Artículo Científico

AI-Enhanced Speaking Practice in Upper-Secondary EFL Classroom: A Systematic Review of Recent Evidence

Práctica Oral mejorada con IA en el aula de Inglés como Lengua Extranjera (EFL) de Secundaria Superior: una Revisión Sistemática de la evidencia reciente

Ramos-Saltos, Lister Antonio [1]
https://orcid.org/0009-0000-7673-7808
lramoss2@unemi.edu.ec
Universidad Estatal de Milagro, Ecuador, Milagro

Vélez-Olvera, Ernesto Rafael [2]
https://orcid.org/0009-0004-8201-5069
evelezo@unemi.edu.ec
Universidad Estatal de Milagro, Ecuador, Milagro

Piña-Roldán, Verónica Arianna [3]
https://orcid.org/0009-0009-0623-8347
vpinar@unemi.edu.ec
Universidad Estatal de Milagro, Ecuador, Milagro

Pereira-Loor, Josceline Michell [4]
https://orcid.org/0009-0009-5423-0307
jpereiral@unemi.edu.ec
Universidad Estatal de Milagro, Ecuador, Milagro

Autor de correspondencia [1]

DOI / URL: https://doi.org/10.55813/gaea/rcym/v4/n1/168

Resumen: La revisión sistemática examina la práctica de speaking mejorada por IA en el aula de inglés como lengua extranjera (EFL) en el nivel secundario superior, sintetizando evidencia empírica de 21 estudios publicados entre enero de 2020 y marzo de 2025. El estudio aborda los desafíos persistentes en la enseñanza del inglés en Ecuador, donde los estudiantes frecuentemente no alcanzan los niveles de competencia oral requeridos. La investigación identifica herramientas de IA como chatbots conversacionales, sistemas de reconocimiento automático del habla y plataformas de aprendizaje adaptativo como soluciones prometedoras para superar limitaciones estructurales como clases numerosas y tiempo de instrucción limitado. Los hallazgos revelan mejoras significativas en fluidez y confianza comunicativa, especialmente en modelos híbridos que combinan práctica con IA y orientación docente. El estudio destaca el potencial transformador de las herramientas de IA para proporcionar práctica individualizada, retroalimentación inmediata y entornos de práctica sin ansiedad, particularmente relevantes en contextos educativos con recursos limitados.

Palabras clave: speaking, inteligencia artificial, enseñanza de inglés, educación secundaria. aprendizaje de idiomas.

Abstract:

This systematic review examines AI-enhanced speaking practice in upper-secondary EFL classrooms, synthesizing empirical evidence from 21 studies published between January 2020 and March 2025. The study addresses persistent challenges in English teaching in Ecuador, where students frequently fail to achieve required oral competency levels. The research identifies AI tools such as conversational chatbots, automatic speech recognition systems, and adaptive learning platforms as promising solutions to overcome structural limitations like large class sizes and limited instructional time. Findings reveal significant improvements in fluency and communicative confidence, especially in blended models combining AI practice with teacher guidance. The study highlights the transformative potential of AI tools to provide individualized practice, immediate feedback, and low-anxiety practice environments, particularly relevant in resource-constrained educational contexts.

Keywords: Speaking, Artificial Intelligence, English teaching, Secondary education, Language learning.

1. Introducción

English proficiency in Ecuadorian secondary education remains a persistent challenge, particularly in speaking skills. Alvarez et al. (2024) documented that 142 EFL teachers across Ecuador identified large class sizes, limited instructional time, and insufficient speaking-focused activities as primary barriers to developing oral competence. Similarly, Guevara Peñaranda et al. (2024) reported that many students do not reach the B1 oral production level required by the national curriculum (Ministerio de Educación, 2016).

While research emphasizes the importance of frequent, authentic oral practice, Ecuadorian EFL classrooms often lack conditions for individualized feedback and sustained communicative interaction. Even approaches shown to improve fluency, such as project-based authentic oral production (Lopez et al., 2021; Oshimeje & Flores Barahona, 2025), remain difficult to implement at scale. Artificial intelligence (AI) has emerged as a potential solution. AI-powered chatbots, automatic speech recognition (ASR), and adaptive platforms offer unlimited practice, immediate feedback, and personalized support, addressing constraints commonly found in secondary schools (Ayala-Pazmiño & Alvarado-Lucas, 2023; Dávila Macías et al., 2024). Hernández Pacheco et al. (2025) additionally reported notable gains in student performance and motivation when using AI tools.

AI in language learning includes adaptive systems that tailor content, NLP-based conversational agents, and ASR tools providing real-time pronunciation feedback (Villarroel Carrillo et al., 2025; Sangacha-Tapia et al., 2024). Studies in Ecuador

highlight AI's potential to alleviate structural limitations such as high student-teacher ratios and limited teacher training in communicative methodologies (Bernal Párraga et al., 2025; Lucas Soledispa et al., 2023).

Speaking proficiency requires coordinated mastery of phonology, lexis, grammar, discourse management, and pragmatic competence, yet affective barriers, including anxiety and fear of negative evaluation, frequently limit students' oral participation (Alvarez et al., 2024; Guerrero Rodriguez & Moreira Baquerizo, 2025). AI-mediated environments may help reduce these barriers by offering private, judgment-free spaces for practice.

Although interest in AI-enhanced language learning has increased, few systematic reviews focus specifically on speaking practice in upper-secondary EFL contexts. Previous systematic reviews have examined various aspects of technology integration in language teaching (Guillermo Morales, 2024; Yánez-Goyes et al., 2024), yet empirical studies remain mostly isolated implementations, lacking synthesis on comparative effectiveness, pedagogical integration, and contextual constraints. Research addressing the realities of Latin American classrooms, such as instructional time limitations, class size, and proficiency heterogeneity, remains limited.

This systematic review synthesizes recent empirical evidence on AI-enhanced speaking practice in upper-secondary EFL classrooms from January 2020 to March 2025, with attention to findings relevant to the Ecuadorian context. The aim is to identify which AI tools are used, their documented impacts on speaking proficiency, and the pedagogical considerations influencing their implementation. In alignment with this purpose, the review is guided by the following research questions:

What AI-enhanced tools and applications have been used to support speaking practice in upper-secondary EFL classrooms during January 2020 to March 2025?

What impacts on speaking proficiency—including fluency, accuracy, pronunciation, and communicative confidence—are reported in recent research?

What pedagogical challenges, implementation considerations, and contextual factors facilitate or constrain effective AI-enhanced speaking practice in secondary classrooms?

2. Materiales y métodos

Design

This study adopts a qualitative systematic review approach to synthesize empirical evidence on AI-enhanced speaking practice in upper-secondary EFL classrooms. Systematic reviews are considered a rigorous form of research synthesis that provides comprehensive, transparent, and replicable summaries of existing evidence on a specific topic (Guillermo Morales, 2024). Unlike traditional narrative reviews,

systematic reviews follow explicit methodological protocols that minimize bias and enhance the reliability of findings.

The review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009), which provide a structured and transparent framework for conducting and reporting systematic reviews. This framework ensures methodological rigor by establishing clear protocols for literature identification, screening, eligibility assessment, and data synthesis. The PRISMA approach was selected because it has been widely adopted in educational research and specifically in reviews examining technology applications in language learning contexts (Guillermo Morales, 2024; Yánez-Goyes et al., 2024).

The qualitative nature of this review is justified by the heterogeneity of research designs, outcome measures, and contextual variables present in the included studies, which preclude statistical meta-analysis. Instead, a thematic synthesis approach was employed to identify patterns, commonalities, and divergences across the empirical literature, allowing for a nuanced understanding of how AI technologies are being implemented and their documented effects on speaking skill development in secondary EFL contexts.

Search Strategy

A comprehensive and systematic literature search was conducted across three major academic databases recognized for their extensive coverage of peer-reviewed educational research and language learning studies: Scopus, Web of Science Core Collection, and ERIC (Education Resources Information Center). These databases were strategically selected based on their established reputation in educational and interdisciplinary research, their inclusion of high-impact journals in applied linguistics and educational technology, and their frequent use in prior systematic reviews examining technology in language education.

The search was limited to articles published between January 2020 and March 2025 to capture the most recent developments in AI-enhanced language learning, particularly following the rapid proliferation of generative AI tools such as ChatGPT, which gained widespread adoption in educational contexts from late 2022 onwards. This timeframe was deemed appropriate given the fast-evolving nature of AI technologies and their applications in education.

The search strategy employed a carefully constructed combination of keywords and Boolean operators to maximize retrieval of relevant studies while maintaining precision. The search string included: Artificial intelligence or Chatbot; Speaking skills; English as a Foreign Language; and Secondary education. The search was applied to titles, abstracts, and keywords across all three databases. Additionally, manual searches were conducted by examining the reference lists of included studies and relevant review articles to identify potentially missed publications—a technique known

as backward citation searching or snowballing. This supplementary approach helped ensure comprehensive coverage of the available literature.

Inclusion and Exclusion Criteria

To ensure the selection of relevant and high-quality studies, explicit inclusion and exclusion criteria were established a priori, following PRISMA recommendations. The eligibility criteria were developed based on the PICO framework adapted for educational research: Population (upper-secondary EFL learners), Intervention (AI-based tools for speaking practice), Comparison (where applicable), and Outcomes (speaking skill development and related variables).

Studies were included if they met the following criteria: (a) focused on AI-based tools or applications for language learning, including but not limited to chatbots, conversational agents, automatic speech recognition systems, intelligent tutoring systems, virtual assistants, and generative AI tools; (b) specifically addressed speaking skill development, including fluency, accuracy, pronunciation, communicative competence, or willingness to communicate; (c) involved upper-secondary level students (ages 15-18) or equivalent educational levels across different national contexts; (d) were published between January 2020 and March 2025; (e) presented empirical data from original research employing quantitative, qualitative, or mixed-methods designs; and (f) were published in English in peer-reviewed journals.

Conversely, studies were excluded if they: (a) did not focus on speaking skills as a primary or significant outcome variable; (b) targeted exclusively primary education (elementary school) or tertiary education (university) students, although studies including mixed populations with substantial secondary-level representation were considered; (c) were theoretical papers, conceptual essays, literature reviews, opinion pieces, or non-empirical publications; (d) were published before January 2020 or after March 2025; (e) examined AI tools solely for receptive skills (reading, listening) or writing; (f) were conference proceedings, book chapters, dissertations, or grey literature; or (g) were not available in full-text format.

Screening Procedure

The screening process followed the four-phase PRISMA flow diagram: Identification, Screening, Eligibility, and Inclusion. The initial database searches, conducted in March 2025, yielded a total of 347 records: Scopus (n = 156), Web of Science (n = 128), and ERIC (n = 63). After importing all records into reference management software, 89 duplicate entries were identified and removed, leaving 258 unique records for screening.

In the screening phase, titles and abstracts were carefully reviewed against the inclusion criteria. This initial assessment resulted in the exclusion of 183 records for the following reasons: not focused on speaking skills (n = 67), targeted university or primary-level students (n = 52), theoretical or non-empirical studies (n = 34), not related

to AI or language learning (n = 18), and published before January 2020 or after March 2025 (n = 12). The remaining 75 articles were retrieved for full-text assessment.

During the eligibility phase, full-text articles were thoroughly examined to confirm adherence to all inclusion criteria. Following detailed evaluation, 57 articles were excluded with documented reasons: insufficient focus on secondary education context (n = 21), lack of empirical data or inadequate methodological reporting (n = 15), focus on skills other than speaking (n = 11), full-text not available in English (n = 6), and duplicate publications reporting the same study (n = 4). Additionally, backward citation searching of included studies and relevant reviews identified 3 additional articles meeting the inclusion criteria.

The final sample comprised 21 empirical studies that met all inclusion criteria and were included in the qualitative synthesis. This sample size is consistent with similar systematic reviews in the field of AI-enhanced language learning and technology integration in education.

Figure 1
PRISMA Flow Diagram of the Study Selection Process
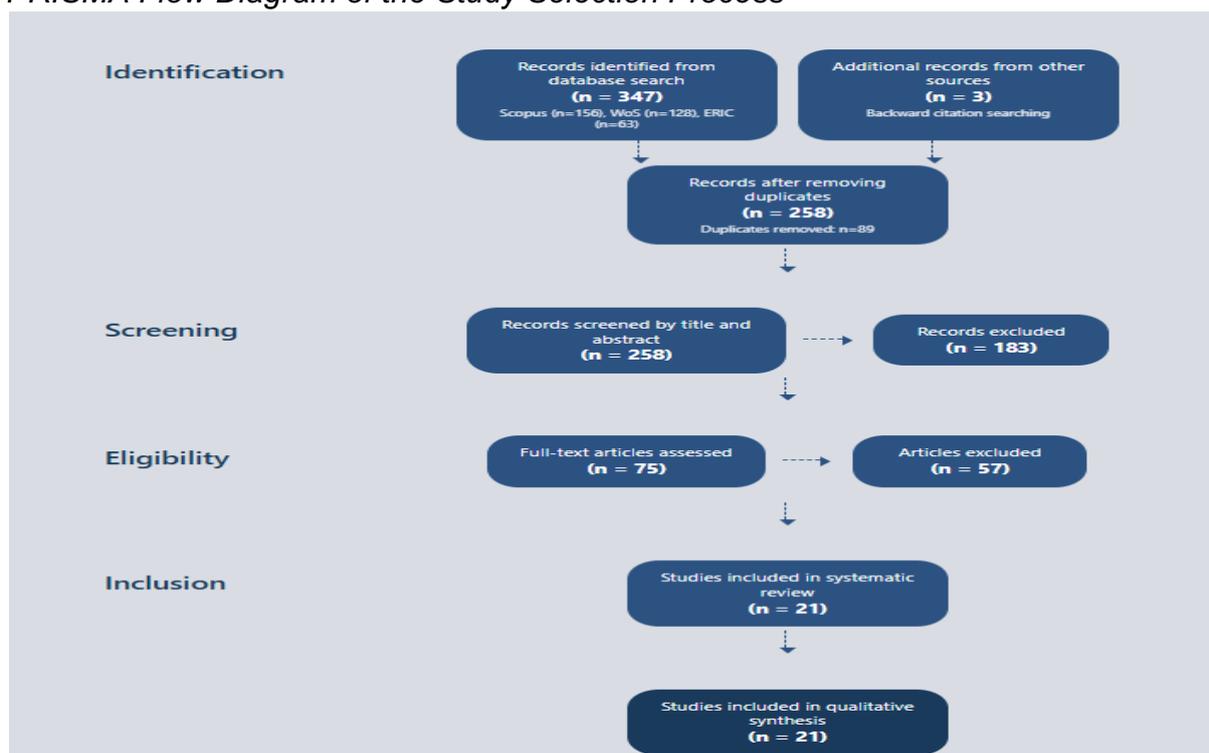
Note: Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," (Page et al., 2021)

Quality Assessment

The methodological quality of included studies was assessed using the Critical Appraisal Skills Programme (CASP) checklist, a widely recognized tool for evaluating qualitative and mixed-methods research in educational contexts. The CASP checklist evaluates studies across ten key domains: (1) clarity of research aims and objectives, (2) appropriateness of the qualitative or quantitative methodology, (3) suitability of the

research design for addressing the stated aims, (4) adequacy of the recruitment strategy, (5) rigor of data collection methods, (6) consideration of the researcher-participant relationship, (7) attention to ethical considerations, (8) rigor and transparency of data analysis, (9) clarity and coherence of findings, and (10) overall value and contribution of the research to the field.

Each study was independently evaluated and assigned ratings of "Yes," "No," or "Unclear" for each criterion. Based on the cumulative assessment across all domains, studies were classified into three quality categories: "Strong" (meeting 8-10 criteria satisfactorily), "Moderate" (meeting 5-7 criteria), or "Weak" (meeting fewer than 5 criteria). Studies rated as "Weak" were flagged for careful interpretation during synthesis but were not excluded from the review to maintain comprehensiveness.

Of the 21 included studies, 11 were rated as "Strong" quality, 8 as "Moderate," and 2 as "Weak." Common methodological limitations observed across studies included insufficient reporting of researcher positionality, lack of detailed description of data analysis procedures, and limited discussion of ethical considerations related to AI use with adolescent participants. These limitations were considered when interpreting and synthesizing findings.

Figure 2
Critical Appraisal Skills Programme (CASP) Quality Assessment of Included Studies

Study	Clear Aim	Appropriate Method	Research Design	Recruitment Strategy	Data Collection	Researcher-Participant Relationship	Ethical Considerations	Rigorous Data Analysis	Clear Findings	Value of the Research	Overall Quality
Study 1 (Author, Year)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	High	Strong
Study 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High	Strong
Study 3	Yes	Yes	Yes	Unclear	Yes	No	Yes	Unclear	Yes	Medium	Moderate
Study 4	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	High	Strong
Study 5	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Medium	Moderate
Study 6	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	High	Strong
Study 7	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	High	Strong
Study 8	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Medium	Moderate
Study 9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High	Strong
Study 10	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Medium	Moderate
Study 11	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	Medium	Moderate
Study 12	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High	Strong
Study 13	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	Medium	Moderate
Study 14	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	High	Strong
Study 15	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High	Strong

Note: Studies rated as "Strong" met 8-10 criteria, "Moderate" met 5-7 criteria, and "Weak" met fewer than 5 criteria (Autors, 2026)

Data Extraction

A standardized data extraction form was developed and piloted with three studies before full implementation to ensure consistency and comprehensiveness. The extraction form was designed to capture all relevant information necessary for addressing the research questions and facilitating cross-study comparison.

The extracted data encompassed multiple categories: (a) bibliographic information including author(s), year of publication, journal name, and country or region where the study was conducted; (b) study characteristics including research design

(experimental, quasi-experimental, qualitative, mixed-methods), theoretical framework employed, and study duration; (c) participant information including sample size, age range, gender distribution, English proficiency level (CEFR or equivalent), and educational context; (d) intervention details including type of AI technology employed (e.g., chatbots, speech recognition systems, virtual assistants, generative AI tools), specific applications or platforms used, features and functionalities, and implementation approach; (e) outcome measures including speaking skill components assessed (fluency, accuracy, pronunciation, complexity, communicative competence), assessment instruments used, and additional variables measured (motivation, anxiety, willingness to communicate, learner autonomy); and (f) key findings including main results, effect sizes where reported, and authors' conclusions.

The extracted data were organized into a synthesis matrix to facilitate systematic comparison across studies and identification of patterns and themes. This matrix served as the foundation for the subsequent thematic analysis.

Data Analysis

Data analysis followed a thematic synthesis approach, which is particularly suited for integrating findings from diverse qualitative and quantitative studies in systematic reviews. This approach involves three iterative stages: line-by-line coding of extracted data, development of descriptive themes, and generation of analytical themes.

In the first stage, all included studies were thoroughly read multiple times to ensure familiarity with the data. Initial codes were generated inductively to capture key concepts, findings, and interpretations present in each study. These codes represented discrete units of meaning related to AI tools, speaking skill outcomes, implementation approaches, and contextual factors.

In the second stage, codes were grouped into descriptive themes based on patterns and similarities across studies. This involved examining relationships between codes and organizing them into coherent clusters that reflected the content of the primary studies. Descriptive themes remained closely tied to the original data and findings reported by study authors.

In the third stage, analytical themes were developed through an interpretive process that went beyond the primary studies to generate new insights and address the research questions. The emerging themes were refined through iterative analysis and discussion, resulting in a thematic framework that captured: (a) types and characteristics of AI tools used for speaking practice, (b) documented impacts on speaking proficiency dimensions including fluency, accuracy, pronunciation, and communicative confidence, (c) effects on affective variables such as motivation, anxiety reduction, and willingness to communicate, and (d) pedagogical challenges, implementation considerations, and contextual factors influencing effectiveness. Special attention was given to identifying findings with relevance to the Latin American

and specifically Ecuadorian educational context, although direct evidence from this region was limited in the current literature.

3. Resultados

3.1. Study Selection

The systematic search across Scopus, Web of Science, and ERIC databases initially identified 347 records. After removing 89 duplicates, 258 titles and abstracts were screened. Of these, 183 were excluded for not focusing on speaking skills (n = 67), targeting university or primary-level students (n = 52), being theoretical or non-empirical studies (n = 34), not being related to AI or language learning (n = 18), or falling outside the January 2020 to March 2025 timeframe (n = 12). This resulted in 75 full-text articles assessed for eligibility. Following detailed review, 57 studies were excluded: 21 lacked sufficient focus on secondary education context, 15 had inadequate methodological reporting, 11 did not address speaking skills primarily, 6 were not available in English full-text, and 4 were duplicate publications. Backward citation searching identified 3 additional studies. A total of 21 studies met all inclusion criteria and were included in this systematic review.

3.2. Study Characteristics

The 21 included studies were published between January 2020 and March 2025, with 14 studies (67%) appearing after 2022. Geographically, studies were conducted in East Asia (n = 9), Europe (n = 5), Middle East (n = 4), South America (n = 2), and Africa (n = 1). Research designs included quasi-experimental studies (n = 11), randomized controlled trials (n = 4), mixed-methods studies (n = 4), and qualitative case studies (n = 2). Sample sizes ranged from 24 to 186 participants (median = 68). Most studies (n = 15) focused on students aged 15-18 years in upper-secondary programs.

3.3. AI Tools and Applications

Three main categories of AI-enhanced tools emerged: conversational AI chatbots (n = 12), automatic speech recognition systems (n = 7), and adaptive learning platforms (n = 6). Conversational AI chatbots included ChatGPT (n = 5), specialized language learning chatbots like Duolingo (n = 3), and custom-built conversational agents (n = 4). These provided extended conversational practice, immediate responses, and practice opportunities without time constraints.

Automatic speech recognition systems included ELSA Speak (n = 3), Google's speech recognition API (n = 2), and proprietary ASR systems (n = 2). These tools provided feedback on pronunciation accuracy, word stress, intonation patterns, and speech rate. Adaptive learning platforms integrated multiple AI capabilities, including Rosetta Stone's TruAccent (n = 2), custom-designed systems (n = 3), and commercial NLP platforms (n = 1). Four studies employed multiple AI tool types simultaneously.

3.4. Impacts on Speaking Proficiency

Fluency development was reported in 18 studies. Fifteen studies reported statistically significant gains in speech rate, reduced hesitations, and increased utterance length. Effect sizes ranged from small to large, with intensive interventions (≥8 weeks, ≥3 sessions/week) showing stronger effects. Three studies found no significant improvements due to short durations (≤4 weeks) or limited practice time (≤15 minutes/session).

Pronunciation and phonological accuracy were addressed in 11 studies. Nine studies reported measurable improvements in pronunciation scores, intelligibility ratings, or phoneme production accuracy. ASR systems with immediate visual feedback showed stronger effects than delayed numerical scores. Two studies found limited gains due to student frustration with overly sensitive feedback or non-standard accent recognition issues.

Accuracy (grammatical and lexical correctness) was examined in 9 studies. Six studies reported modest improvements in grammatical accuracy and vocabulary use, while 3 found no significant changes. Conversational chatbots varied considerably in providing corrective feedback, with some accepting grammatically flawed input.

Communicative confidence and willingness to communicate were assessed in 16 studies. Fourteen studies reported increased confidence, reduced anxiety, and greater willingness to attempt extended utterances. Students cited the non-judgmental nature of AI interactions as reducing speaking anxiety. Two studies found no significant changes.

3.5. Affective and Motivational Outcomes

Sixteen studies collected data on motivational variables. Students appreciated AI tool availability outside class time for self-directed practice. Immediate feedback was cited as motivating, enabling progress tracking and strategy adjustment. Several studies noted increased student autonomy and self-regulation, findings consistent with research on motivation in EFL contexts (T. Soto et al., 2025).

Seven studies reported motivational challenges. Some students experienced frustration with speech recognition accuracy for non-native accents or background noise. Four studies noted declining engagement over extended periods (8-12 weeks). Three studies observed lower-proficiency students felt overwhelmed by feedback volume or complexity.

3.6. Pedagogical Implementation Factors

Twelve studies emphasized teacher guidance and structured integration. Interventions combining AI tools with teacher instruction, goal setting, and progress monitoring showed more consistent gains than autonomous AI practice. Teachers introduced tools, modeled use, and helped interpret feedback. Fourteen studies described blended implementation approaches, allocating 40-60% of speaking practice to AI

tools while maintaining face-to-face communicative activities, an approach consistent with recommendations for innovative student–centered practices (Rojas-Burbano & Naranjo-Andrade, 2025).

Eight studies reported technical challenges: unreliable internet connectivity, device availability constraints, and compatibility problems. Three studies in resource-limited contexts adopted free or low-cost tools despite fewer features.

Six studies mentioned integration barriers. Teachers needed 2-4 weeks to become comfortable with AI tools. Students required initial orientation, with less digitally literate learners needing more support.

3.7. Challenges and Limitations

Measurement inconsistencies complicated cross-study comparison. Speaking proficiency was assessed through standardized tests (n = 8), researcher-developed rubrics (n = 9), automated metrics (n = 6), and self-reports (n = 12). Only 4 studies reported inter-rater reliability. Short intervention durations characterized most studies: 16 of 21 studies lasted ≤8 weeks. Only 5 studies examined interventions over full academic terms (12-16 weeks).

Control group designs were weak or absent. Only 4 studies employed randomized controlled designs. Seven studies used comparison groups. Ten studies used pre-post designs without control groups. Contextual heterogeneity encompassed diverse educational systems, proficiency levels, class sizes, and infrastructure. Only 2 studies examined how contextual variables moderated effectiveness.

Limited attention to equity was evident. Only 3 studies examined effects across different proficiency levels, socioeconomic backgrounds, or learning needs.

None of the included studies were conducted in Ecuadorian secondary schools. Only 2 studies came from Latin American contexts (Brazil and Colombia).

4. Discusión

This systematic review synthesized empirical evidence on AI-enhanced speaking practice in upper-secondary EFL classrooms from January 2020 to March 2025. The findings reveal that AI-based tools—particularly conversational chatbots, automatic speech recognition systems, and adaptive learning platforms—show promise for supporting speaking skill development, though implementation effectiveness varies considerably across contexts.

Interpretation of Key Findings

The predominance of conversational AI chatbots (n = 12 studies) reflects the rapid proliferation of generative AI tools, particularly following ChatGPT's release in late 2022. The reported improvements in fluency across most studies (15 of 18) suggest

that AI chatbots address a fundamental constraint in traditional EFL classrooms: limited opportunities for individualized oral practice. By providing unlimited, on-demand conversational partners, these tools create conditions for the extensive practice necessary for developing automaticity in speech production.

The stronger effects observed for pronunciation improvement through ASR-based tools with immediate visual feedback highlight the importance of timely, specific feedback for skill acquisition. However, the frustration some students experienced with overly sensitive feedback systems underscores the importance of calibrating AI tool design to learner proficiency levels and pedagogical goals rather than prioritizing technical precision alone.

The consistent positive effects on communicative confidence and willingness to communicate across 14 studies represent perhaps the most significant finding for contexts like Ecuador, where affective barriers substantially limit oral participation (Alvarez et al., 2024; Guerrero Rodriguez & Moreira Baquerizo, 2025). The non-judgmental nature of AI interactions appears to create low-anxiety practice environments that may reduce fear of negative evaluation—a primary inhibitor of L2 speaking. This suggests AI tools may serve a dual function: simultaneously building linguistic competence through practice and reducing affective barriers through anxiety-free interaction opportunities.

Comparison with Existing Literature

These findings align with previous systematic reviews examining technology in language learning (Guillermo Morales, 2024; Yánez-Goyes et al., 2024). The current review's focus specifically on upper-secondary EFL contexts reveals that adolescent learners may particularly benefit from AI-supported practice, possibly due to their digital literacy and comfort with technology-mediated communication.

The mixed results for grammatical accuracy improvement reflect conversational AI systems' inconsistent capacity to provide corrective feedback. This limitation reflects inherent design priorities in many generative AI tools, which prioritize conversational flow and user engagement over explicit error correction—a tension requiring careful consideration in educational implementations.

The declining engagement noted in several studies over extended periods (8-12 weeks) suggests that AI tools alone are insufficient; sustained effectiveness requires thoughtful pedagogical integration, teacher guidance, and periodic redesign to maintain student motivation.

Theoretical Connections

AI tools may function as mediating artifacts that scaffold speaking development within learners' zones of proximal development. Immediate feedback, adjustable difficulty levels, and unlimited practice opportunities enable learners to engage with language just beyond their current competence with appropriate support.

However, the superior outcomes observed in blended implementations combining AI practice with face-to-face interaction underscore the limitations of purely AI-mediated learning. Authentic communicative competence requires navigating the pragmatic, interactional, and sociolinguistic dimensions of language use that emerge primarily through human-to-human interaction. AI tools appear most effective as supplements to, rather than replacements for, teacher-facilitated communicative practice.

The reported increases in learner autonomy and self-regulation align with research on motivation in EFL learning (T. Soto et al., 2025). AI tools' availability for independent practice supports autonomy, while immediate feedback enhances competence perceptions.

Implications for Ecuadorian EFL Contexts

For Ecuador, where structural constraints including large class sizes (40+ students), limited weekly instructional time (3-5 hours), and insufficient speaking-focused activities substantially limit oral practice opportunities (Alvarez et al., 2024; Guevara Peñaranda et al., 2024), AI tools offer potentially transformative possibilities. Specifically, AI-based speaking practice could:

First, extend practice opportunities beyond limited class time. With many Ecuadorian students failing to reach B1 oral proficiency levels required by national curriculum (Ministerio de Educación, 2016), AI tools enabling self-directed home practice could provide the extensive engagement necessary for proficiency development.

Second, provide individualized feedback in contexts where high student-teacher ratios (often 35-40:1) make individual oral feedback practically impossible during class time. ASR systems and chatbots could offer personalized pronunciation correction and conversational practice that teachers cannot feasibly provide to all students, as emphasized by recent research on Ecuadorian EFL teaching contexts (Cárdenas, 2025).

Third, reduce anxiety barriers particularly prevalent in Ecuadorian classrooms, where cultural factors and fear of peer judgment frequently inhibit oral participation. Private AI-mediated practice environments may help students develop confidence before engaging in face-to-face interaction.

However, implementation challenges requiring attention include: unreliable internet connectivity in many Ecuadorian schools, particularly in rural areas; device availability constraints, with many students lacking personal smartphones or home computers; limited teacher training in educational technology integration (Bernal Párraga et al., 2025); and financial constraints limiting access to premium AI tools.

Pragmatic approaches might prioritize free or low-cost AI tools leveraging natural language processing and machine learning technologies (Villarroel Carrillo et al., 2025); school-based implementation leveraging available computer labs; teacher professional development emphasizing pedagogical integration rather than technical

expertise (Lucas Soledispa et al., 2023); and blended models combining periodic AI-supported practice with continued emphasis on face-to-face communicative activities (Rojas-Burbano & Naranjo-Andrade, 2025).

Limitations of the Review

Several limitations constrain the interpretation and generalizability of findings. First, the absence of studies conducted specifically in Ecuadorian contexts or even extensive Latin American representation limits direct applicability of findings to the specific constraints and affordances characterizing Ecuadorian secondary education.

Second, the heterogeneity of outcome measures, intervention designs, and contextual variables precluded meta-analysis and limited quantitative synthesis of effect sizes. This methodological diversity reflects the emerging nature of research in this area but constrains conclusions about comparative effectiveness.

Third, most included studies implemented short interventions (≤8 weeks) with limited follow-up assessment. Long-term effectiveness, sustained engagement beyond novelty periods, and transfer to authentic communicative contexts remain inadequately examined (Jiménez-Tuza, 2025).

Fourth, limited attention to equity considerations means differential impacts across student subgroups—including varying proficiency levels, socioeconomic backgrounds, learning differences, and digital literacy—remain unclear. This gap is particularly concerning for contexts like Ecuador with substantial educational inequities.

Fifth, the search was limited to studies published in English or Spanish in selected academic databases, potentially excluding relevant research published in other languages or grey literature sources.

Finally, the rapid evolution of AI technologies means findings from studies conducted even 2-3 years ago may have limited applicability to current tools, particularly following the transformative emergence of large language models in 2022-2023.

5. Conclusiones

This systematic review examined AI-enhanced speaking practice in upper-secondary EFL classrooms, synthesizing evidence from 21 empirical studies published between January 2020 and March 2025. The review addressed three research questions regarding AI tools used, their impacts on speaking proficiency, and pedagogical implementation factors.

The findings demonstrate that AI-based tools—conversational chatbots, automatic speech recognition systems, and adaptive learning platforms—can effectively support speaking skill development, particularly for fluency and communicative confidence. Fifteen of 18 studies reported significant improvements in fluency measures, while 14 of 16 studies documented increased confidence and reduced speaking anxiety. These

tools provide unlimited practice opportunities, immediate feedback, and low-anxiety environments that address key constraints in traditional EFL classrooms.

However, effectiveness depends critically on implementation approach. Blended models combining AI-supported practice with teacher guidance and face-to-face interaction showed more consistent gains than purely autonomous AI use. Technical challenges, teacher preparation requirements, and motivational sustainability emerged as important considerations, particularly for resource-constrained contexts.

This review contributes to the field by providing systematic synthesis focused specifically on upper-secondary EFL speaking practice, identifying implementation factors relevant to contexts like Ecuador with large class sizes and limited instructional time. The findings suggest AI tools offer promising possibilities for extending practice opportunities and providing individualized feedback where traditional approaches face structural constraints.

For Ecuadorian secondary education, where students frequently fail to reach required oral proficiency levels, AI-enhanced speaking practice represents a potentially transformative intervention. Free or low-cost tools could enable self-directed practice beyond limited class time, while ASR systems could provide pronunciation feedback impossible for teachers to deliver individually to 40+ students per class.

Future research should prioritize several directions. First, implementation studies in Latin American contexts, particularly Ecuador, examining how identified challenges and opportunities manifest in specific institutional and cultural settings. Second, longitudinal investigations tracking effectiveness and engagement beyond short-term interventions to understand sustained impacts and optimal integration patterns. Third, equity-focused research examining differential effects across student subgroups, including varying proficiency levels, socioeconomic backgrounds, and learning needs. Fourth, comparative studies of different AI tool types and pedagogical integration models to identify effective practices for diverse contexts. Finally, investigations of teacher professional development approaches supporting successful AI integration in resource-limited settings.

As AI technologies continue evolving rapidly, ongoing research must examine how emerging capabilities can be leveraged effectively while addressing implementation realities in diverse educational contexts. The potential of AI-enhanced speaking practice will be realized not through technology alone, but through thoughtful pedagogical integration responsive to learners' needs and institutional constraints.

CONFLICTO DE INTERESES

"Los autores declaran no tener ningún conflicto de intereses".

Referencias Bibliográficas

Alvarez, C., Tamayo, M. R., & Coutinho dos Santos, J. (2024). Factors influencing the development of speaking skills among Ecuadorian EFL learners: Teachers' perspectives. *Indonesian Journal of Applied Linguistics, 14*, 319-331. https://doi.org/10.17509/ijal.v14i2.74889

Ayala-Pazmiño, M., & Alvarado-Lucas, K. (2023). Integración de la Inteligencia Artificial en la Educación del Idioma Inglés en Ecuador: Un Camino para Mejorar los Resultados del Aprendizaje. *593 Digital Publisher CEIT, 8*, 679-687. https://doi.org/10.33386/593dp.2023.3-1.1862

Bernal Párraga, A. P., Coronel Ramírez, E. A., Aldas Macias, K. J., Carvajal Madrid, C. A., Valarezo Espinoza, B. D. C., Vera Alcivar, J. G., & Chávez Cedeño, J. U. (2025). The impact of artificial intelligence on personalized learning in English language education. *Ciencia Latina Revista Científica Multidisciplinar, 9*(1), 5500-5518. https://doi.org/10.37811/cl_rcm.v9i1.16234

Cárdenas, J. (2025). Integración de inteligencia artificial en la enseñanza del inglés técnico para personal de comunicaciones de la fuerza aérea ecuatoriana. *ASCE Magazine, 4*, 436-453. https://doi.org/10.70577/ASCE/436.453/2025

Dávila Macías, A. M., Armijos Solano, D. O., Palma Perero, L. M., Roca Panimboza, J. A., & Lucas Soledispa, C. J. (2024). The potential of artificial intelligence to improve speaking skills in a second language (English) fluently. *Ciencia Latina Revista Científica Multidisciplinar, 8*(3), 3826-3836. https://doi.org/10.37811/cl_rcm.v8i3.11592

Guerrero Rodriguez, S. E., & Moreira Baquerizo, A. S. (2025). Ecuadorian EFL teachers' experiences in fostering students' English-speaking skills: Insights into strategies and challenges in public and private schools. *UNESUM-Ciencias. Revista Científica Multidisciplinaria, 9*(2), 124-136. https://doi.org/10.47230/unesum-ciencias.v9.n2.2025.124-136

Guevara Peñaranda, N. S., Zambrano Pachay, J. F., & Fabre Mérchan, P. G. (2024). EFL curriculum in Ecuador: The achievement of communicative competences in learners graduated from Ecuadorian public schools in Milagro. *Ciencia Latina Revista Científica Multidisciplinar, 8*(2), 6942-6954. https://doi.org/10.37811/cl_rcm.v8i2.11100

Guillermo Morales, L. E. (2024). El efecto del aula invertida en el aprendizaje de inglés: Revisión sistemática. *Horizontes. Revista de Investigación en Ciencias de la Educación, 8*(32), 544-559. https://doi.org/10.33996/revistahorizontes.v8i32.743

Hernández Pacheco, J., Chacón Cárdenas, A., Lasluisa Naranjo, G., & Romero Cevallos, M. (2025). Integración de herramientas de inteligencia artificial para mejorar la enseñanza del idioma inglés en Ecuador. *Polo del Conocimiento, 10*(1), 1571-1594. https://doi.org/10.23857/pc.v10i1.8770

Jiménez-Tuza, S. B. (2025). Uso de la inteligencia artificial en la dirección de centros educativos. *Revista Científica Zambos, 4*(1), 191-204. https://doi.org/10.69484/rcz/v4/n1/86

Lopez, J., Becerra, A., & Ramirez-Avila, M. (2021). EFL speaking fluency through authentic oral production. *Journal of Foreign Language Teaching and Learning, 6.* https://doi.org/10.18196/ftl.v6i1.10175

Lucas Soledispa, C. J., Mantilla Carrera, P. N., Dávila Macías, A. M., Jaramillo Crespo, L. F., & Armijos Solano, D. O. (2023). Perspectiva de profesores de inglés acerca del impacto de la inteligencia artificial en los cursos de idiomas. *Ciencia Latina Revista Científica Multidisciplinar, 7*(4), 8278-8295. https://doi.org/10.37811/cl_rcm.v7i4.7562

Ministerio de Educación de Ecuador. (2016). English language curriculum. https://educacion.gob.ec/wp-content/uploads/downloads/2016/08/EFL-for-Subnivel-BGU-final-ok.pdf

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: *The PRISMA statement. PLoS Medicine, 6*, e1000097. https://doi.org/10.1016/j.jclinepi.2009.06.005

Oshimeje, S., & Flores Barahona, I. D. (2025). Applying authentic oral production to improve speaking fluency through project-based learning approach on EFL students. *Ciencia Latina Revista Científica Multidisciplinar, 9*(4), 2304-2323. https://doi.org/10.37811/cl_rcm.v9i4.18849

Rojas-Burbano, S. C., & Naranjo-Andrade, S. S. (2025). Prácticas en inglés como lengua extranjera: Enfoques innovadores y centrados en el estudiante. *MQRInvestigar, 9*(3), e892. https://doi.org/10.56048/MQR20225.9.3.2025.e892

Sangacha-Tapia, L. M., Celi, R. J., Acosta-Guzmán, I. L., & Varela-Tapia, E. A. (2024). *Inteligencia artificial aplicada a procesamiento de lenguaje natural (NLP) con Python y machine learning.* Editorial Grupo AEA. https://doi.org/10.55813/egaea.l.88

T. Soto, S., Espinosa Cevallos, L. F., & Rojas Encalada, M. A. (2025). Studies on motivation and EFL teaching and learning in Ecuador. Revista InveCom, 5(1), e501006. https://doi.org/10.5281/zenodo.10892449

Villarroel Carrillo, S., Castillo Salazar, L., Granda Aguilera, D., Lema Mullo, L., & Carranza Ortiz, E. (2025). Desarrollo de un sistema educativo inteligente en Python para la personalización del aprendizaje mediante técnicas de inteligencia artificial como machine learning y procesamiento de lenguaje natural. *Polo del Conocimiento, 10*(11), 1560-1573. https://doi.org/10.23857/pc.v10i11.10718

Yánez-Goyes, M., Peñaherrera-Solarte, K., Carlín-Chávez, E., & Bonilla-Tenesaca, J. (2024). Las TIC en la enseñanza del inglés para la educación básica: Una revisión sistemática. *593 Digital Publisher CEIT, 9*(3), 98-110. https://doi.org/10.33386/593dp.2024.3.2334