.
Conversation partner / interaction facilitator	Providing simulated dialogue, speaking practice, or opportunities to rehearse language.	Defines communicative goals, monitors appropriateness and accuracy, and supplements rather than replaces authentic interaction.	Some evidence concerns confidence and anxiety perceptions; direct evidence of speaking proficiency remains limited in this corpus.
Tutor / instructional assistant	Explaining language concepts, clarifying reading passages, and supporting grammar or vocabulary practice.	Scaffolds learning, checks explanations against reliable sources, and assesses whether learners can apply concepts without AI assistance.	Evidence spans reading performance and language-learning processes; the small and varied evidence base limits broad conclusions.
Learning companion	Supporting self-directed practice, personalized assistance, and learner questions beyond a single classroom task.	Sets boundaries, encourages learner agency, and monitors dependence, verification, and the quality of self-regulation.	Studies report motivation, autonomy, and self-directed learning-related findings; long-term effects require further study.
AI-literacy / metacognitive support	Evaluating generated responses, improving prompts, verifying claims, and reflecting on appropriate use.	Explicitly teaches critical evaluation, privacy, authorship, bias awareness, and responsible use.	Evidence includes AI-literacy and critical-use outcomes; validated measures and longitudinal evidence remain limited.

Note. The table distinguishes pedagogical functions from demonstrated effects. Outcomes are described only at the level supported by the included studies; learner perceptions, process observations, and measured achievement are not treated as interchangeable.

5.3 RQ2: Language skills and instructional activities

Writing remains the most represented primary language focus, accounting for 11 of the 22 studies. Writing-related activities include drafting, paragraph production, argumentative writing, continuation tasks, revision, feedback, and comparison of machine and human feedback. The newer studies broaden the writing evidence by examining argumentative writing

(Mai, 2026), communicative competence in continuation writing (Li, 2026), and topic-specific writing feedback (Ozan & Azap, 2026).

The same GenAI function can support different learning activities: generated text may serve as a model for analysis, a draft for critical revision, or an answer that students are asked to verify. These activities place different demands on learners and should not be grouped as equivalent instances of AI-supported writing.

Reading is represented through chatbot-supported reading activities, including work examining reading anxiety and reading performance. Speaking is represented by Farhane (2026), while vocabulary is directly examined by Pham and Pham (2026). Grammar/vocabulary and broader integrated-language activities are also represented. The distribution indicates that the secondary evidence base is broader than writing alone but remains uneven across language skills, with direct speaking, reading, and vocabulary evidence each represented by one primary-focus study in the coded corpus.

A notable feature is the emergence of instructional activities concerned with critical AI use: evaluating outputs, verifying information, prompting strategically, recognizing limitations, and reflecting on appropriate use. These activities shift GenAI from being merely a learning aid to becoming part of the language-learning environment that students must learn to navigate critically.

5.4 RQ3: Learning outcomes

The outcome evidence spans four domains. Linguistic outcomes include writing quality, grammar, vocabulary-related learning, reading performance, communicative competence, and dimensions of writing complexity, accuracy, and fluency. Cognitive outcomes include cognitive engagement, self-regulation, and AI literacy. Affective outcomes include motivation, satisfaction, confidence, anxiety, and attitudes. Behavioral outcomes include participation, revision behavior, interaction, autonomy, and patterns of AI use. Importantly, not all studies measured achievement outcomes; some primarily examined perceptions, motivation, attitudes, teacher practice, or learning processes.

For interpretation, measured language performance, observed learning processes, and self-reported perceptions are kept analytically distinct: confidence, satisfaction, or motivation may indicate how learners experience an activity, but they do not by themselves establish gains in linguistic competence.

Several intervention-oriented studies report improvements in selected language outcomes, but the findings are not uniform. Mai (2026) reported early advantages for the ChatGPT group in aspects of argumentative writing alongside concerns about formulaic writing and reduced critical engagement. Li (2026) reported greater gains for the DeepSeek group across several communicative-competence dimensions, while noting weaker improvement in idea organization. Ozan and Azap (2026) found early task-level gains that converged by the post-test, with no significant overall post-test difference. Lariosa et al. (2025) reported no significant relationship between ChatGPT use and measured English-language skills in a correlational study. These findings illustrate why GenAI use should not be equated with learning effectiveness.

The evidence also suggests that affective and cognitive outcomes may move in different directions. GenAI can be associated with convenience, confidence, satisfaction, motivation, or autonomy while also raising concerns about cognitive load, dependence, reliability, critical engagement, authorship, and authenticity. Pham and Pham (2026), for example, reported moderate motivation and generally positive attitudes toward ChatGPT-supported vocabulary learning, while also identifying reliability and over-reliance concerns. Farhane (2026) reported perceived gains in speaking confidence and reduced anxiety but also concerns about accuracy, dependence, and authenticity. The appropriate synthesis is therefore conditional rather than uniformly positive.

5.5 RQ4: Teacher-mediated practices, implementation conditions, and challenges

Teacher mediation is a recurring feature of the evidence. Teachers participate in task design, prompt design, scaffolding, monitoring, verification, feedback, assessment redesign, AI-literacy development, critical evaluation, and discussion of ethics and academic integrity. The teacher is therefore positioned in the evidence as an important orchestrator of GenAI-supported learning rather than as a role made unnecessary by the technology. Teacher-focused studies further show that preparedness, professional identity, and practical classroom conditions shape how GenAI is integrated.

The implementation challenges reported across the evidence can be grouped into pedagogical, technological, ethical, institutional, and contextual dimensions. Accuracy and hallucination concerns require verification. Overreliance and student misuse require explicit guidance and monitoring. Academic integrity and authorship concerns affect assessment design. Privacy, bias, access, and infrastructure raise institutional and equity questions. Teacher competence and AI literacy influence whether GenAI is integrated critically rather than used as an unexamined shortcut. Speaking-focused evidence additionally raises questions about authenticity of interaction, while vocabulary evidence highlights concerns about passive use and reliability.

These challenges are not peripheral to implementation. They influence the design of learning activities themselves. A writing task, for example, may need to require independent drafting, documented prompting, source verification, revision histories, and reflective commentary rather than merely asking students to submit a final AI-assisted product.

5.6 Methodological characteristics and interpretation

The included studies use heterogeneous designs, including qualitative, survey, correlational, case-based, quasi-experimental, and repeated-measures approaches. These designs answer different kinds of questions: intervention studies may inform changes in selected outcomes, while qualitative and survey studies are particularly useful for understanding learners' and teachers' experiences, perceptions, and implementation conditions. The evidence should therefore be interpreted in light of each study's design, context, sample, measures, and reported limitations rather than as a uniform body of causal evidence.

Experimental and quasi-experimental studies provide more direct evidence about selected learning outcomes than studies based only on perceptions, but the included interventions, samples, durations, comparison conditions, and measures vary. Findings from perception-based studies are relevant to acceptability and experience but should not be treated as proof of improved language performance.

Because of this methodological heterogeneity, the synthesis does not calculate an aggregate quality score or pool effects statistically. Instead, the review qualifies conclusions according to the type and limits of evidence available for each theme.

5.7 RQ5: Research gaps

Five higher-order gaps emerge from the 22-study corpus. First, secondary-specific evidence remains smaller and less mature than the broader GenAI language-learning literature. Second, teacher-mediated integration is less extensively studied than student perceptions and short-term outcomes. Third, longitudinal evidence and stronger comparative designs remain limited. Fourth, English/EFL and selected geographical contexts dominate, while multilingual and non-English language contexts remain less visible. Fifth, implementation conditions are often examined separately from measurable learning outcomes, limiting understanding of how instructional design and institutional conditions shape outcomes. The expanded corpus also shows the need for more direct evidence on speaking, vocabulary achievement, and broader communicative competence.

5.8 Cross-dimensional synthesis

Table 4. ***Cross-dimensional synthesis of pedagogical patterns and evidence gaps***

Theme	Supporting evidence	Qualification / contradiction	Methodological pattern	Gap / implication
Writing support	Multiple secondary EFL studies use GenAI for drafting, feedback, revision, and writing development.	Benefits are not uniform; correlational evidence includes null findings.	Quasi-experimental, mixed, case, correlational.	Need longer and better-controlled writing studies.
Feedback	GenAI can provide rapid feedback and support revision.	Machine feedback requires verification and may not align with instructional goals.	Intervention and qualitative studies.	Need systematic comparison of human, AI, and combined feedback.
Conversation / tutoring	GenAI provides interaction, explanation, and practice opportunities.	Quality, dependence, and appropriateness of interaction vary.	Qualitative and mixed-method evidence.	Speaking and interaction outcomes need more direct measurement.
Learning companion	GenAI can support self-directed learning and autonomy.	Autonomy may coexist with overreliance and weak verification.	Mixed-method evidence.	Need longitudinal autonomy and AI-use studies.
AI literacy	Explicit instruction can develop critical and responsible GenAI use.	AI literacy is still unevenly integrated into language curricula.	Quasi-experimental and qualitative evidence.	Need validated measures and curriculum-level implementation studies.
Teacher mediation	Teachers design tasks, monitor use, verify outputs, and facilitate reflection.	Teacher preparedness and identity can constrain integration.	Qualitative, case, mixed-method evidence.	Teacher development should be studied alongside student outcomes.

Note. Themes summarize recurring patterns across the included evidence set; they are not prevalence estimates and should not be interpreted as pooled effects.

5.9 Discussion

Across the evidence, the most consistent point is that pedagogical purpose matters more than adoption by itself. GenAI has been used as a source of feedback, a conversation partner, a writing assistant, a tutor, and a learning companion, but these roles are not interchangeable. What students learn from the technology depends on the task, the expectations placed on them, and the guidance they receive. The findings therefore favor a pedagogical starting point: teachers identify the language-learning purpose first and then decide whether GenAI can contribute meaningfully to that purpose.

The distinction between application and effectiveness is particularly important when interpreting these roles. A system that generates a model paragraph gives learners access to sample language, but the generated product does not demonstrate that learners can independently organize ideas or compose an effective text. By contrast, a feedback-oriented task can require learners to diagnose errors, judge the usefulness of suggestions, and revise their own writing. A conversation partner may create additional opportunities for rehearsal, but positive reports of confidence or reduced anxiety should not be interpreted as direct evidence of improved speaking proficiency. Likewise, a learning companion may support self-directed activity while also creating risks of dependence or uncritical acceptance. The instructional task therefore determines what learners are asked to do with the output, while teacher mediation helps ensure that the activity remains aligned with the language objective. The reviewed

evidence supports this differentiated interpretation, but it does not establish a universal causal effect for any single GenAI role.

Writing is the dominant focus in the evidence, which is understandable given the capacity of current GenAI systems to generate, revise, and comment on text. It is also a limitation of the literature. Research on speaking, interaction, reading, and vocabulary is less extensive, making it difficult to conclude that the patterns found in writing apply equally across language learning. The concentration on written products may partly reflect what is easiest to assign and assess, rather than the full range of ways in which language develops in secondary classrooms.

The outcome evidence does not point to a single effect. Differences in intervention design, learner characteristics, duration, comparison conditions, and outcome measures help explain why some studies report gains while others report mixed or null findings. The newer studies illustrate this variation: one reported broader gains across several dimensions of communicative competence, whereas another found early writing gains that did not persist as a significant overall post-test difference. GenAI use, therefore, should not be treated as the outcome-producing intervention in itself. Its contribution has to be considered alongside the instructional design and the particular learning outcome being measured.

Teacher mediation is central rather than secondary. Teachers determine the purpose of GenAI use, establish boundaries, design tasks, teach prompting and verification, monitor student behavior, and interpret the appropriateness of generated content. This pattern is consistent with the teacher-focused evidence in the corpus and with broader K–12 teacher-intervention literature. However, teacher mediation should be treated as a recurring explanatory theme rather than a proven universal causal moderator.

AI literacy also emerges as a language-education issue rather than merely a technical skill. Students increasingly need to evaluate generated language, identify inaccuracies, understand authorship and integrity, make informed prompting decisions, and recognize when independent language production is educationally necessary. In this sense, critical interaction with GenAI can itself become a language-learning activity.

The geographical and linguistic distribution of the included evidence requires attention. The corpus includes studies from China, the Philippines, Hong Kong, Türkiye, Morocco, Vietnam, Korea, and Malta. The distribution of the 22 included studies describes this identified evidence set only; it is not an estimate of the worldwide distribution of research. The concentration in selected contexts limits transferability to other school systems, languages, and policy environments. The evidence outside English/EFL remains comparatively limited in this corpus, so broader claims about language education should be cautious.

Taken together, the evidence supports a conceptual interpretation in which GenAI is embedded within a mediated instructional system: GenAI affordances influence available pedagogical roles; roles shape language-learning activities; activities structure learner–AI interaction; teacher mediation guides that interaction; and implementation conditions influence the processes through which linguistic, cognitive, affective, and behavioral outcomes may emerge. This is a conceptual synthesis, not a causal model.

5.10 Implications

5.10.1 For language teachers

- Define the instructional purpose of GenAI before introducing the tool.
- Require learners to evaluate, verify, revise, and reflect on AI-generated language rather than merely submit generated products.
- Use GenAI selectively for feedback, brainstorming, interaction, and practice while preserving opportunities for independent language production.
- Teach prompting, verification, source evaluation, privacy, authorship, and academic-integrity practices as part of AI literacy.

5.10.2 For schools and school leaders

- Develop clear guidance on acceptable and unacceptable GenAI use.

- Align assessment procedures with the reality of AI-supported work.
- Provide teacher professional development focused on pedagogy and critical AI literacy rather than tool operation alone.
- Address privacy, access, infrastructure, equity, and responsible-use requirements.

5.10.3 For researchers

- Report the GenAI system and version, task design, prompting procedures, teacher role, learner proficiency, intervention duration, and verification procedures.
- Use larger and more diverse samples and longer follow-up periods where feasible.
- Expand evidence beyond writing and English/EFL.
- Study teacher mediation and implementation conditions alongside student outcomes.

5.11 Limitations

The search was conducted across accessible scholarly sources rather than through exhaustive institutional database retrieval. Scopus and Web of Science were not directly searched during final evidence assembly. Complete historical query strings, record-level exports, and retrieval totals were not retained for every source, so the review cannot establish a complete census of eligible studies. The 22 included studies should be understood as the verified evidence set identified through the available search and citation-chasing process. Any additional search undertaken after the review should be documented as a supplementary search, with its own date, strategy, and results. Requiring accessible full text may also have introduced availability bias. Because the historical records do not preserve complete identification, duplicate-removal, screening, and full-text exclusion counts, a conventional numerical PRISMA flow diagram cannot be reconstructed. Figure 1 therefore presents the documented selection pathway and the verified final inclusion count without assigning unsupported numbers to intermediate stages.

The included studies vary in design, population, intervention, language skill, outcome measure, duration, and context. Screening, extraction, and synthesis were conducted by one reviewer, and independent duplicate screening was not undertaken; selection and interpretation bias therefore cannot be ruled out. These differences also limit direct cross-study comparison and require caution when drawing conclusions about generalizability, long-term outcomes, or causal effectiveness. Some recent studies have not yet accumulated substantial replication.

6. Conclusion

Generative artificial intelligence is becoming a visible component of secondary language education, particularly in writing, feedback, conversational interaction, self-directed learning, vocabulary learning, and AI-literacy activities. The 22 included studies indicate potential across these roles but do not support a uniform claim of learning effectiveness. The educational contribution of GenAI appears to depend on the task, the language-learning activity, learner characteristics, teacher mediation, and implementation conditions.

Ethics statement

This review synthesizes previously published studies and involved no direct recruitment, intervention, or interaction with human participants by the review author. Institutional research ethics approval was therefore not required for the review itself.

Conflict of interest

The author declares no known financial or non-financial conflicts of interest that could have influenced the conduct, interpretation, or reporting of this review.

Funding

No external funding was received for this review.

Data availability

The evidence synthesized in this review is drawn from the studies cited in the reference list. The manuscript presents the study-characteristics and cross-dimensional synthesis tables used to summarize the included evidence.

Author contribution

The author was responsible for the review conception, methodology, literature identification and source verification, data extraction, synthesis, and manuscript preparation.

7. References

Cabato, J. U. (2025). From awareness to practice: Exploring the knowledge, attitudes, and practices of secondary ESL teachers in the Philippines toward ChatGPT in education. *LatIA*, *3*, Article 360. https://doi.org/10.62486/latia2025360

Farhane, H. (2025). The impact of ChatGPT as a formative feedback tool on Moroccan secondary school students' paragraph writing skills: A quasi-experimental design. *Journal of English Language Teaching and Applied Linguistics*, *7*(3), 258–275. https://doi.org/10.32996/jeltal.2025.7.3.26

Farhane, H. (2026). Moroccan secondary school students' perceptions and attitudes toward using ChatGPT-assisted feedback to improve English speaking skills. *Journal of English Language Teaching and Applied Linguistics*, *8*(8), 6–19. https://doi.org/10.32996/jeltal.2026.8.8.2

Gok Acan, Y., & Liman Kaban, A. (2026). Human vs. machine feedback: Evaluating ChatGPT-4 in the assessment of secondary EFL learners' writing. *Asia Pacific Journal of Educators and Education*, *41*(1), 159–187. https://doi.org/10.21315/apjee2026.41.1.8

Javier, D. R. C., & Moorhouse, B. L. (2024). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, *15*(2), e755. https://doi.org/10.1002/tesj.755

Kök, M. (2026). Generative artificial intelligence literacy training for 12th grade English as a foreign language students: Effects on artificial intelligence literacy, ChatGPT usage, and autonomous English learning. *HAYEF: Journal of Education*, *23*, 1–10. https://doi.org/10.5152/hayef.2026.26029

Lariosa, J. J., Bajao, K. S., Talaroc, J. D., Amancio, E. D., Buna, J. C., & Villarin, S. J. B. (2025). Usage of ChatGPT on English language skills of students at Maputi Senior High School. *Psychology and Education: A Multidisciplinary Journal*, *43*(3), 368–378. https://doi.org/10.70838/pemj.430309

Lee, H. (2026). Teacher intervention in K-12 AI-based instruction: A systematic review of processes, strategies, and effects. *Smart Learning Environments*, *13*, 35. https://doi.org/10.1186/s40561-026-00461-1

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, *34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

Li, B., Tan, Y. L., Wang, C., & Lowell, V. L. (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

Li, Y., Shadiev, R., & Chiu, T. K. F. (2026). 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, Z. (2026). The impact of GenAI chatbots on communicative competence in EFL writing in Chinese upper secondary school. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2026.2685753

Ma, Q., Chen, S., Lo, W. M. S., Zou, D., & Pérez-Paredes, P. (2026). Generative AI as a learning companion for self-directed second language learning among secondary school students: A mixed-methods study. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70375

Maghamil, M., & Sieras, S. (2024). Impact of ChatGPT on the academic writing quality of senior high school students. *Journal of English Language Teaching and Applied Linguistics, 6*(2), 115–128. https://doi.org/10.32996/jeltal.2024.6.2.14

Mai, H. T. (2026). Evaluating the impact of using ChatGPT on EFL students' argumentative skills. *Discover Education, 5*, Article 189. https://doi.org/10.1007/s44217-026-01153-1

Ozan, O., & Azap, S. (2026). Employing ChatGPT to improve high school students' writing skills by providing feedback on topic-specific writing tasks. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2026.2613219

Pei, L., Jong, M. S.-Y., Huang, B., Pang, W.-C., & Shang, J. (2025). Formally integrating generative AI into secondary education: Application of ChatGPT in EFL writing instruction. *Educational Technology & Society, 28*(3), 281–297. https://doi.org/10.30191/ETS.202507_28(3).TP05

Pérez-Montesdeoca, H., Rodríguez-Rodríguez, D., & Fernández-Sogorb, A. (2026). The integration of generative artificial intelligence in secondary education: A systematic review. *Contemporary Educational Technology, 18*(3), Article ep667. https://doi.org/10.30935/cedtech/18746

Pham, T. A., & Pham, C. H. (2026). Vietnamese secondary school EFL students' motivation and attitudes toward using ChatGPT for vocabulary learning. *International Journal of Evaluation and Research in Education, 15*(3), 2708–2715. https://doi.org/10.11591/ijere.v15i3.38603

Shim, D., & Kim, H. (2024). The effects of using ChatGPT as a feedback tool on high school students' English writing. *Multimedia-Assisted Language Learning, 27*(1), 25–50. https://doi.org/10.15702/mall.2024.27.1.25

Wang, H., Cui, Y., & Fang, H. (2026). Identity tensions of EFL teachers in navigating GenAI-empowered language teaching: A case study of two secondary school teachers in China. *International Journal of Applied Linguistics, 36*(3), 2648–2659. https://doi.org/10.1111/ijal.70106

Wang, Y., & Shi, H. (2026). Exploring GenAI affordance in EFL continuation task instruction for developing Chinese high school students' writing complexity, accuracy and fluency. *Discover Computing, 29*, Article 373. https://doi.org/10.1007/s10791-026-10241-7

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Woo, D. J., Guo, K., & Susanto, H. (2025). EFL secondary students' use of ChatGPT for writing task completion pathways. *The Journal of Educational Research*, *118*(6), 596–609. https://doi.org/10.1080/00220671.2025.2510382

Woo, D. J., Wang, D., Guo, K., & Susanto, H. (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

Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing*, *67*, 101179. https://doi.org/10.1016/j.jslw.2025.101179

Żammit, J. (2025). Secondary school teachers' experiences with generative AI in Maltese language teaching. *Technology, Knowledge and Learning*, *30*, 2103–2121. https://doi.org/10.1007/s10758-025-09861-7

Zheng, Y. (2024). The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. *Computers and Education: Artificial Intelligence*, *7*, 100271. https://doi.org/10.1016/j.caeai.2024.100271