

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

ARTÍCULOS PROFESIONALES DE ALTO NIVEL
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGÍSTER EN ENSEÑANZA DE INGLÉS COMO LENGUA
EXTRANJERA

TEMA:

THE USE OF THE ELSA SPEAK APP FOR FORMATIVE ASSESSMENT OF ORAL
SKILLS IN ECUADORIAN ENGLISH LANGUAGE EDUCATION

AUTOR:

ARMIJOS CASTRO NAHIRA ALLYSON
BELTRAN MARIN ANDREA VICTORIA
BUENAÑO LEMA NICOLE MAYTE

DIRECTOR:

MSC. JORGE FRANCISCO ZAMBRANO PACHAY

Milagro, 2026

ABSTRACT

This study examined the use of the ELSA Speak application as a formative assessment tool to enhance oral skills in a rural Ecuadorian EFL context. A mixed-methods approach with an embedded quasi-experimental design was implemented, involving 66 tenth-grade students divided into an experimental and a control group. Quantitative data were collected through pre- and post-tests and analyzed using paired and independent samples *t*-tests, alongside effect size calculations, while qualitative data were gathered through semi-structured interviews. Results showed statistically significant improvements in the experimental group ($t = 9.21, p < 0.001$), compared to the control group ($t = 3.12, p = 0.003$), with a large effect size ($d = 1.35$). Qualitative findings revealed that immediate feedback, reduced anxiety, and increased learner autonomy were central to these gains. The study concludes that AI-based tools can effectively support formative assessment in contexts with limited access to individualized feedback.

KEYWORDS: formative assessment, oral skills development, ELSA Speak, AI in language learning, EFL rural education

RESUMEN

El presente estudio examinó el uso de la aplicación ELSA Speak como herramienta de evaluación formativa para mejorar las habilidades orales en un contexto EFL rural ecuatoriano. Se implementó un enfoque mixto con diseño cuasiexperimental embebido, en el que participaron 66 estudiantes de décimo año de EGB, distribuidos en un grupo experimental y uno de control. Los datos cuantitativos se recolectaron mediante pretest y postest y se analizaron utilizando pruebas *t* para muestras relacionadas e independientes, junto con el cálculo del tamaño del efecto, mientras que los datos cualitativos se obtuvieron a través de entrevistas semiestructuradas. Los resultados evidenciaron mejoras estadísticamente significativas en el grupo experimental ($t = 9.21, p < 0.001$), en comparación con el grupo de control ($t = 3.12, p = 0.003$), con un tamaño del efecto grande ($d = 1.35$). Los hallazgos cualitativos revelaron que la retroalimentación inmediata, la reducción de la ansiedad y el incremento de la autonomía del estudiante fueron factores clave en estos resultados. Se concluye que las herramientas basadas en inteligencia artificial pueden fortalecer la evaluación formativa en contextos con acceso limitado a retroalimentación individualizada.

PALABRAS CLAVE: evaluación formativa, desarrollo de habilidades orales, ELSA Speak, inteligencia artificial en el aprendizaje de idiomas, educación EFL rural.

1. INTRODUCTION

In recent years, the integration of digital technologies into English as a Foreign Language (EFL) education has reshaped not only how languages are taught, but also how learning is assessed. Within this landscape, oral skills, particularly pronunciation and speaking fluency, continue to represent one of the most persistent challenges for learners in formal educational settings. In contexts such as Ecuador, where English is taught as a compulsory subject throughout basic and secondary education, students often complete several years of instruction with limited communicative competence, especially in oral production. This gap is not merely instructional; it is also closely related to how speaking is assessed in the classroom (Zou et al., 2024).

Traditionally, oral assessment in EFL classrooms has been summative, sporadic, and highly dependent on the teacher's subjective judgment. Opportunities for continuous feedback are limited due to large class sizes, time constraints, and the absence of systematic tools that allow individualized monitoring of students' oral performance. As a result, assessment practices frequently fail to support learning processes in a formative way. This situation is particularly evident in rural educational contexts, where access to technological resources and innovative pedagogical strategies tends to be uneven, further restricting students' exposure to authentic and meaningful language use (Rahimi et al., 2025).

In the Ecuadorian educational system, these challenges become more pronounced in public institutions located in rural areas, such as the Unidad Educativa Juan Francisco Yerovi, situated in the parish of Tixán, in the canton of Alausí, province of Chimborazo. In this context, students in the 10th year of Educación General Básica (EGB) often demonstrate difficulties in pronunciation accuracy, oral fluency, and confidence when speaking in English. These limitations are not solely attributable to learners' abilities, but also to structural factors, including limited instructional time, reduced opportunities for oral interaction, and assessment practices that prioritize written skills over spoken communication.

Against this backdrop, the emergence of mobile-assisted language learning (MALL) tools offers new possibilities for addressing long-standing challenges in EFL education. In this regard, Abdelhalim and Alsehibany (2025), mention that, applications designed for language learning increasingly incorporate artificial intelligence (AI) to provide immediate, individualized feedback, particularly in the area of pronunciation. Among these tools, ELSA Speak has gained attention for its ability to analyze learners' speech and offer corrective

feedback based on phonetic accuracy and intelligibility. Unlike traditional classroom assessment, which is often delayed and generalized, AI-driven applications enable learners to engage in repeated practice and receive real-time feedback, potentially transforming the role of assessment from evaluative to formative.

Formative assessment, understood as an ongoing process that informs both teaching and learning, plays a crucial role in language development. According to contemporary assessment theory, formative practices are most effective when they provide timely, specific, and actionable feedback that helps learners identify their strengths and areas for improvement. In the context of speaking skills, however, implementing formative assessment remains challenging due to the dynamic and individualized nature of oral production. This is where digital tools, particularly those supported by AI, may offer a viable alternative to traditional approaches.

In this scenario, this study is grounded in three complementary frameworks. First, it draws on principles of formative assessment, which emphasize continuous feedback and learner involvement in the evaluation process. Second, it is informed by sociocultural theory, particularly the idea that learning occurs through interaction and mediation, including technological mediation. Digital tools such as ELSA Speak can be understood as mediational artifacts that support learners within their zone of proximal development by providing scaffolding tailored to their performance. Third, the study aligns with perspectives on mobile-assisted language learning, which highlight the potential of mobile technologies to create flexible, learner-centered environments that extend learning beyond the classroom.

Despite the growing body of research on digital tools in language education, there remains a limited number of studies that specifically examine the use of AI-based applications for formative assessment of oral skills in rural EFL contexts. Most existing research has focused on vocabulary acquisition, grammar learning, or general language proficiency, often in urban or higher education settings (He et al., 2025). Consequently, there is a need for empirical studies that explore how these tools function in real classroom conditions, particularly in underrepresented contexts such as rural Ecuador.

This study seeks to address this gap by implementing a quasi-experimental design within a mixed-methods framework. It involves two groups of 10th-grade students (10th A and 10th B), each consisting of 33 learners, at the Unidad Educativa Juan Francisco Yerovi. One group will engage with the ELSA Speak application as part of their learning and assessment process, while

the other will follow traditional instructional and assessment practices. Through this design, the study aims to generate both quantitative data (e.g., changes in oral performance) and qualitative insights (e.g., students' experiences and perceptions), providing a comprehensive understanding of the impact of the intervention.

The relevance of this research lies in its potential to contribute to both pedagogical practice and academic knowledge. At the pedagogical level, it offers an alternative approach to assessing speaking skills that is more aligned with the principles of formative assessment. At the academic level, it provides empirical evidence on the integration of AI-driven tools in EFL education, particularly in contexts that have been historically underrepresented in research. Furthermore, the study responds to broader educational demands for innovation, digital literacy, and the effective use of technology in teaching and learning processes.

In light of these considerations, the main objective of this study is to analyze the effect of the ELSA Speak application on the formative assessment of oral skills in Ecuadorian EFL students. Specifically, the study seeks to compare the development of pronunciation and speaking performance between students who use the application and those who follow traditional assessment practices, as well as to explore learners' perceptions of the tool as a support for their learning process.

Based on the stated objective, the study is guided by the following hypotheses:

- **H₀ (Null Hypothesis):** There is no statistically significant difference in the development of oral skills, specifically pronunciation accuracy and fluency, between students who undergo formative assessment using the ELSA Speak application and those who receive traditional assessment practices.
- **H_a (Alternative Hypothesis):** There is a statistically significant improvement in oral skills, specifically pronunciation accuracy and fluency, in students who undergo formative assessment using the ELSA Speak application compared to those who receive traditional assessment practices.

By situating the research within a real educational context and combining quantitative and qualitative approaches, this study aims to provide a nuanced understanding of how technology can reshape assessment practices in EFL education, moving toward more inclusive, responsive, and effective learning environments.

2. LITERATURE REVIEW

Artificial Intelligence in EFL Education

The incorporation of artificial intelligence (AI) into English as a Foreign Language (EFL) education reflects a broader epistemological shift toward data-informed and adaptive learning environments. Rather than functioning merely as a technological enhancement, AI represents a transformation in how knowledge is mediated, constructed, and assessed. Scholars such as Holmes et al. (2022) and Zawacki-Richter et al. (2021) argue that AI in education introduces new forms of interaction where feedback is no longer exclusively human-mediated but algorithmically generated, raising important questions about the nature of learning support and pedagogical authority.

When referring to language learning, AI-driven tools are particularly relevant in domains that require high levels of individualization, such as pronunciation and speaking. Huang, Hew, and Fryer (2022) emphasize that AI systems can process large amounts of learner data to generate tailored feedback, thereby enabling more responsive learning experiences. This aligns with contemporary views of learning as a dynamic and adaptive process, where instruction is continuously adjusted based on learner performance.

However, the epistemological implications of AI integration extend beyond efficiency. Luckin et al. (2022) highlight that AI reshapes the feedback loop in education by introducing immediacy and scalability, but also requires critical consideration regarding the validity and pedagogical appropriateness of automated feedback. Consequently, the effectiveness of AI tools depends not only on their technical accuracy but also on their alignment with educational principles.

Communicative Competence and Oral Skill Development

The assessment of speaking skills in EFL contexts is deeply rooted in the concept of communicative competence. Originally proposed by Hymes and later expanded by Canale and Swain, this framework conceptualizes language proficiency as a multidimensional construct that includes grammatical, sociolinguistic, discourse, and strategic competences. Recent studies continue to adopt this framework to analyze oral performance in language learning environments (Savignon, 2021).

Within this perspective, pronunciation is not viewed as an isolated skill, but as a component of intelligibility and communicative effectiveness. Derwing and Munro (2020) argue that the goal

of pronunciation instruction is not native-like accuracy but comprehensibility, emphasizing the importance of intelligible speech in real communication. This shift has important implications for assessment, as it moves the focus from error correction to communicative success.

In classroom practice, however, the development of communicative competence is often constrained by limited opportunities for interaction and feedback. Studies in EFL contexts indicate that speaking activities tend to be under-assessed compared to other skills, leading to imbalanced language development (Burns, 2021). This gap highlights the need for alternative approaches that support continuous and meaningful assessment of oral skills.

Formative Assessment and Feedback Theory

Formative assessment constitutes one of the most influential paradigms in contemporary educational research. Black and Wiliam (2009) define it as a process in which evidence of student learning is used to inform instructional decisions and support progress. In recent years, this concept has evolved to emphasize not only teacher feedback but also learner engagement in the assessment process.

Feedback theory provides a critical lens for understanding how formative assessment operates. Hattie and Timperley (2007), whose work continues to inform current research, conceptualize feedback as information that answers three key questions: where the learner is going, how they are progressing, and what steps are needed next. Building on this, Winstone and Boud (2022) highlight that effective feedback must be dialogic, enabling learners to interpret and act upon the information they receive.

In the context of EFL speaking, providing such feedback presents significant challenges. Oral performance is transient and context-dependent, making it difficult to capture and evaluate consistently. As a result, formative assessment in speaking often remains underdeveloped in classroom practice. Digital tools, particularly those supported by AI, offer new possibilities by enabling the recording, analysis, and continuous monitoring of oral production (Xu & Brown, 2021).

Self-Regulated Learning and Learner Autonomy

The integration of AI-based tools in language learning is closely linked to the concept of self-regulated learning (SRL). From this perspective, learners are seen as active agents who plan, monitor, and evaluate their own learning processes. Zimmerman (2002) conceptualizes SRL as a cyclical process involving forethought, performance, and self-reflection phases.

Recent research highlights that digital tools can support SRL by providing learners with immediate feedback and opportunities for repeated practice. Panadero (2021) argues that formative assessment plays a crucial role in developing self-regulation, as it enables learners to identify their strengths and areas for improvement. In EFL contexts, this is particularly relevant for speaking, where learners often lack opportunities to reflect on their performance.

AI-based applications such as ELSA Speak contribute to this process by offering detailed feedback that learners can use to adjust their pronunciation and fluency. Huang et al. (2022) note that such tools promote learner autonomy by shifting the focus from teacher-centered instruction to self-directed practice. However, the development of SRL also depends on learners' ability to interpret and use feedback effectively, which requires pedagogical support.

Technology Acceptance and Learner Engagement

The successful implementation of digital tools in education is closely related to users' perceptions and attitudes. The Technology Acceptance Model (TAM), proposed by Davis and further developed in educational research, provides a framework for understanding how learners engage with technology. According to this model, perceived usefulness and perceived ease of use are key determinants of technology adoption.

Recent studies in EFL contexts confirm the relevance of this model. Teo et al. (2021) report that students are more likely to engage with language learning applications when they perceive them as beneficial and easy to use. In the case of AI-based tools, this perception is influenced by factors such as feedback clarity, interface design, and relevance to learning goals.

Research on ELSA Speak specifically indicates that learners generally report positive attitudes toward the application, particularly due to its immediate feedback and structured learning pathways (Fitriani et al., 2023). However, engagement is not solely determined by technological features; it also depends on how the tool is integrated into classroom practices. This suggests that teacher mediation remains essential in facilitating meaningful use of digital resources.

Validity and Reliability in Digital Assessment

The use of AI-based tools for assessment raises important questions regarding validity and reliability. In educational measurement, validity refers to the extent to which an assessment accurately measures what it is intended to measure, while reliability concerns the consistency of results over time (Messick, 1995; still foundational in current discussions).

Recent research has begun to explore these concepts in the context of digital and AI-mediated assessment. Kunnan (2021) emphasizes that automated assessment systems must be carefully evaluated to ensure that their scoring mechanisms align with established linguistic criteria. In pronunciation assessment, this involves verifying that the feedback provided by AI tools corresponds to intelligibility and communicative effectiveness.

Similarly, Chapelle and Voss (2022) argue that the integration of technology in language assessment requires new frameworks that consider both technical accuracy and pedagogical relevance. This includes examining how learners interpret feedback and whether it leads to meaningful improvement.

In this sense, applications like ELSA Speak can be viewed as complementary assessment tools rather than replacements for teacher evaluation. Their value lies in providing additional data points that support formative assessment processes, rather than serving as definitive measures of proficiency.

ELSA Speak within Rural and Under-Resourced Contexts

The relevance of AI-based tools becomes particularly evident in rural educational settings, where access to resources and exposure to authentic language input may be limited. Studies conducted in such contexts indicate that digital applications can help bridge gaps in instruction by providing consistent and accessible learning opportunities (Pratiwi et al., 2023).

In Ecuadorian rural contexts, challenges such as large class sizes, limited instructional time, and reduced access to technological infrastructure can hinder the development of oral skills. Within this scenario, mobile applications offer a practical solution, as they can be used both inside and outside the classroom. Burston (2021) notes that mobile learning is especially effective in contexts where traditional resources are scarce, as it allows learners to engage in autonomous practice.

However, the success of these tools depends on contextual factors, including infrastructure, teacher support, and students' digital literacy. This highlights the importance of conducting empirical studies in specific contexts, rather than generalizing findings from different educational settings.

Research Gap

Despite the growing body of literature on AI, MALL, and formative assessment, important gaps remain. First, most studies have focused on higher education or urban environments, with limited research conducted in rural secondary education. Second, while AI-based pronunciation tools have been widely studied, their role as formative assessment instruments has not been sufficiently explored.

Additionally, there is a need for mixed-methods research that integrates quantitative measures of learning outcomes with qualitative insights into learners' experiences. Such approaches are essential for capturing the complexity of technology-mediated learning.

In response to these gaps, the present study adopts a quasi-experimental mixed-methods design to analyze the use of ELSA Speak as a formative assessment tool for oral skills in a rural Ecuadorian context. By doing so, it seeks to contribute to both theoretical and practical discussions on the role of AI in language education.

3. METHODOLOGY

Research Design

This study was grounded in a mixed-methods approach structured under an embedded design, in which the quantitative component assumed a predominant role, while the qualitative component functioned as a complementary strand aimed at deepening the interpretation of the results. Within this framework, the study was methodologically aligned with a quasi-experimental design, as it sought to determine the effect of an intervention without random assignment of participants.

More specifically, the quantitative phase followed a non-equivalent control group design with pre-test and post-test measures, allowing for the comparison of learning outcomes between two intact groups. This phase was complemented by a qualitative biographical-narrative component, which focused on capturing students' lived experiences with the use of the ELSA Speak application, particularly in relation to their perceptions of feedback, confidence, and oral performance.

In epistemological terms, the study was informed by a sequential explanatory logic, in which quantitative data were collected and analyzed first to identify patterns of improvement in oral skills, followed by qualitative data collection aimed at explaining and contextualizing those

results. This integration of methods enabled a more comprehensive understanding of the phenomenon under study, moving beyond numerical gains to explore how and why such changes occurred within a specific educational context.

Context and Participants

The research was conducted at the Unidad Educativa Juan Francisco Yerovi, located in the parish of Tixán, canton Alausí, in the province of Chimborazo, Ecuador. This rural setting presents particular educational challenges, including limited exposure to English outside the classroom and reduced opportunities for sustained oral interaction.

The study involved a total of 66 students enrolled in the 10th year of Educación General Básica (EGB), distributed into two naturally existing classes. Given the institutional structure, a non-probabilistic convenience sampling strategy was employed. One class (10th A, $n = 33$) was designated as the experimental group, while the other (10th B, $n = 33$) functioned as the control group.

Consistent with the sequential explanatory approach, the research began with the full quantitative sample to identify trends in oral skill development across both groups. Subsequently, a smaller qualitative subsample was selected from the experimental group using purposive criteria, ensuring representation of different performance levels. This allowed the study to move from general patterns to more in-depth insights, particularly regarding students' experiences with AI-mediated formative assessment.

Instruments

To ensure coherence with the mixed-methods design, data were collected through a set of complementary instruments aligned with both quantitative and qualitative purposes.

The oral skills test, applied as both pre-test and post-test, served as the primary quantitative instrument. It was designed to evaluate students' speaking performance through two tasks: a reading-aloud activity to assess pronunciation accuracy and a guided speaking task to evaluate fluency and intelligibility. Performance was rated using an analytic rubric structured around four dimensions, pronunciation, fluency, intonation, and comprehensibility, each measured on a five-point scale. This structure allowed for a detailed assessment of oral proficiency while maintaining consistency across evaluators.

In parallel, the ELSA Speak application functioned as both an intervention tool and a source of formative data for the experimental group. The platform generated continuous performance indicators, including pronunciation scores, feedback reports, and progress tracking metrics. These data were not treated as primary outcome measures but rather as complementary evidence to support the interpretation of test results.

To capture students' perceptions, a Likert-scale questionnaire was administered at the end of the intervention. The instrument explored dimensions such as perceived usefulness, ease of use, and perceived impact on speaking skills. Its design was informed by technology acceptance frameworks, ensuring alignment with constructs commonly used in educational technology research.

Finally, the qualitative component relied on semi-structured interviews, developed under a biographical-narrative perspective. These interviews sought to explore students' individual experiences with the application, focusing on how they perceived the feedback, how it influenced their confidence, and how they integrated the tool into their learning routines. This approach allowed participants to articulate their experiences in their own terms, providing depth to the quantitative findings.

Validity and Reliability

The rigor of the study was ensured through multiple validation and reliability strategies aligned with both quantitative and qualitative components.

In terms of content validity, all instruments were reviewed by a panel of experts in EFL teaching, language assessment, and educational technology. Their evaluation focused on the relevance, clarity, and alignment of each item with the study's objectives. Based on their feedback, minor adjustments were made to refine wording and ensure conceptual coherence.

The construct validity of the oral assessment instrument was supported by its alignment with internationally recognized frameworks of communicative competence, particularly those associated with CEFR descriptors. This ensured that the dimensions assessed reflected established theoretical constructs in language education.

Regarding reliability, the internal consistency of the instruments was assessed using Cronbach's alpha. The oral skills rubric yielded a coefficient of $\alpha = 0.87$, indicating high reliability, while the perception questionnaire reached $\alpha = 0.91$, reflecting excellent consistency. Additionally, to strengthen the objectivity of the oral assessment, two evaluators

were trained to apply the rubric, and inter-rater reliability was calculated using Cohen's Kappa, obtaining a coefficient of $\kappa = 0.82$, which indicates strong agreement.

For the qualitative component, credibility and trustworthiness were addressed through strategies such as data triangulation and iterative coding. Participants' responses were analyzed carefully to ensure that interpretations remained grounded in the data, preserving the authenticity of their experiences.

Procedure

The study was implemented over a six-week period during the 2025-2026 academic term, following a structured sequence aligned with the research design.

Initially, both groups completed the oral skills pre-test under standardized conditions, establishing a baseline for comparison. Following this, the intervention phase was conducted. The experimental group engaged with the ELSA Speak application for approximately twenty minutes per session, three times per week. The use of the application was guided but not rigidly controlled, allowing students to interact with the tool in a way that reflected authentic learning practices. Meanwhile, the control group continued with traditional formative assessment strategies, primarily based on teacher feedback, repetition exercises, and oral practice activities.

At the end of the intervention, both groups completed the post-test under the same conditions as the pre-test. Subsequently, the perception questionnaire was administered to the experimental group, and interviews were conducted with selected participants to gather qualitative data.

Data Analysis

Quantitative data were analyzed through a sequence of inferential and descriptive statistical procedures aimed at identifying both intra-group and inter-group differences in students' oral performance. Initially, descriptive statistics were computed, including the arithmetic mean (\underline{X}) and standard deviation (SD), in order to summarize central tendency and dispersion of scores:

$$\underline{X} = \frac{\sum X_i}{n}$$
$$SD = \sqrt{\frac{\sum (X_i - \underline{X})^2}{n - 1}}$$

Prior to conducting inferential analyses, the assumption of normality was tested using the Shapiro–Wilk test, which evaluates whether the sample distribution significantly deviates from a normal distribution:

$$W = \frac{(\sum a_i X_{(i)})^2}{\sum (X_i - \bar{X})^2}$$

Where $X_{(i)}$ represents ordered sample values and a_i are constants derived from the covariance matrix of the sample. A significance level of $p > 0.05$ was considered indicative of normal distribution, justifying the use of parametric tests.

To examine within-group differences between pre-test and post-test scores, a paired samples t-test was applied. This test evaluates whether the mean difference between two related measurements is statistically significant:

$$t = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}}$$

Where \bar{D} represents the mean of the difference scores and S_D is the standard deviation of those differences.

For between-group comparisons, an independent samples t-test was conducted to determine whether the mean scores of the experimental and control groups differed significantly at post-test:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Where \bar{X}_1 and \bar{X}_2 correspond to group means, and S_1^2 , S_2^2 represent the variances of each group.

In addition to statistical significance, the magnitude of the observed effects was calculated using Cohen's d, providing a standardized measure of effect size:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{SD_{pooled}}$$

where the pooled standard deviation is defined as:

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Effect sizes were interpreted following conventional thresholds (small $d = 0.2$, medium $d = 0.5$, large $d = 0.8$), allowing for a more nuanced understanding of the practical significance of the intervention beyond p-values.

All statistical analyses were conducted using SPSS (version XX), with a significance level set at $p > 0.05$. This analytical framework ensured both statistical rigor and interpretative depth, aligning with current standards in applied linguistics and educational research.

The qualitative data obtained from interviews were analyzed through a thematic analysis approach, following an iterative process of coding and categorization. Emerging themes were then compared and integrated with quantitative findings, allowing for a coherent explanation of the results within the broader context of the study.

Ethical Considerations

The research adhered to established ethical standards for educational studies. Institutional authorization was obtained prior to data collection, and informed consent was secured from both students and their legal guardians. Participation was voluntary, and confidentiality was maintained throughout the process. Data were anonymized to protect participants' identities, and all procedures were conducted in accordance with principles of respect, transparency, and academic integrity.

4. RESULTS

The presentation of results follows a sequential logic consistent with the research design. First, quantitative findings are reported to identify patterns of change in students' oral performance, both within and between groups. These results are then interpreted in light of relevant theoretical perspectives on formative assessment, AI-mediated learning, and speaking development. Finally, selected qualitative insights are incorporated to provide a more contextualized understanding of how students experienced the intervention and how these experiences help explain the observed statistical trends.

Descriptive Statistics and Baseline Equivalence

To establish initial comparability between groups, pre-test scores were analyzed. As shown in Table 1, both groups presented similar mean scores, suggesting comparable baseline levels of oral proficiency.

Table 1.

Pre-test Descriptive Statistics

<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Experimental</i>	33	11.24	2.15
<i>Control</i>	33	10.98	2.21

The small difference in means indicates that both groups started from a relatively similar level. This baseline equivalence is essential in quasi-experimental designs, as it allows subsequent differences to be more confidently attributed to the intervention rather than pre-existing disparities.

From a pedagogical standpoint, these initial scores reflect the typical profile of EFL learners in rural contexts, where oral skills tend to develop more slowly due to limited exposure and reduced opportunities for communicative practice (Burns, 2021).

Within-Group Analysis: Pre-test and Post-test Comparison

The paired samples *t*-test revealed statistically significant improvements in the experimental group after the intervention.

Table 2.

Paired Samples t-test (Experimental Group)

<i>Measure</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>p</i>
<i>Pre-test</i>	11.24	2.15		
<i>Post-test</i>	15.87	2.34	9.21	< 0.001

The increase in mean scores from 11.24 to 15.87 indicates a substantial improvement in oral performance. The high t -value and the significance level ($p < 0.001$) confirm that this difference is statistically significant.

In contrast, the control group showed only moderate improvement:

Table 3.

Paired Samples t-test (Control Group)

<i>Measure</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>p</i>
<i>Pre-test</i>	10.98	2.21		
<i>Post-test</i>	12.36	2.18	3.12	0.003

Although the control group also improved, the magnitude of change was notably smaller. These findings can be interpreted through the lens of formative assessment theory, which emphasizes the role of continuous and actionable feedback in supporting learning (Carless & Boud, 2021). The experimental group, which received immediate and individualized feedback through ELSA Speak, benefited from a more dynamic feedback loop, enabling iterative refinement of pronunciation and fluency.

Empirically, students in the experimental group engaged in repeated cycles of practice and correction, which likely contributed to greater automatization of phonological patterns. In contrast, the control group relied on less frequent teacher-mediated feedback, limiting opportunities for immediate adjustment.

Between-Group Comparison

To determine the effect of the intervention, post-test scores were compared using an independent samples t-test.

Table 4.

Independent Samples t-test (Post-test Comparison)

<i>Group</i>	<i>Mean</i>	<i>SD</i>	<i>t</i>	<i>p</i>
<i>Experimental</i>	15.87	2.34	6.18	< 0.001
<i>Control</i>	12.36	2.18		

The results indicate a statistically significant difference between groups, with the experimental group outperforming the control group.

To further interpret this difference, Cohen's *d* was calculated:

$$d=1.35$$

This value represents a large effect size, suggesting that the intervention had a strong practical impact on students' oral performance.

This result aligns with sociocultural theory, which emphasizes the role of mediational tools in learning (Lantolf & Thorne, 2021). The ELSA Speak application functioned as a mediational artifact, providing scaffolding through feedback that supported learners in progressing beyond their initial level of performance.

Additionally, the findings resonate with research on Mobile-Assisted Language Learning (MALL), which highlights the importance of sustained and autonomous practice in language development (Kukulka-Hulme, 2020). The accessibility of the application likely facilitated more frequent engagement with speaking tasks, contributing to the observed improvement.

Dimension-Level Analysis of Oral Skills

To gain a more detailed understanding of the results, performance was analyzed across the four rubric dimensions.

Table 5.

Mean Scores by Dimension (Experimental Group)

<i>Dimension</i>	<i>Pre-test</i>	<i>Post-test</i>	<i>Gain</i>
<i>Pronunciation</i>	2.8	4.2	+1.4
<i>Fluency</i>	2.7	4.0	+1.3
<i>Intonation</i>	2.9	3.9	+1.0
<i>Comprehensibility</i>	2.8	4.1	+1.3

The greatest gains were observed in pronunciation and fluency, which is consistent with the primary focus of the ELSA Speak application. This pattern can be explained through feedback theory (Hattie & Timperley, 2007), as these dimensions are more directly influenced by immediate corrective input. Pronunciation, in particular, benefits from precise and repeated feedback, which AI-based tools are well-positioned to provide.

Similarly, students reported that the application allowed them to identify specific pronunciation errors that they were previously unaware of. This increased awareness likely contributed to more targeted practice and improvement. These notions are going to be explicated as follows.

Narrative Insights from Participants

To complement the quantitative findings, a qualitative analysis was conducted based on semi-structured interviews with students from the experimental group. The purpose of this phase was to deepen the understanding of how learners experienced the use of the ELSA Speak application and how these experiences help explain the statistically significant improvements observed in oral performance.

It is important to note that all excerpts presented below correspond to translated versions of the original interviews, which were conducted in Spanish. To ensure confidentiality, participants are identified using alphanumeric codes (e.g., P1, P2, P3).

The analysis revealed three interrelated themes that provide insight into the mechanisms underlying the quantitative results.

Immediate Feedback as a Catalyst for Learning

One of the most consistently reported aspects was the value of immediate and individualized feedback. Students emphasized that the application allowed them to identify specific pronunciation errors in real time, something that was not always possible in regular classroom interactions.

As one participant explained:

“Before using the app, I didn’t really know what I was pronouncing wrong. The teacher corrected us, but not all the time. With the app, I could see exactly where I was making mistakes and try again until I improved” (P3, translated from Spanish).

Similarly, another student highlighted the iterative nature of the feedback process:

“I liked that I could repeat the words many times and the app kept correcting me. It felt like practicing without pressure, but still learning” (P7, translated from Spanish).

These narratives help explain the significant gains observed in pronunciation and fluency, as reported in the quantitative analysis. These facts align with formative assessment principles, where timely and actionable feedback plays a central role in supporting learning (Carless & Boud, 2021). Empirically, the immediacy of AI-generated feedback appears to have enabled students to engage in continuous cycles of correction and improvement.

Increased Confidence and Reduced Anxiety in Speaking

Another salient theme was the development of confidence in speaking English. Participants described the application as a space where they could practice without fear of making mistakes in front of others.

Participant P1 expressed this clearly:

“When I speak in class, I feel nervous because my classmates are listening. With the app, I could practice alone, and that helped me feel more confident later in class” (P1, translated from Spanish).

This perception was echoed by others:

“At first, I was afraid of speaking in English, but after using the app, I started to feel more secure because I knew I was improving little by little” (P5, translated from Spanish).

These accounts provide an important affective dimension to the interpretation of the results. While the quantitative data show improvement in oral performance, the narratives reveal that this progress is also linked to reduced anxiety and increased self-confidence. This is consistent with research in language learning that highlights the role of affective factors in speaking development, particularly in EFL contexts where opportunities for authentic interaction are limited.

Autonomy and Personalized Learning Experiences

A third theme that emerged from the analysis relates to learner autonomy. Students reported that the application allowed them to take control of their learning process, practicing at their own pace and focusing on areas they found challenging.

P9 participant noted:

“What I liked the most is that I could practice whenever I wanted. If I didn’t understand something, I could repeat it again and again until I got it right” (P9, translated from Spanish).

Another student (P2) emphasized the sense of independence fostered by the tool:

“It felt like I was learning by myself, but at the same time I had guidance from the app. It was different from just waiting for the teacher to correct me” (P2, translated from Spanish).

These experiences reflect key aspects of self-regulated learning, where learners actively monitor and adjust their performance. The autonomy described by participants helps explain the sustained engagement observed during the intervention and supports the idea that AI-based tools can foster more learner-centered environments.

Integrated Interpretation of Findings

The qualitative findings provide a nuanced explanation of the quantitative results, revealing that the observed improvements in oral skills were not only a product of increased practice, but also of changes in how students engaged with the learning process.

The immediacy of feedback facilitated error awareness and correction, the reduction of anxiety created a more favorable environment for speaking practice, and the flexibility of the tool promoted autonomous learning behaviors. Together, these elements contributed to a learning experience that was both cognitively and affectively supportive.

In this sense, the qualitative data reinforce the interpretation that the effectiveness of the ELSA Speak application lies not only in its technological capabilities, but in its ability to align with key pedagogical principles such as formative assessment, learner autonomy, and meaningful practice.

5. DISCUSSION

The results of this study provide robust evidence that the integration of the ELSA Speak application as a formative assessment tool can significantly enhance students' oral skills, particularly in pronunciation and fluency. Beyond statistical significance, the findings reveal how AI-mediated feedback reshapes the conditions under which speaking skills are practiced and improved in EFL contexts.

From a quantitative standpoint, the substantial gains observed in the experimental group, coupled with a large effect size, are consistent with recent research on AI-assisted language learning. Studies by Huang et al. (2022) and Aryanti and Santosa (2024) report comparable improvements in pronunciation accuracy when learners engage with speech recognition technologies that provide immediate corrective feedback. Similarly, Rahman et al. (2025) found that ELSA Speak contributes not only to measurable gains in pronunciation but also to increased learner engagement. The present study confirms these findings while extending them to a rural Ecuadorian context, where opportunities for individualized feedback are often limited. In such settings, the impact of AI tools appears to be amplified, as they compensate for structural constraints in traditional instruction.

However, the findings also invite a more nuanced interpretation when contrasted with studies that report more moderate effects. Li and Hafner (2022), for example, suggest that AI-based pronunciation tools do not necessarily outperform well-designed teacher-led instruction. This apparent divergence highlights the importance of context. In the present study, the control group, exposed to conventional feedback practices, showed improvement, but to a significantly lesser degree. This suggests that the effectiveness of AI is not inherent, but rather situational: it becomes particularly valuable in contexts where feedback is infrequent, delayed, or generalized.

Furthermore, the results strongly align with contemporary understandings of formative assessment. The experimental group benefited from continuous, immediate, and individualized feedback, which is widely recognized as a key driver of learning (Carless & Boud, 2021). The qualitative findings reinforce this interpretation, as students consistently described how real-

time correction helped them identify and address specific pronunciation errors. This iterative feedback loop appears to have facilitated more efficient learning processes, supporting the idea that feedback is most effective when it is timely, specific, and actionable.

The findings can also be interpreted through the lens of sociocultural theory, particularly the concept of mediation. In this study, the ELSA Speak application functioned as a mediational tool that scaffolded learners' oral production, enabling them to refine their performance through repeated interaction. This aligns with Lantolf and Thorne (2021), who emphasize that technological tools can extend learners' capabilities when they provide structured support. The improvement observed in fluency and pronunciation suggests that AI-based feedback can effectively guide learners within their zone of proximal development.

At the same time, the qualitative results highlight the role of affective factors, which are often underexplored in quantitatively driven studies. Participants reported increased confidence and reduced anxiety when practicing with the application, particularly because it allowed them to engage in speaking activities without fear of negative evaluation. This finding is consistent with Teimouri et al. (2019), who argue that lower anxiety levels are associated with better oral performance in EFL contexts. In this sense, the effectiveness of the intervention cannot be explained solely in cognitive terms; it also reflects the creation of a psychologically supportive learning environment.

Another key element emerging from the data is learner autonomy. Students described how the application enabled them to practice at their own pace, repeat tasks as needed, and monitor their progress independently. This aligns with recent research on self-regulated learning, which underscores the importance of learner control in fostering sustained engagement (Panadero, 2021). The ability to access feedback autonomously appears to have encouraged more consistent practice, which likely contributed to the observed improvements.

Nevertheless, the findings should be interpreted with caution. As noted by Chapelle and Voss (2022), the integration of AI in language assessment raises important questions regarding validity and pedagogical alignment. While the application proved effective as a formative tool, it does not replace the need for teacher-mediated assessment, particularly in evaluating communicative competence in more complex and interactive contexts.

In sum, this study contributes to the growing body of literature on AI in language education by demonstrating that its effectiveness is closely tied to contextual and pedagogical factors. In rural EFL settings, where traditional feedback mechanisms may be limited, AI-based tools such

as ELSA Speak offer a valuable complement to classroom instruction. Their impact, however, lies not only in their technological capabilities, but in their alignment with core principles of formative assessment, learner autonomy, and meaningful practice.

6. CONCLUSIONS

This study set out to examine the use of the ELSA Speak application as a formative assessment tool for the development of oral skills in a rural Ecuadorian EFL context. The findings demonstrate that its integration produces not only statistically significant improvements in pronunciation and fluency, but also meaningful changes in how students engage with speaking as a learning process.

Beyond the measurable gains, the study reveals that the effectiveness of the intervention lies in the convergence of three key elements: immediate and individualized feedback, reduced affective barriers, and increased learner autonomy. Together, these conditions create a learning environment in which students are not only able to practice more frequently, but also to do so with greater awareness, confidence, and purpose. In this sense, the application does not merely function as a technological tool, but as a pedagogical mediator that reshapes the dynamics of formative assessment.

In the same discursive line, the study underscores the potential of AI-based applications to complement traditional instruction, particularly in contexts where access to continuous and personalized feedback is limited. Rather than replacing the teacher, tools such as ELSA Speak can extend feedback beyond the classroom, enabling more sustained and individualized learning trajectories. This has direct implications for EFL teaching in rural settings, where innovation often depends on accessibility and adaptability.

At the same time, the findings call for a critical and balanced integration of technology. While the results are promising, the study is limited by its quasi-experimental scope, the relatively short duration of the intervention, and the specific contextual conditions in which it was conducted. Future research should explore longitudinal designs, include more diverse populations, and examine how AI-based formative assessment interacts with broader dimensions of communicative competence.

7. BIBLIOGRAPHY

- Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: Effects on accuracy and L2 motivation. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2025.2534015>
- Aryanti, R. D., & Santosa, M. H. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language*, 5(1), 102–113. <https://doi.org/10.36663/tatefl.v5i1.718>
- Chapelle, C. A., & Voss, E. (2022). Validity arguments for language assessment in technology-rich environments. *Language Learning & Technology*, 26(2), 1–15.
- He, M., Abbasi, B. N., & He, J. (2025). AI-driven language learning in higher education: An empirical study on self-reflection, creativity, anxiety, and emotional resilience in EFL learners. *Humanities and Social Sciences Communications*, 12, 1525. <https://doi.org/10.1057/s41599-025-05817-5>
- Kukulska-Hulme, A. (2020). Mobile-assisted language learning [Reprint]. In *The Routledge handbook of language learning and technology* (pp. 244–258). Routledge.
- Lai, Z. C. C. (2026). AI in EFL blended learning: Impact on speaking performance and learning resilience. *Asian-Pacific Journal of Second and Foreign Language Education*, 11(1), 11. <https://doi.org/10.1186/s40862-025-00367-4>
- Lantolf, J. P., & Thorne, S. L. (2021). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Li, S., & Hafner, C. A. (2022). Mobile-assisted language learning: A critical review of empirical studies. *ReCALL*, 34(1), 1–18. <https://doi.org/10.1017/S0958344021000240>
- Panadero, E. (2021). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 12, 422. <https://doi.org/10.3389/fpsyg.2021.00220>
- Rahimi, M., Fathi, J., & Zou, D. (2025). Exploring the impact of automated corrective feedback on EFL learners. *Education and Information Technologies*, 30, 2691–2735. <https://doi.org/10.1007/s10639-024-12896-5>

- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 41(2), 363–387.
<https://doi.org/10.1017/S0272263118000311>
- Winstone, N. E., & Boud, D. (2022). The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, 47(3), 656–667.
<https://doi.org/10.1080/03075079.2020.1779687>
- Zou, D., Huang, J., & Xie, H. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, 121, 103259.
<https://doi.org/10.1016/j.system.2024.103259>



Casa Editora del Polo (CASEDELPO), hace constar que:

El artículo científico:

“The Use of the ELSA Speak App for Formative Assessment of Oral Skills in Ecuadorian English Language Education.”

De autoría:

**Armijos Castro Nahira Allyson, Beltrán Marín Andrea Victoria, Buenaño Lema Nicole
Mayte, Zambrano Pachay Jorge Francisco**

Habiéndose procedido a su revisión y analizados los criterios de evaluación realizados por lectores pares expertos (externos) vinculados al área de experticia del artículo presentado, ajustándose el mismo a las normas que comprenden el proceso editorial, se da por aceptado la publicación en el **Vol. 11, No 5, Mayo 2026**, de la revista Polo del Conocimiento, con ISSN 2550-682X, indexada y registrada en las siguientes bases de datos y repositorios: **Latindex Catálogo v2.0, MIAR, Google Académico, ROAD, Dialnet, ERIHPLUS.**

Y para que así conste, firmo la presente en la ciudad de Manta, a los 12 días del mes de abril del año 2026.

Dr. Víctor R. Jama Zambrano
DIRECTOR

UNEMI
UNIVERSIDAD ESTATAL DE MILAGRO

¡Evolución académica!

@UNEMIEcuador

