Journal of Interdisciplinary Studies in Education

Volume 15, Issue 6 (2026), pp. 47-78
Journal of Interdisciplinary Studies in Education
ISSN: 2166-2681Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/w506dk33

Artificial Intelligence in School Language Education: A Systematic Review of Trends, Challenges, and Future Directions

Muhammad Imran
Educational Research Lab, Prince Sultan University, Saudi Arabia
English Language and Literature Department, Khazar University, Azerbaijan
ORCID: 0000-0002-8754-2157

Norah Almusharraf
Education Research Lab, Prince Sultan University, Saudi Arabia
ORCID: 0000-0002-6362-4502

Khurram Shehzad
Software College, Northeastern University, Shenyan, China
FG Public School Boys No.1, Okara Cantt., Pakistan
ORCID: 0000-0002-7129-5876

ABSTRACT

Artificial intelligence (AI) is increasingly transforming language education, yet research on its pedagogical integration in school settings remains limited. This systematic review synthesizes findings from 20 studies published between 2018 and 2025, selected using the PRISMA framework from Web of Science, Scopus, and Google Scholar. The review examined AI technologies, pedagogical integration, and developmental outcomes in elementary and secondary language education. Results identified four major AI functions: evaluative feedback, adaptive scaffolding, generative language support, and conversational assistance. Most studies focused on writing, with limited attention to speaking, listening, and reading. AI proved most effective when integrated into teacher-guided instruction, while learner self-regulation influenced outcomes. Key challenges included algorithm reliability, institutional readiness, data privacy, academic integrity, and

equitable implementation, highlighting the need for human-centered, ethically guided AI integration.

Keywords: Adolescent learners, AI-assisted language learning, artificial intelligence in education, school language learning, self-regulated learning, systematic review

INTRODUCTION

In the twenty-first century, language learning has become an important element for individuals and for nations as a whole (Liang et al., 2023). Some individuals learn a second language to fulfill educational requirements, others to meet workplace requirements, and others to communicate effectively with people from other nations. The integration of Artificial Intelligence (AI) into educational contexts represents a paradigm shift that offers unprecedented opportunities to personalize learning, automate administrative tasks, and provide real-time, actionable feedback (Benkhalfallah et al., 2024). Within this broad landscape, language education stands as a particularly fertile ground for AI application (Sharmin et al., 2026). Technologies such as Natural Language Processing (NLP), Intelligent Tutoring Systems (ITS), and conversational agents (chatbots) directly target core language skills like listening, speaking, reading, and writing, with a strong impact on the learning experience (Gutiérrez, 2023; Gull et al., 2020).

Scholars have explored the potential of AI in elementary and secondary education. The youth of this digital era are more proficient technology users than people aged above 40 years (Roll & Wylie, 2016). The cognitive demands of language education have shifted from basic comprehension to critical analysis and fluent production, making it an ideal context for AI-driven support (Wiboolyasarin et al., 2025). Tools for automated writing evaluation can provide iterative feedback on essays, speech recognition software can aid pronunciation practice, and adaptive platforms can tailor vocabulary acquisition to individual students' needs; these tools address persistent challenges in student engagement and achievement gaps (Dekhakhena, 2025; Pennington & Rogerson-Revell, 2018). Moreover, these tools provide instant feedback, 24/7 availability, and personalized learning opportunities, which are crucial for developing cognitive, behavioral, and

emotional engagement throughout the language-learning process (Wiboolyasarin et al., 2025).

Over the last decade, from 2015 to 2025, advances in AI tools and their greater availability have prompted numerous studies exploring the effects and applications of these tools in secondary school language education. A growing number of systematic and literature reviews have begun to synthesize evidence on the effectiveness of AI in education broadly (Mustafa et al., 2024) and in language learning specifically (Zhai, 2025). Existing syntheses often focus narrowly on quantitative learning outcomes or specific technologies in isolation, frequently within higher education or adult learning contexts (Martin et al., 2020). Despite numerous research studies in this domain, there is a conspicuous lack of a comprehensive, systematic review in the elementary and secondary education context to measure AI trends in language learning, identify research gaps, and outline future applications.

To achieve efficient benefits from AI Technology, particularly in language learning in the school education context, it is necessary to overcome technical, socio-technical, and stakeholder-related challenges. For the successful implementation of educators' willingness, compatibility of the curriculum with new technology, data privacy issues, and equitable access to technological tools are crucial; these are the main barriers that should be understood; otherwise, fruitful results cannot be achieved (Waqar et al., 2023). In the literature, these issues were discussed, but these studies did not provide a clear way to resolve them, thereby enabling the smooth implementation of AI for language learning in elementary and secondary-level settings.

Therefore, we conducted this systematic review to fill this research gap by synthesizing research trends from 2018 to 2025. This systematic review paper moves beyond questions of "what works" to explore how AI is being implemented, which challenges are being addressed in the school teaching environment, and, after identifying research gaps, where future research should be conducted. To achieve this goal, this systematic review paper aims to answer the following research questions:

RQ1: What are the technological and pedagogical trends of AI in language education for elementary and secondary Schools?

RQ2: What challenges are reported in the literature regarding the integration of AI tools into elementary and secondary school language learning environments?

RQ3: What research gaps and future directions have been identified in the literature concerning the integration of AI in elementary and secondary school language education?

With the help of these research questions, we contribute in three different ways: firstly, this research will provide a state-of-the-art map of the current AI-driven language learning landscape for students at the secondary and elementary levels, secondly, this study offers researchable insights for teachers, policymakers, and curriculum designers, and thirdly, by establishing a robust research agenda to guide future scholarly inquiry in this rapidly evolving domain. The remainder of this paper is structured as follows: the methodology section outlines the systematic protocol; the results present the findings of the review; the discussion interprets these findings and sets a future agenda; and the conclusion summarizes the main implications.

LITERATURE REVIEW

In the modern age of cutting-edge technology, the use of AI tools has increased across all educational levels, particularly in language learning, which has emerged as a pressing field for second-language learning. Numerous studies have explored the impact of AI on language learning, providing valuable guidance for future research directions. This section critically examines existing syntheses across three overarching domains: (1) the broader landscape of AI in Education (AIED) to find the technological and pedagogical trends in this domain, (2) the specific application of AI in language learning, and (3) the focus on elementary and secondary school education contexts. The systematic review aims to crystallize the distinct contributions by analyzing the scope and limitations of past research in this domain.

Syntheses of Artificial Intelligence in Language Education (AILE)

Regarding AI in language education, researchers have conducted systematic reviews from various perspectives. Scholars have conducted systematic reviews on AI in language education from different perspectives. Ji et al. (2023) investigated the role of teachers in AI-integrated language teaching by examining 24 empirical studies on the collaboration between conversational AI and language teachers from 2015 to 2021. They found that although there is extensive discussion of the benefits of AI-integrated classrooms, there remains a notable paucity of empirical evidence delineating specific strategies for leveraging AI-human teacher collaboration to improve teachers' teaching efficiency.

In recent years, with the emergence of AI tools like large language models and Generative AI, interest in language learning with AI assistant tools has increased manyfold. Studies conducted between 2020 and 2025 demonstrate that AI chatbots play a key role in enhancing learners' speaking and writing skills, as well as their motivation (Zhang et al., 2025). For instance, the study by Zhang et al. (2024) found that AI-driven speaking exercises improved learners' motivation

and willingness to communicate (WTC); similarly, the study by TRÂN (2025) revealed significant benefits for both vocabulary growth and grammar accuracy. Another study explores the trends and potential of AI in STEM and language education, demonstrating AI's role in supporting personalized learning for students (Imran et al., 2024a). Although these studies show the potential of AI, their scope is limited, revealing a clear gap in the exploration of domain-specific challenges, such as pedagogical considerations in language acquisition. Moreover, most recent studies have focused on the overall effects of AI in language learning across all academic levels and age groups, from primary to university, but a special focus on the elementary and secondary levels has remained obscured.

The Convergence of AI and Language Learning

AI and language learning are often conflated with AI-assisted Language Learning, a common misconception. Imran and Almusharraf (2023) conducted a systematic review to explore the role of AI-based applications in language learning, particularly automated writing evaluation (AWE) and intelligent tutoring systems. Several reviews have focused on single technologies. For example, Barrot (2023) provided a comprehensive review of AWE tools, analyzing their impact on writing performance and students' perceptions. While insightful, such a singular focus cannot capture the ecosystem of interdependent tools (e.g., chatbots for speaking practice and AI-powered vocabulary apps) emerging in modern classrooms. Other reviews (e.g., Kessler, 2024) have taken a more pedagogical approach, discussing how AI aligns with established theories of second language acquisition. These reviews are strong on theory but often lack a systematic analysis of the empirical evidence base, particularly regarding implementation challenges.

The Neglected Context: School Education

The essence of this gap that this review aims to address is apparent in the educational context. Most existing syntheses of AIALL primarily draw on research on university or adult learning contexts (Ouyang et al., 2022). This leaves a significant limitation for two reasons. First, the school students are at a different cognitive and developmental stage than adults. Their metacognitive skills, self-regulation, and academic motivations vary, which influences how they utilize and benefit from AI tools (Dahri et al., 2024). An AI tool that may work well for adult, autonomous learners may overwhelm or be abused by adolescents unless it is effectively scaffolded (Bailey et al., 2025). Second, there are unique structural constraints of the elementary and secondary school environment that are largely non-existent in higher education. These include standardized curricula, high-stakes testing, fixed class schedules, and multiple degrees of technology infrastructure

and teacher readiness (Imran et al., 2024b). Integrating AI effectively is therefore not only a matter of effectiveness but also one of feasibility in this complex ecosystem. Issues of data privacy and algorithmic bias are even more pronounced when users are minors; ethical questions are often neglected when reviews focus on adult learners.

RESEARCH METHOD

Literature Search

This important section describes the overall adopted research design for this systematic review in the field of AI in language education. For this review study, we selected papers from 2018 to 2025 in the field of AI in language education from the Web of Science (WOS), Scopus, and Google Scholar databases, which are highly cited in academia. We adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Principles due to its wide range of acceptable ratios for systematic review papers to ensure the transparency and reproducibility of this research work (O'Dea et al., 2021). To collect data for this systematic review paper, we searched the "Topic" field, including "title, "abstract," authors keyword," and "keywords plus," in the Scopus database on AI in Language Education at the secondary level between 2018 and 2025. All the search strings we used to identify studies in this domain are listed below. The search string was selected to meet the criteria for articles frequently cited on AI in Language Education. The Scopus search returned 447 records. The search Strings used for the selection process are mentioned below.

From Scopus: TITLE-ABS-KEY ("artificial intelligence" OR AI OR "machine learning" OR elearning OR "personalized learning" OR "generative AI") AND TITLE-ABS-KEY ("elementary school" OR "secondary school" OR "junior high school" OR "junior high schools" OR "junior high school students") AND TITLE-ABS-KEY ("language learning" OR "second language" OR "foreign language" OR reading OR "writing instruction" OR "automated assessment") AND PUBYEAR > 2018 AND PUBYEAR < 2026

From Web of Science: TS=(("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "natural language processing" OR NLP OR "intelligent tutoring system*" OR chatbot* OR "conversational agent*" OR "automated writing evaluation" OR AWE OR "generative AI" OR "large language model*" OR LLM* OR "speech recognition" OR "adaptive learning") AND ("secondary school*" OR "high school*" OR "middle school*" OR "junior high" OR "K-12" OR adolescen*) AND ("language learning" OR "language education" OR "second language" OR "foreign language" OR EFL OR ESL OR "language acquisition"

OR writing OR speaking OR reading OR vocabulary)) AND PY=(2018-2025) AND LA=(English) AND DT=(Article OR Proceedings Paper)

From Google Scholar: ("artificial intelligence" OR "generative AI" OR chatbot OR "automated writing evaluation" OR "intelligent tutoring system") AND ("secondary school" OR "high school" OR adolescent) AND ("language learning" OR EFL OR ESL OR "foreign language")("large language model" OR "ChatGPT" OR "generative AI") AND ("secondary school" OR "high school") AND (writing OR speaking OR "language education")

Inclusion and Exclusion Criteria

According to the PRISMA 2020 guidelines, the systematic search of the literature in three large databases, namely Web of Science (n = 416), Scopus (n = 447), and Google Scholar (n = 322), was conducted, resulting in 1185 records. After eliminating 237 duplicate records, 948 distinct records were screened based on titles and abstracts. At the title and abstract screening stage, 891 records were filtered out based on the preset inclusion and exclusion criteria. The following reasons were used to exclude them: the wrong population was used (not elementary and secondary school students), the technology used was not artificial intelligence, the study design was non-empirical, or there were no measurable language-learning results. Of the other 57 full-text articles, the eligibility was evaluated. Following full-text analysis, 37 articles were eliminated due to the following reasons: not carried out in the context of elementary and secondary schools (n =16), non-AI technologies (n = 9), review or theoretical articles (n =3), inappropriate or non-measurable outcomes (n =7), and non-English publications (n =2). This resulted in 20 studies being used in the final synthesis of the systematic review (Figure 1), as they all satisfied the inclusion criteria.

Coding Scheme

The authors developed a structured coding scheme based on established classification schemes in the field of Artificial Intelligence in Education and on prior systematic reviews of AI-assisted language learning. The coding categories were explicitly linked to the research questions, thereby ensuring transparency, consistency, and methodological replicability in data extraction.

Publication Characteristics

Each study was coded for year of publication, country/region of origin, research design (e.g., randomized controlled trial, quasi-experimental, mixed methods, case study), and sample characteristics (grade level and sample size).

Educational Context

The inclusion criteria limited the corpus to studies conducted in the elementary and secondary education setting (Grades 6-12 or equivalent, ages 11-19). Studies conducted only in a setting (primary, tertiary, or adult education) were systematically excluded.

AI Technology Typology

The identified AI artifacts were classified under Automated Writing Evaluation (AWE), Automated Written Feedback (AWF), conversational agents or chatbots, speech recognition systems, adaptive learning systems or Intelligent Tutoring Systems (ITS), generative AI/Large Language Models (LLMs), explainable AI (XAI), and AI-driven analytics tools. In instances where studies included more than one concomitant technology, a key pedagogical AI function was assigned to the principal coding category.

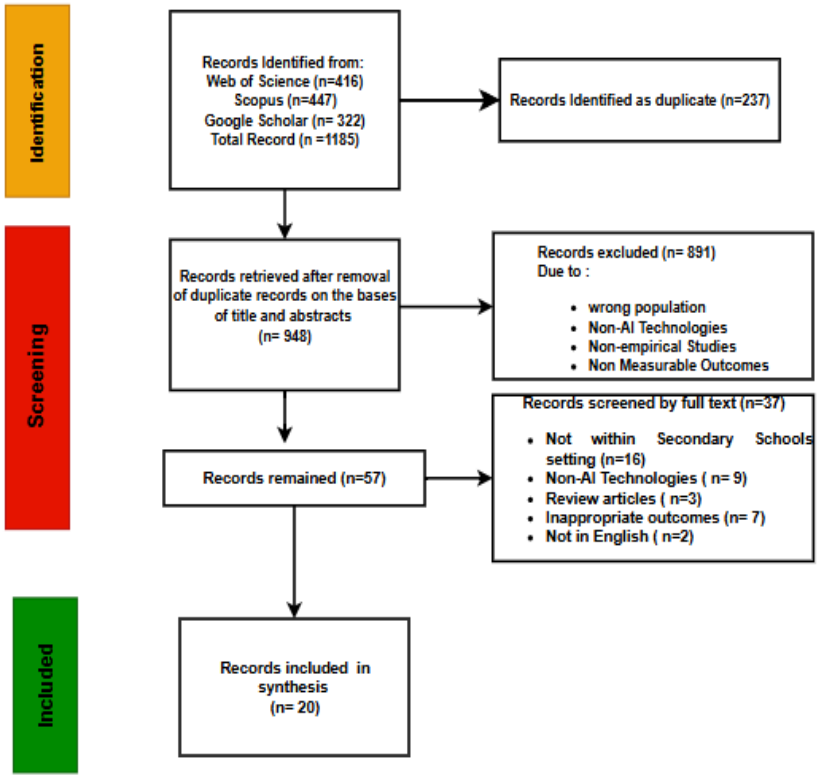

Figure 1: Articles search and selection based on PRISMA

Core AI Function

Instructional AI function activities were divided into automated feedback, error detection, text generation, adaptive recommendations, proficiency estimation, dialogue simulation, discourse analytics, and support for engagement or self-regulated learning.

Pedagogical Integration Mode

Studies were coded using the implementation model, including core curricular (teacher-mediated), hybrid or structured instructional, supplementary or student-directed, targeted intervention in class, informal or unsanctioned, and system-level or professional development integration.

Targeted Language Focus

The main language focus was coded as writing, speaking, reading, vocabulary, integrated skills, learning processes (such as metacognition, engagement), or a teacher's discourse.

Learning Outcomes

Reported learning outcomes were categorized as cognitive (achievement or performance), affective (e.g., motivation, anxiety), behavioral (engagement or usage patterns), or metacognitive/self-regulatory.

Reported Challenges

Implementation obstacles were identified as technical, pedagogical, ethical, equity-related, privacy or security-related, teacher preparation or readiness, or policy misalignment.

A structured extraction matrix was used for coding all studies. Coding discrepancies were resolved by discussion to ensure conceptual consistency.

Inter-Coder Reliability

All of the included studies were coded by two independent reviewers. Cohen's Kappa coefficient on primary coding categories was used to compute inter-rater reliability. The overall consensus was high ($k = 0.86$), indicating a high level of reliability by existing standards. The value of k at the category level ranged from 0.82 to 0.91. The coding differences were solved by discussion until agreement was reached.

Quality Appraisal

All the included studies were evaluated based on the 2018 version of the Mixed Methods Appraisal Tool (MMAT)(Hong et al., 2018). The analysis covered the clarity of the research questions, the transparency of the sampling process, the appropriateness of the data collection procedures, the appropriateness of the outcome measures, and the coherence between the data and the resulting conclusions. The overall quality of methodology was moderate to high. Commonly encountered limitations included short intervention periods, limited longitudinal follow-up, and low ecological validity due to single-site implementation. Quality appraisal was not used to exclude any studies; however, quality considerations were used to interpret the strength of evidence during the synthesis.

RESULTS

General Findings

The findings show that the geographic distribution of the 20 studies included in this systematic review is highly diverse. The distribution shows a very strong concentration of research activity in East and Southeast Asia, with China dominating most of it. A second group of works is connected to Norway, Indonesia, Japan, and the United States and proves moderate involvement in the Western and Asian academic environment. Conversely, donations in other parts of the world, such as Africa, the Middle East, and Latin America, are limited to solitary research, indicating a disjointed research presence. Taken together, the uneven geographic dispersion suggests that the existing evidence base on the integration of AI in school language education is dominated by a small number of national settings, underscoring the need for more geographically varied and comparative studies.

The time distribution of the included studies in that review shows a positive, steadily accelerating growth curve in AI-related studies in the field of language education in schools between 2018 and 2025. The first publication output remained low in 2018 and 2019, as it was an exploratory period during which initial attempts were made with AI-based instructional tools. Since 2020, the sector has shown consistent growth, with a sharp rise since 2022 and a peak between 2024 and 2025. This positive trend indicates a transition from peripheral uptake to the more formal implementation of AI technologies in secondary language situations. The faster pace of growth in later years is presumably driven by broader technological development and greater institutional involvement in generative and adaptive AI systems. Notably, the lack of decline over the study period suggests sustained scholarly momentum rather than momentary technological fervor,

making AI-based secondary language education a developing and stabilizing field of study.

The findings further illustrate that the study sample in the review was based on pedagogical learning modalities, namely those in which artificial intelligence was operationalized. These findings suggest that core curricular integration through the teacher is the most commonly used, followed by hybrid or designed instructional models that integrate AI tools with teacher scaffolding. The additional student-initiated implementations and specific classroom interventions are found in the middle frequency, with possible exploratory or skill use. Surprisingly, an equal number of studies document the informal or unauthorized use of AI tools by students, indicating a divergence between the institution's policy and students' behavior. System-level or professional development-oriented deployments are infrequently represented, suggesting that the existing body of research has mostly focused on classroom-level integration rather than scalable institutional adoption models.

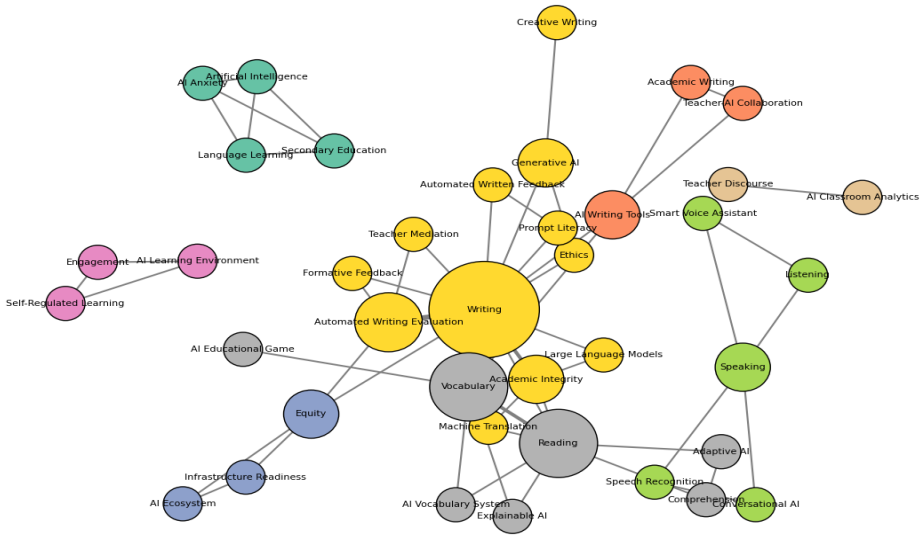

Figure 2: Clustered keyword co-occurrence knowledge map generated using Louvain modularity optimization. Node size reflects keyword frequency, edge thickness indicates co-occurrence strength, and colors represent thematic clusters.

The co-occurrence network keyword was created from the author keywords extracted from the 20 included studies, as shown in Figure 2. Only those words used at least twice were selected in order to minimize noise and improve conceptual interpretability. The co-occurrence relations were calculated on a

document level. A directed graph was generated where the nodes were represented by keywords, whereas the values of the edges were the frequency of co-occurrence. They were identified using the Louvain modularity algorithm. To identify conceptually structurally influential concepts, network measures, such as degree centrality, betweenness centrality, and eigenvector centrality, were computed. A force-directed layout was used to create the visualizations, which were then exported at 600 DPI to meet publication requirements.

Moreover, the findings identify the key language skills that AI-based interventions should address in secondary school language education. Writing is the predominant focus, as most of the studies reviewed emphasize it, reflecting the maturity and accessibility of AI technologies for providing text-based feedback and assessment. Conversely, comparatively little attention is given to speaking and reading, as well as vocabulary growth; in addition, very few of the studies are multi- and process-oriented, with the focus on the regulation or engagement of learning. There is a major gap in the literature, as little has been explored on teacher discourse and instructional practices. All in all, the distribution suggests a significant research bias in writing outcomes and underscores the need for more comprehensive studies that encompass oral skills, receptive competencies, and teacher-mediated processes.

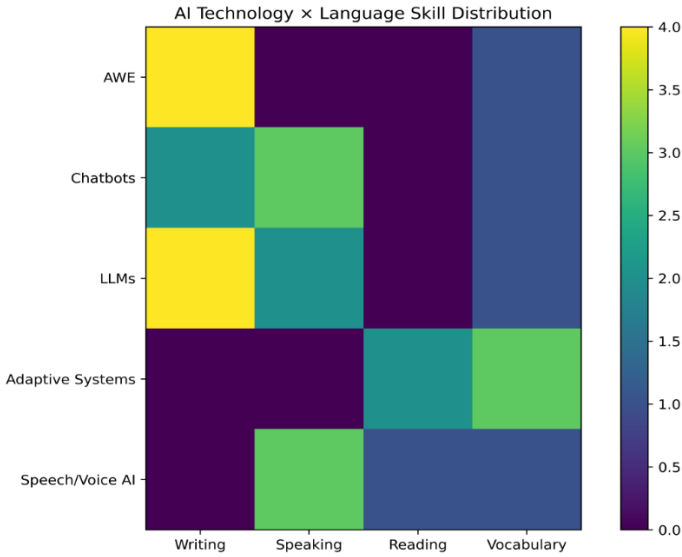

Figure 3: Heatmap showing concentration of AI applications in writing, especially AWE and LLMs, with comparatively limited attention to reading and integrated skill development.

The prevalence of AI technologies in language skills (Figure 3) shows a clear imbalance, both in technological application and in pedagogical focus. Writing is cited as the most prevalent field, where Automated Writing Evaluation systems and Large Language Models show the highest concentration of applications; thus, the field focuses on text-based performance and automated feedback systems.

On the other hand, speaking-related applications are secured by chatbots and speech-based AI and, on the other hand, are fairly represented but not as developed. Most AI types involve few direct readings, which creates a significant gap in current research on AI-supported comprehension procedures in secondary schools. The adaptive systems show somewhat better vocabulary development, as they are more adaptable to personalized lexical practice, but the overall integration of various abilities remains limited. The Heatmap suggests that modern studies have disproportionately focused on measurable written production and, as a result, have placed less emphasis on integrated communicative abilities, thereby heightening the need to adopt more multi-skilled AI interventions in elementary and secondary language education.

Main Findings

RQ1: What are the technological and Pedagogical Trends of AI in Language Education for Secondary and elementary Schools?

As shown in Table 1, the corpus of studies (20 studies) has a decisive turn toward disaggregated, instrument-focused applications of artificial intelligence, with a focus on models of pedagogical mediation and systematic integration into elementary and secondary language education. Early and subsequent applications focused more on using AI as a student-friendly assistance tool, such as machine translation, automated writing feedback, and gamified vocabulary training, which mostly operated outside formal learning environments (e.g., Wen et al., 2024; Organ, 2023; Andreani & Ying, 2019). Although these interventions showed short-term benefits in linguistic accuracy and student engagement, they also raised enduring concerns about academic dishonesty, cognitive overdependence, and unfair learning among students when teacher mediation was either absent or limited.

Table 1: *Technological–Pedagogical Trends of AI Use in Elementary and Secondary School Language Education (n = 20)*

No.	Study (Country)	AI Technology	Core AI Function(s)	Pedagogical Integration Model	Target Language Skill(s)	Dominant Research Trend
1	(Wen et al., 2024) China	AI Language Learning Apps	MT, writing correction, personalized feedback	Supplementary, student-directed	Writing, Listening, Speaking, Vocabulary	AI acceptance & anxiety
2	(Krange et al., 2023) Norway	AWE System	Error detection, formative feedback	Core curricular, teacher-mediated AfL	Writing (Mechanics)	Teacher-scaffolded AI feedback
3	(Gupta & Yadav, 2025) India	Multi-AI Ecosystem	SR, ITS, chatbots, adaptive platforms	Proposed systemic integration	All four skills	AI for equity & infrastructure readiness
4	(Junaidi, 2020) Indonesia	Conversational AI	Speech recognition, dialogue practice	Targeted classroom intervention	Speaking	Direct efficacy testing
5	(Organ, 2023) UK	Free Online MT	Text translation	Informal, unsanctioned use	Reading, Writing	Academic integrity & policy gap
6	(Engeness, 2025) Norway	AWE System	Structured formative feedback	Core curricular, theory-driven	Writing	AI as cultural–mediational tool
7	(Obateru & Daramola, 2025) Nigeria	Generative AI Tools	Text generation, idea expansion	Informal student use	Writing	AI misuse & ethics literacy
8	(Zhu et al., 2024) USA	LLMs	Essay generation, task assistance	Informal, prohibited use	Writing, Literacy	Pervasive unsanctioned adoption
9	(Alghasab, 2025) Kuwait	Multi-AI Writing Tools	Ideation, vocabulary, accuracy	Supplementary, student-initiated	Writing	Strategic multi-tool orchestration
10	(Lee & Son, 2024) South Korea	AWF Tools	Automated written feedback	Structured instructional intervention	Writing	Prompt literacy dependency
11	(Takii et al., 2024) Japan	XAI Recommender	Difficulty estimation, explanations	Integrated ER program	Reading, Vocabulary	Trust-building via explainable AI
12	(Jaramillo et al., 2025) Colombia	AI Writing Tools	Grammar, coherence, citation support	Hybrid instructional model (WWIM)	Academic writing	Teacher–AI pedagogical synergy
13	(Shi et al., 2025) China	AI Learning Environment	SRL & engagement support	Longitudinal program integration	Learning processes	Process-oriented AI evaluation
14	(Lei, 2022) China	AI Vocabulary System	Cognitive structuring, processing	Core instructional replacement	Vocabulary, Reading	Comparative instructional efficacy

No.	Study (Country)	AI Technology	Core AI Function(s)	Pedagogical Integration Model	Target Language Skill(s)	Dominant Research Trend
15	(Chu et al., 2024) China	AI Classroom Analytics	Discourse mining, DL analysis	Teacher professional development	Teacher discourse	AI-driven teaching analytics
16	(Woo et al., 2024) Hong Kong	Generative AI	Idea generation, co-creation	Structured workshop-based use	Creative writing	AI as co-creative partner
17	(Wilson et al., 2024) USA	AWE System	Trait scoring, formative feedback	System-wide benchmark integration	Writing	Equitable engagement hypothesis
18	(Alqarni & Alhramelah, 2025) Saudi Arabia	Smart Voice Assistant	Interactive oral feedback	Designed affective intervention	Listening, Speaking	Anxiety reduction via AI
19	(Andreani & Ying, 2019) Indonesia	AI Educational Game	Gamified feedback, tracking	Supplementary mobile learning	Vocabulary	Gamified micro-learning
20	(Malakul & Park, 2023) Japan	Adaptive AI Reading System	Proficiency estimation, content adaptation	Integrated subtitle video reading support	Reading, comprehension	Limitations of adaptive algorithms

RQ2: What challenges are reported in the literature regarding the integration of AI tools into elementary and secondary school language learning environments?

Table 2 summarizes the implementation issues documented in the 20 studies included in this review. Rather than presenting single instances of technical shortages, these dilemmas are presented as a multi-level mechanism that spans pedagogical, developmental, institutional, ethical, and algorithmic levels. Together, the facts suggest that the implementation of AI in school language education is not, in the first place, limited by technological possibilities; rather, it is limited by the complexity of relations with adolescent cognition, teacher mediation, the curriculum system, and governance schemes.

Table 2: Reported Implementation Challenges in AI-Supported Language Education

No.	Study	Reported Challenge(s)	Severity / Context
1	Wen et al. (2024)	Over-reliance on automated corrective feedback, reduced learner reflection, and limited independent transfer	Pedagogically Conditional: Requires structured teacher scaffolding
2	Krange et al. (2024)	Teacher interpretation difficulties of AI analytics; integration complexity; infrastructural demands	Institutional Readiness Barrier: Effective use depends on training and system support
3	Gupta & Yadav (2025)	Uneven learner engagement; dependency on AI prompts; variability in autonomy	Developmental Constraint: Outcomes mediated by self-regulated learning capacity

No.	Study	Reported Challenge(s)	Severity / Context
4	Junaidi et al. (2020)	Limited contextual transfer beyond AI-supported tasks; short-term intervention scope	Transferability Limitation: Sustainability beyond experimental setting is unclear
5	Organ (2022)	Ethical ambiguity in AI-mediated assessment; authorship transparency concerns	Academic Integrity Issue: Requires explicit governance policies
6	Engeness (2025)	Tension between AI scaffolding and independent cognitive development; need for reflective mediation	Instructional Balance Challenge: AI must complement—not replace—human guidance
7	Alqarni & Alhramelah (2025)	Data privacy and security concerns; system transparency limitations	Ethical–Governance Concern: Institutional data protection frameworks required
8	Obateru & Daramola (2025)	Technical reliability constraints; uneven personalization accuracy	Algorithmic Limitation: Adaptive precision varies across learner profiles
9	Wilson et al. (2024)	Cognitive overload from multimodal AI feedback; fragmented learner attention	Cognitive Load Risk: Feedback calibration necessary
10	Zhu et al. (2024)	Limited long-term evidence of skill retention; controlled research conditions	Research Scope Limitation: Broader ecological validation required
11	Alghasab (2025)	Potential reduction in critical thinking when AI suggestions are passively adopted	Cognitive Dependency Risk: Reflective engagement essential
12	Lee & Son (2024)	Teacher preparedness gaps; curricular alignment challenges	Implementation Barrier: Professional development required
13	Takii et al. (2024)	Short-term motivational gains without sustained longitudinal tracking	Sustainability Limitation: Long-term impact remains uncertain
14	Jaramillo et al. (2025)	Superficial revision when AI feedback is uncritically accepted; need for guided reflection	Pedagogically Conditional: Structured mediation supports deep learning
15	Shi et al. (2025)	Unequal metacognitive readiness; inconsistent long-term engagement	Developmental Variability: AI effectiveness linked to learner self-regulation
16	Lei (2022)	Limited ecological validity; insufficient longitudinal retention evidence	Contextual Limitation: Findings require classroom-based replication
17	Chu et al. (2024)	Technical complexity of AI discourse analytics; teacher training demands; data governance issues	System-Level Barrier: Institutional capacity determines success
18	Woo et al. (2024)	Blurred authorship boundaries in generative AI writing; diminished student voice risk	Ethical–Creative Tension: Responsible usage frameworks required
19	Andreani & Ying (2019)	Limited durability of gamified vocabulary gains; restricted contextual transfer	Motivational but Limited Depth: Curriculum integration necessary
20	Malakul & Park (2023)	Imperfect reading proficiency estimation; cognitive load from adaptive subtitles	Technical–Cognitive Constraint: Algorithm refinement needed for equitable personalization

First: Pedagogical Mediation and Cognitive Dependency Risks. The concurrent theme in the corpus concerns the conditional usefulness of AI-generated feedback. Some studies claim that automated corrective systems, especially in writing situations, can reduce learners' reflection when adopted passively (e.g., Wen et al., 2024; Jaramillo et al., 2025). Thoughtful revisions, blind following of AI recommendations, and overreliance on automated grammar correction were common in areas where formal teacher support was restricted.

Likewise, Engeness (2025) emphasizes the tension between AI-assisted scaffolding and the nurturing of independent cognitive processes, arguing that AI should be used as a mediating tool rather than a replacement for higher-level thinking.

The implications of these results are a more general dependence on cognition: if AI outputs are accepted as authoritative or definitive, learners can lose engagement in more metacognitive processing. This matter is especially relevant in the context of generative AI, where idea-building and drafting services may blur the distinction between support and authorship (Woo et al., 2024). The evidence thus highlights the need to incorporate AI into reflective pedagogical cycles, which need to be justified, compared, and revised AI-generated feedback.

Second: Developmental Variability and Self-Regulated Learning Constraints. Table 1 also presents significant variability in learners' responsiveness, which is mediated by inequalities in self-regulated learning (SRL) capacity, autonomy, and engagement levels (Gupta, 2025; Shi et al., 2025). Even though AI-supported environments can facilitate planning, monitoring, and revision procedures, the outcomes vary across learner profiles. Some students demonstrate statistically significant changes in engagement and metacognition awareness, while others plateau or show no change.

Such heterogeneity in development is further exacerbated in the adolescent environment of elementary and secondary schools. School students are still in the stage of consolidating executive functioning and regulatory competencies (unlike adult students). As a result, AI systems, which rely heavily on learner autonomy, can potentially discriminate against students with lower metacognitive preparedness. These statistics thus suggest that AI efficacy is not universal or invariant but rather conditional on the development stage and prior digital literacy, highlighting the necessity of differentiated scaffolding approaches.

Third: Technical and Algorithmic Limitations. Despite the common adaptive character of artificial intelligence technologies, multiple empirical studies indicate inherent technical limitations that undermine the accuracy of personalization and the reliability of feedback. For example, Malakul and Park (2023) report constraints in proficiency estimation in an adaptive subtitle delivery system, where discrepancies between subtitle delivery and a learner's ability can magnify cognitive load. Similarly, Obateru and Daramola (2025) confirm that personalization accuracy is not uniform across learner types, casting doubt on the algorithm's reliability and fairness.

Chu et al. (2024) considered both technical complexity and infrastructure as relevant challenges to successful implementation in classroom discourse analytics. These findings stress the need to go beyond algorithm refinement to effective deployment: not only is congruity with school-level technological resources and educational professionalism in the interpretation of analytic outputs required, but also, beyond that, further action. In this regard, technical

sophistication should be complemented by usability and transparency considerations to reduce cognitive overload and eliminate the risk of incorrect application.

Fourth: Institutional and Professional Readiness Barriers. One of the prominent clusters of challenges concerns institutional capacity and teacher preparedness. According to Krange et al. (2023), there is a challenge in interpreting AI-generated analytics and incorporating them into formative assessment practices. On the same note, Lee and Son (2024) reveal that gaps in teachers' AI literacy and curriculum fit cannot support sustainable integration. The findings align with broader systemic issues of infrastructural preparedness, professional development, and policy coherence.

Notably, Table 1 shows that the majority of reviewed research has been conducted at the classroom level, and there is little indication of system-wide or district-wide integration models used. In the absence of coordinated training or governance structures when AI systems are implemented, the implementation will be fragmented and reliant on teacher initiative. Therefore, institutional preparedness appears to be a structural factor in AI effectiveness rather than a peripheral logistical issue.

Fifth: Ethical, Privacy, and Academic Integrity Concerns. Ethical tensions are another important aspect of the challenges that were reported in the literature. The issues described in Organ (2023) and Zhu et al. (2024) highlight problems with the unsanctioned use of machine translation systems and large language models, which have blurred the boundaries of authorship and created policy uncertainties. In generative AI settings, the issues extend beyond plagiarism to include weaker student voice and originality (Woo et al., 2024).

Concerns related to data privacy and security are also anticipated, especially when AI systems gather and process data on learners' performance (Alqarni & Alhramelah, 2025). Since school students are minors, there is a greater ethical urgency in such concerns. However, few studies present in-depth governance structures that address data protection, transparency, and algorithmic bias. This gap suggests that ethical governance is not proactive; instead, it is responsive and not systematically integrated into AI plans.

Sixth: Research Design and Ecological Validity Constraints. Lastly, Table 2 highlights the methodological limitations that undermine the strength of the evidence base. Some studies acknowledge a limited sample size, a single-site experiment, or a short-term intervention (Malakul & Park, 2023; Lei, 2022). Limiting conclusions about long-term skill retention, extrapolation outside the experimental environment, and trends in long-term engagement is due to the paucity of longitudinal tracking. Furthermore, this fact raises concerns about ecological validity, as controlled designs and quasi-experiments predominate. The results of AI tools tested in a structured workshop environment might differ when scaled across classrooms with varying teacher experience and contrasting

infrastructure. Therefore, the existing evidence base has a promising future, but it should be replicated across more heterogeneous and broader educational settings.

RQ.3: What research gaps and future directions have been identified in the literature concerning the integration of AI in elementary and secondary school language education?

In light of the synthesis of empirical findings and identified implementation obstacles, the review outlines strategic directions for further research on the adoption of artificial intelligence (AI) in secondary school language education. Although current research evidence shows promising pedagogical gains, especially in AI-aided writing, adaptive vocabulary systems, and self-regulated learning environments, the current evidence base is imbalanced in scope, timing, and situational coverage. To overcome these structural constraints, future research should go beyond the tool-focused, experiment-only, isolated investigation and engage in theoretically informed, developmentally sensitive, and system-level studies.

First of all, the application of AI-assisted language skills and learning processes should be extended in the future. The current body of evidence is relatively concentrated on the instruction of writing, specifically, Automated Writing Evaluation (AWE), Automated Written Feedback (AWF), and generative AI systems (e.g., Woo et al., 2024; Jaramillo et al., 2025). Although such interventions show improvements in grammatical accuracy, revision behavior, and learner confidence, other areas of language, especially speaking, listening, reading comprehension, and teacher discourse, are less thoroughly researched. The literature on speech recognition systems and intelligent voice assistants (e.g., Alqarni and Al-hramelah, 2025) suggests that they can lead to decreased anxiety and oral proficiency, but there is limited longitudinal evidence. Similarly, AI-supported adaptive subtitle systems and reading environments (Malakul & Park, 2023) should be further explored across different proficiency levels and curriculum settings. Based on this, future studies should use integrated skill models to determine the extent to which AI supports multimodal communicative competence, rather than assessing discrete linguistic outcomes.

Second, the longitudinal and developmental research design has to be reinforced. Most of the literature uses short-term quasi-experimental or design-based designs, which are usually limited to individual classrooms or schools. Despite recent longitudinal modeling studies showing that the co-development of self-regulated learning (SRL) and task engagement (TE) in AI-assisted learning settings is modest but significant (Shi et al., 2025), growth patterns appear heterogeneous and depend on initial learner traits. There are ceiling effects among highly self-regulated students and variation in digital literacy profiles, suggesting that AI effectiveness depends on development. Subsequent research ought to

employ longitudinal designs spanning semesters or academic years, combining data from behavioral logs, discourse analytics, and classroom observation to triangulate self-report measures. These refinements in the methodology could help researchers obtain sustained improvements in learning, examine the effects of transfers beyond AI-mediated activities, and identify possible nonlinear patterns of development.

Third, the design and assessment of adaptive and explainable AI architectures that fit the needs of adolescent learners should be given a priority in future research. Some studies emphasize the role of personalization mechanisms (Lei, 2022; Takii et al., 2024), although adaptive accuracy varies across learner profiles. The higher level of challenge posed by static scaffolding might be insufficient for higher-ability learners and excessively challenging for lower-ability learners, due to superior metacognitive capacity (Shi et al., 2025). Besides, generative AI systems, although effective at extending ideas and as drafting aids (Woo et al., 2024), may encourage uncritical adherence to suggestions without careful consideration (Jaramillo et al., 2025). Further studies should thus consider adaptive feedback systems that adjust task complexity, feedback granularity, and metacognitive prompts in real time based on learners' changing profiles. They can meet explainable AI (XAI) capabilities that explain algorithmic suggestions, which can also build trust, reduce cognitive load, and enhance processing, particularly in a secondary context when people are still acquiring critical literacy.

Fourth, teacher mediation and professional integration should be investigated systematically. As has been shown, AI tools have been most effective when integrated into pre-structured pedagogies and teacher-directed reflection cycles (Jaramillo et al., 2025; Woo et al., 2024). On the other hand, unauthorized or unlicensed use of machine translation and large language models is associated with issues of academic honesty, authorship, and superficial revision (Organ, 2023; Zhu et al., 2024). Future studies should thus focus on teacher AI co-regulation models that incorporate automated feedback, dialogic scaffolding, and collaborative revision. The models of professional development that will improve teachers' AI literacy, data-interpretation competencies, and ethical governance capabilities will also be empirically tested. AI will not be adopted sustainably in support and alignment with the curriculum without institutional support.

Fifth, responsibility-use models and ethical governance should move to the forefront of the research agenda. As generative artificial intelligence becomes more involved in the ideation and creation of text, questions of authorship, originality, and learner agency are also magnified (Woo et al., 2024; Obateru & Daramola, 2025). Although some interventions include plagiarism detection and systematic ethics guidance (Jaramillo et al., 2025), comprehensive, systemic policies addressing data privacy, algorithmic bias, and adolescents' digital rights have not been well written. Since the impact of AI-based data collection, transparency mechanisms, and bias mitigation in language assessment and

feedback systems is of interest to scholarship, and because elementary and secondary learners are minors, this should be explicitly investigated. Pedagogical innovation should be guided by ethical design principles rather than by post hoc regulation.

Sixth, contextual equity and cross-cultural validation should receive more attention from scholars. There is a geographic distribution of studies under review, with a concentration in East and Southeast Asia and relatively low representation from Africa, Latin America, and low-resource educational settings. Several publications highlight infrastructural inequalities, the uneven distribution of digital connectivity, and inconsistent institutional preparedness as barriers to successful AI implementation (Gupta & Yadav, 2025; Obateru & Daramola, 2025). Future studies ought to be comparative and multi-sited to examine how socio-economic settings, school rigidity, language policy, and technological infrastructure mediate AI outcomes. Furthermore, AI systems should be tested across various linguistic contexts, including non-Indo-European languages, to reduce the risk of algorithmic bias and ensure fair personalization.

Lastly, disciplinary cross-fertilization on theoretical grounds is key to the field's development. Current research often separates technological functionality from the linguistic theory, developmental psychology, and the learning sciences. Further research should be grounded in theories of second-language learning, self-determination theory, and socio-cultural frameworks to conceptualize AI not as a feedback mechanism but as a mediation tool in dynamic learning ecologies. In practice, this will require interdisciplinary collaboration among computer scientists, linguists, educators, and policy scholars to develop AI systems that are pedagogically valuable, morally accountable, and developmentally sound.

Overall, the next step in research on AI applications in elementary and secondary school language education must shift from brief, tool-focused experimentation to thorough, theoretically informed, and equity-oriented research. By augmenting skill coverage, strengthening longitudinal designs, advancing adaptive and explainable designs, foregrounding teacher mediation, institutionalizing ethical governance, and addressing contextual variations, future research can provide a sustainable, humanistic framework for AI-enhanced language learning in school education.

DISCUSSION

The rapid uptake of artificial intelligence in language classrooms has outpaced the field's ability to systematically account for how, why, and under what conditions these tools actually support learning. Understanding this trajectory matters not only for researchers tracking an emerging technology but also for the teachers, administrators, and policymakers now making real-time decisions about tools whose pedagogical and developmental implications are still being established.

This review study explores the technological and pedagogical trends that are a foundation to AI-aided language education in elementary and secondary schools (RQ1), the multidimensional challenges of implementation as reported by a variety of schools (RQ2) and the research gaps and future direction as emerging from an evidence-informed synthesis of what is a rapidly evolving field (RQ3). The corpus demonstrates that the field is marked by accelerating innovation in tools and interfaces, but also exposes residual structural unevenness in the pedagogical embedding, developmental calibration, and institutional governance of these tools. Rather than a simple, expository description of AI's potential for language learners, the evidence, taken together, recasts the question as more fundamental: under what pedagogical, developmental, and systemic conditions can AI contribute meaningfully and equitably to language education in elementary and secondary schools? This discussion responds to that question by situating the findings within the broader scholarly contexts of Artificial Intelligence in Education and Second Language Acquisition (SLA), with the aim of shedding light on the theoretical contribution, practical consequences, and boundary conditions of the review's conclusions.

From Instrumental Adoption to Structured Pedagogical Integration

The corpus suggests a major shift from independent, additional use of artificial intelligence to a well-organized, mediated use by teachers. Informal applications were commonly used that were not within the scope of purposeful instruction design, most notably machine translation and unregulated generative writing tools (Organ, 2023; Zhu et al., 2024). Although these had short-term positive effects on accuracy or engagement (Wen et al., 2024; Andreani & Ying, 2019), they were always associated with superficial revision, cognitive dependency, and incomplete skill transfer (Alghasab, 2025; Jaramillo et al., 2025). This trend suggests that the greatest constraint is not technological ability but pedagogical practice.

Conversely, research reports longer-term results of incorporating AI into teacher-assessment feedback loops and formalized workshops (Krange et al., 2023; Engeness, 2025; Woo et al., 2024; Lee & Son, 2024). In these environments, automated writing assessment, automated feedback, and generative tools wore the role of mediators, as opposed to a single teacher, which can be described as in line with the principle of AIED that teachers should be used in learning in the form of pedagogical rather than isolated mediation (Martin et al., 2020; Mustafa et al., 2024). Teacher mediation as part of AI efficacy is a necessary, not an additional, element of metacognitive regulation, which is in its nascent stages in elementary and secondary education (Dahri et al., 2024).

Skill Distribution and Writing-Centric Bias

The corpus represents a high level of focus on writing education. There is a predominance of evidence on AWE, AWF, and generative AI (Krange et al., 2023; Jaramillo et al., 2025; Woo et al., 2024; Wilson et al., 2024), as these AI outcomes are computationally tractable (Liang et al., 2023). Although the enhancements in grammatical accuracy, revision behavior, and argumentative coherence have been sufficiently established, this bias narrows the skill ecology of language learning with AI assistance and may exaggerate the centrality of writing to communicative competence.

Intervention of speaking and listening is promising but underresearched. Smart voice assistants and conversational agents demonstrated the potential to improve oral fluency and reduce anxiety (Junaidi, 2020; Alqarni & Alhramelah, 2025; Zhang et al., 2025). Adaptive reading systems, vocabulary tools (Lei, 2022; Malakul & Park, 2023; Takii et al., 2024), and teacher discourse analytics (Chu et al., 2024) also have a narrow scope. From an SLA perspective, communicative competence is multimodal and interactional; a writing-dominated research agenda prioritizes a product-based measure over a combined communicative activity (Pennington & Rogerson-Revell, 2018). The expansion of AI use across all four language skills is thus critical to building an ecologically valid evidence base.

Developmental Contingency and Self-Regulated Learning

Practical studies invariably demonstrate that the effectiveness of artificial intelligence in education is not distributed equally; it depends on developmental parameters. Students who possess strong self-regulated learning (SRL) capabilities and digital literacy enjoy an asymmetrical advantage in AI-mediated settings (Gupta & Yadav, 2025; Shi et al., 2025), and those with weak metacognitive preparedness are unequally benefited or even opposed to automated feedback (Wen et al., 2024; Alghasab, 2025). The latent growth curve studies also suggest that AI only magnifies existing regulatory capacities rather than creating them, resulting in ceiling effects among high-SRL participants and stagnation among low-SRL participants (Shi et al., 2025). When AI is deployed without differentiated scaffolding, it can create more achievement gaps rather than close them. These observational findings are particularly salient in school education, where adolescents continue to strengthen executive functioning, emotional regulation, and academic autonomy (Bailey et al., 2025). Unlike adult learners, for whom a certain degree of autonomy is often assumed (Ouyang et al., 2022), elementary and secondary students should be explicitly scaffolded to help them internalize AI feedback as learning procedures. To that end, AI design should incorporate concepts from developmental psychology and motivation theory and

ensure that adaptive systems encourage metacognitive interactions rather than passive reliance on automated outputs.

Institutional Readiness and Professional Capacity

In addition to the classroom practice, institutional readiness is a conclusive factor. These reviewed studies have identified gaps in teacher preparedness, difficulties in interpreting AI-generated analytics, and infrastructural limitations (Krange et al., 2023; Lee & Son, 2024; Chu et al., 2024). Digital infrastructure disparities and policy misalignment also make the transition to such technologies even more difficult, particularly in resource-scarce settings (Gupta & Yadav, 2025; Obateru & Daramola, 2025). It is worth noting that there are very few system-level and professional- development-based implementations. Therefore, it seems that AI penetration in elementary and secondary language education is mostly experimental rather than institutional. Adoption Sustainable adoption requires cohesive governance, curriculum integration, and holistic teacher-training models. Without systemic scaffolding, AI could exacerbate current inequalities rather than reduce them. These findings align with larger studies of new educational technologies that highlight the need to align innovation, infrastructure, and professional capacity (Martin et al., 2020).

Ethical Governance and Learner Agency

Ethical issues are a ubiquitous issue, particularly acute in the context of elementary and secondary situations in which learners are underage. Angelic authorship, reduced student agency, and the lack of approved use of tools create apprehension about the nature of the problem of academic integrity and learner agency in generative AI (Woo et al., 2024; Zhu et al., 2024; Obateru & Daramola, 2025; Organ, 2023). On their own, AI systems that gather and process adolescent learning data raise important issues of data privacy and algorithmic transparency, which are the focus of a few studies that discuss them only in limited detail (Alqarni & Alhramelah, 2025).

These strains place AI integration in the realm of a normative problem, rather than simply a technical or pedagogical one. Accountable practices require active governance systems, such as data privacy procedures, transparency in algorithms, and academic integrity frameworks, that are instilled at the outset rather than as an afterthought. Implicitly applying AI ethics to language teaching can also help students develop a more critical approach to AI applications and maintain ownership of their education.

Toward a Human-Centered and Developmentally Responsive AI Framework

The overall systemic perspective on the findings from Research questions 1-3 shows that there is one major principle: it is not inherent in the technology but rather in the systemic context to which it belongs. Only if AI becomes part of pedagogical models led by teachers, is adjusted to suit adolescents' developmental capabilities, is supported by a well-developed institutional layer, and is governed by a well-established code of ethics, will substantial educational advantages be achieved. Other interventions, e.g., writing feedback, adaptive reading, self-regulated learning (SRL)-supportive environments, have significant promise; however, variability in skill coverage, limited geographic expansion, and a lack of longitudinal data limit their applicability.

Advancement in this sphere requires a paradigm shift, reimagining AI not as a self-driven educational machine but as an intelligence-enhancing partner that should operate within teacher-designed, ethically regulated learning settings (Roll & Wylie, 2016; Mustafa et al., 2024). This framework will unite the theories of second-language acquisition, developmental psychology, work practice, and governance policy into a single, consistent entity. Only in this way, through an integrated alignment of tool design, classroom pedagogy, teacher development, and institutional supervision, can AI continue to play an important role in communicative competence, cognitive development, and learner agency in elementary and secondary-school language education.

Conclusion, Limitations, and Future Work

This systematic review synthesized empirical studies to map the technological and pedagogical trends, implementation challenges, and research directions surrounding AI integration in second-level language education. The results indicate that AI-based tools, including automated writing evaluation, automated writing feedback, generative AI, adaptive reading systems, and conversational agents, have a high potential to improve writing performance, vocabulary acquisition, engagement of learners, and even the elements of self-regulated learning (Krange et al., 2023; Jaramillo et al., 2025; Shi et al., 2025; Takii et al., 2024). Nevertheless, their usefulness is developmentally mediated and pedagogically conditional. When integrated into instructional systems led by teachers (Engeness, 2025; Woo et al., 2024), consistent with the metacognitive developmental abilities of adolescents (Gupta & Yadav, 2025), and the organizational capacity and structures of governance (Lee & Son, 2024; Chu et al., 2024), AI brings about sustainable advantages. Simultaneously, the key issue is the focus on cognitive dependency, academic integrity, data privacy, and algorithmic transparency (Organ, 2023; Zhu et al., 2024; Alqarni and ʿAlhramelah, 2025). Taken together, the facts present AI not as an independent

teaching alternative but as an intelligence-enhancing companion within a socio-technical, ethically controlled educational system.

Despite many substantive contributions, the present review is limited by several constraints. First of all, the corpus was narrowed to English-language research, which is included in major databases since 20182025; this method can also exclude relevant regional or non-indexed research. In addition, the small sample size (n = 20) and geographic concentration in East and Southeast Asia limit the extent to which the results can be extrapolated to other regions of the world. Lastly, the use of short-term or quasi-experimental designs in most of the sampled studies limits the generalizability of long-term skill retention and institutional scalability. Future research should be more focused on longitudinal, multi-site studies that do not limit integrated language competencies to single writing activities, widen the range of representation in geographically underrepresented regions, and conduct systematic assessments of teacher professional development programs and policy frameworks related to them. More emphasis should also be placed on explainable AI architectures, equity-based deployment approaches, and the encouragement of ethical AI literacy among adolescent users. Further exploration of such areas will enable the field to shift not only towards preliminary experimental innovation but also towards sustainable, developmentally responsive, and human-centered applications of artificial intelligence in school language education.

REFERENCES

Alghasab, M. B. (2025). English as a foreign language (EFL) secondary school students' use of artificial intelligence (AI) tools for developing writing skills: unveiling practices and perceptions. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2505304

Alqarni, A., & Alhramelah, A. (2025). Designing a smart voice assistant based on artificial intelligence to develop listening and speaking skills in English for elementary school students. *Educational Process: International Journal*, *17*, Article e2025403. https://doi.org/10.22521/edupij.2025.17.403

Andreani, W., & Ying, Y. (2019). PowPow interactive game in supporting English vocabulary learning for elementary students. *Procedia Computer Science*, *157*, 473-478. https://doi.org/10.1016/j.procs.2019.09.005

Bailey, A. L., Abraham, M. E., Dong, J. Y., Torres, I. M., Wang, Y. P., Zamora, E., & Zhang, X. (2025). Translanguaging: Conceptual underpinnings of equity-oriented instructional and assessment practices with adolescent multilingual learners. *Linguistics and Education*, *87*, Article 101418. https://doi.org/10.1016/j.linged.2025.101418

Barrot, J. S. (2023). Using automated written corrective feedback in writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, *36*(4), 584-607. https://doi.org/10.1080/09588221.2021.1936071

Benkhalfallah, F., Laouar, M. R., & Benkhalfallah, M. S. (2024). Empowering education: Harnessing artificial intelligence for adaptive e-learning excellence. In Proceedings of the International Conference on Artificial Intelligence and Its Applications in the Age of Digital Transformation. https://doi.org/10.1007/978-3-031-71429-0_4

Chu, L., Liu, Y., Zhai, Y., Wang, D., & Wu, Y. (2024). The use of deep learning integrating image recognition in language analysis technology in secondary school education. *Scientific Reports*, *14*(1), Article 2888. https://doi.org/10.1038/s41598-024-52592-5

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A., & Soomro, R. B. (2024). Extended TAM-based acceptance of AI-powered ChatGPT for supporting metacognitive self-regulated learning in education: A mixed-methods study. *Heliyon*, *10*(8), Article eXXXXX. https://doi.org/10.1016/j.heliyon.2024.e29317

Dekhakhena, A. (2025). AI-powered personalized learning in EFL acquisition: Exploring adaptive instruction and feedback systems. *Journal of Studies in Language, Culture and Society*, *8*(1), 111-131. https://asjp.cerist.dz/en/article/266936

Engeness, I. (2025). Cultural-historical perspective to design pedagogical AI for enhancing student writing. *Technology, Knowledge and Learning*, 1-22. https://doi.org/10.1007/s10758-025-09876-0

Gull, A., Imran, M., Yiunas, M., & Afzaal, M. (2020). A descriptive study of challenges faced by English language teachers in integrating information and communication technology (ICT) tools at elementary level in Pakistan. *International Journal of Advanced Science and Technology*, *29*(8), 290-305.

Gupta, M., & Yadav, A. K. (2025). Adoption of artificial intelligence technologies for enhancing English language skills in government senior secondary schools: Opportunities, challenges, and implementation strategies. *International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR)*, *8*(4). https://www.ijamsr.com/issues/6_Volume%208_Issue%204/20250811_093922_8223.pdf

Gutiérrez, L. (2023). Artificial intelligence in language education: Navigating the potential and challenges of chatbots and NLP. *Research Studies in English Language Teaching and Learning*, *1*(3), 180-191. https://doi.org/10.62583/rseltl.v1i3.44

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., & O'Cathain, A. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285-291. https://doi.org/10.3233/EFI-180221

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Imran, M., Qureshi, S. H., Qureshi, A. H., & Almusharraf, N. (2024a). Classification of English Words into Grammatical Notations Using Deep Learning Technique. *Information*, *15*(12), 801. https://doi.org/10.3390/info15120801

Imran, M., Almusharraf, N., & Dalbani, H. (2024b). Future of writing in higher education: A review of prospective use of artificial intelligence and ChatGPT in academia. In Innovative pedagogical practices for higher education 4.0 (1st ed., pp. 108-124). Routledge. https://doi.org/10.1201/9781003400691

Jaramillo, J. J., Chiappe, A., & Delgado, F. S. (2025). From struggle to mastery: AI-powered writing skills in ESL education. *Applied Sciences*, *15*(14). https://doi.org/10.3390/app15148079

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, *55*(1), 48-63. https://doi.org/10.1080/15391523.2022.2142873

Junaidi, J. (2020). Artificial intelligence in EFL context: Rising students' speaking performance with Lyra virtual assistance. *International Journal of Advanced Science and Technology*, *29*(5), 6735-6741. https://repository.unilak.ac.id/1512/

Kessler, G. (2024). Using technology to improve language learners' accuracy. In Practical grammar teaching for the second language classroom (pp. 242-267). Routledge. https://doi.org/10.4324/9781032650883-13

Krange, I., Segaran, M., Gamlem, S., Moltudal, S., & Engeness, I. (2023). A triple challenge: Students' identification, interpretation, and use of individualized automated feedback in learning to write English as a foreign language. *Interaction Design & Architecture(s)*, *56*, 37-61. https://doi.org/10.55612/s-5002-059-001

Lee, J., & Son, H. (2024). Enhancing EFL adolescents' writing skills using AI. *Multimedia-Assisted Language Learning*, *27*(4). http://dx.doi.org/10.15702/mall.2024.27.4.200

Lei, H. (2022). High school students' foreign language vocabulary acquisition in the era of artificial intelligence. In Proceedings of the 2021 International Conference on Education, Language and Art (ICELA 2021). https://doi.org/10.2991/assehr.k.220131.121

Liang, J.-C., Hwang, G.-J., Chen, M.-R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. Interactive Learning Environments, 31(7), 4270-4296. https://doi.org/10.1080/10494820.2021.1958348

Malakul, S., & Park, I. (2023). The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students. Smart Learning Environments, 10(1). https://doi.org/10.1186/s40561-023-00224-2

Martin, F., Dennen, V. P., & Bonk, C. J. (2020). A synthesis of systematic review research on emerging learning environments and technologies. *Educational Technology Research and Development*, *68*(4), 1613-1633. https://doi.org/10.1007/s11423-020-09812-2

Mustafa, M. Y., Tlili, A., Lampropoulos, G., Huang, R., Jandrić, P., Zhao, J., Salha, S., Xu, L., Panda, S., & Kinshuk. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): A roadmap to a future research agenda. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00350-5

Obateru, O. T., & Daramola, C. A. (2025). Effectiveness of an artificial intelligence literacy programme (AILP) in promoting responsible use of AI tools in English language learning among secondary school students in Oye-Ekiti. *UNIZIK Journal of Educational Research and Policy Studies*, *19*(4). https://unijerps.org/index.php/unijerps/article/view/982

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., & Moher, D. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biological Reviews*, *96*(5), 1695-1722. https://doi.org/10.1111/brv.12721

Organ, A. (2023). Attitudes to the use of Google Translate for L2 production: Analysis of chatroom discussions among UK secondary school students. *The Language Learning Journal*, *51*(3), 328-343. https://doi.org/10.1080/09571736.2021.2023896

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893-7925. https://doi.org/10.1007/s10639-022-10925-9

Pennington, M. C., & Rogerson-Revell, P. (2018). Using technology for pronunciation teaching, learning, and assessment. In English pronunciation teaching and research: Contemporary perspectives (pp. 235-286). Springer. https://doi.org/10.1057/978-1-137-47677-7_5

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582-599. https://doi.org/10.1007/s40593-016-0110-3

Sharmin, L., Kalima, R., Imran, M., Chauca, M., & Balladares, J. (2026). Teachers' perceptions of AI in supporting students' learning within globally diverse digital settings. *Discover Education*, *5*, 75. https://doi.org/10.1007/s44217-025-01089-y

Shi, L., Li, S., & Xing, J. (2025). Exploring Chinese secondary EFL students' self-regulated learning and task engagement in AI-assisted classrooms: A latent growth curve modelling study. *European Journal of Education*, *60*(4). https://doi.org/10.1111/ejed.70241

Takii, K., Flanagan, B., Li, H., Yang, Y., Koike, K., & Ogata, H. (2024). Explainable eBook recommendation for extensive reading in K-12 EFL learning. Research and Practice in Technology Enhanced Learning, 20. https://doi.org/10.58459/rptel.2025.20027

Trân, N. H. B. (2025). The impact of task-based language learning on vocabulary development for Vietnamese EFL learners [master's thesis, DALAT UNIVERSITY]. https://share.google/E64E3BkKXAJjpjmTY

Waqar, A., Khan, M. B., Shafiq, N., Skrzypkowski, K., Zagórski, K., & Zagórska, A. (2023). Assessment of challenges to the adoption of IoT for the safety management of small construction projects in Malaysia. *Applied Sciences*, *13*(5), https://doi.org/10.3390/app13053340

Wen, F., Li, Y., Zhou, Y., An, X., & Zou, Q. (2024). A study on the relationship between AI anxiety and AI behavioral intention of secondary school students learning English as a foreign language. *Journal of Educational Technology Development and Exchange*, *17*(1), 130-154. https://doi.org/10.18785/jetde.1701.07

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. Ampersand, 14. https://doi.org/10.1016/j.amper.2025.100224

Wilson, J., Palermo, C., & Wibowo, A. (2024). Elementary English learners' engagement with automated feedback. Learning and Instruction, 91. https://doi.org/10.1016/j.learninstruc.2024.101890

Woo, D. J., Guo, K., & Salas-Pilco, S. Z. (2024). Writing creative stories with AI: Learning designs for secondary school students. *Innovation in Language Learning and Teaching*, 1-13.

Zhai, Y. (2025). Anxiety Reduction Training and EFL Learners' Foreign Language Anxiety and Negative Emotions. *Perceptual and Motor Skills*, *133*(2), 355-380. https://doi.org/10.1177/00315125251357458

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, *127*, Article 103501.

Zhang, Q., Siraj, S. B., & Abdul Razak, R. B. (2025). Effects of AI chatbots on EFL students' critical thinking skills and intrinsic motivation in argumentative writing. *Innovation in Language Learning and Teaching*, 1-29. https://doi.org/10.1080/17501229.2025.2515111

Zhu, T., Zhang, K., & Wang, W. Y. (2024). Embracing AI in education: Understanding the surge in large language model use by secondary students. arXiv Preprint, arXiv:2411.18708. https://doi.org/10.48550/arXiv.2411.18708

MUHAMMAD IMRAN, PhD from the School of Foreign Languages, Shanghai Jiao Tong University, China, is a researcher in the Educational Research Lab at the College of Humanities and Sciences at Prince Sultan University and an adjunct faculty in the Department of English Language and Literature at Khazar University, Azerbaijan. His research interests include digital literacy, language education technologies, AI and English language teaching and learning, and machine translation. Email: mimran@psu.edu.sa (corresponding author)

NORAH ALMUSHARRAF, PhD from Buffalo University, USA, is an Associate Professor and Leader of the Educational Research Lab at Prince Sultan University in Saudi Arabia. She is among the leading researchers in Language Education in the Middle East. Her research interests include foreign language teaching instructional design, educational technology, and change management. Email: nmusharraf@psu.edu.sa

KHURRAM SHEHZAD holds a master's degree in computer science from Northeastern University in Shenyang, China, and a master's degree in Education from Pakistan. He is currently serving as a Secondary School Teacher at FG Public School Boys No. 1, Okara Cantt., Pakistan. His research interests include educational technology, artificial intelligence in education, computer science education, and innovative teaching and learning practices. Email: Khurram.6470@gmail.com

---

Acknowledgment

The authors would like to thank Prince Sultan University for its technical and financial support.

Artificial Intelligence Use

The authors did not use any AI tools to write this draft. However, the authors used Grammarly Premium only for language proofreading.

Ethical Statement

This article does not contain any studies with human participants performed by any of the authors.