Analysis of the Impact of Pictogram-Based Instruction on EFL Speaking Skills Development

Carmen Benitez-Correa[1], Elsa Morocho-Cuenca[1] & Ana Quinonez-Beltran[1]

[1] Universidad Técnica Particular de Loja, Loja, Ecuador

Correspondence: Carmen Benítez-Correa, Universidad Técnica Particular de Loja, Loja, Ecuador, San Cayetano Alto, C. París (07) 370-1444. E-mail: cdbenitez@utpl.edu.ec

Received: June 3, 2026 Accepted: August 29, 2026 Online Published: September 23, 2026

doi:10.5430/wjel.v16n6p571 URL: https://doi.org/10.5430/wjel.v16n6p571

Abstract

This study examined the effect of pictogram-based instruction on the development of speaking skills in primary EFL students, with specific attention to pronunciation and word stress. A quasi-experimental pre-test/post-test design within a mixed methods approach was employed, involving a control group and an experimental group. The participants were 203 EFL students, distributed into a control group (n = 99) and an experimental group (n = 104). The experimental group received pictogram-based guided speaking instruction, while the control group followed conventional instruction. Speaking performance was assessed pre- and post-intervention using parallel oral tasks and an analytic rubric focusing on pronunciation and word stress. Descriptive statistics, baseline comparisons, paired comparisons, gain-score analyses, confidence intervals, and effect sizes were calculated to determine the impact of the intervention. Interview data from a randomly selected subgroup of students in the experimental group were analyzed and integrated with the quantitative findings to explain students' perceptions of how pictograms supported pronunciation and word stress. The results showed that both groups improved from pre-test to post-test; however, the experimental group achieved significantly greater gains in pronunciation, word stress, and overall speaking performance. Pronunciation showed a large effect size, while word stress demonstrated an even stronger improvement, with a very large effect size. Within the experimental group, word stress improved significantly more than pronunciation. These findings suggest that pictograms can serve as effective visual scaffolds for oral language development, particularly in supporting suprasegmental features such as stress, rhythm, and comprehensibility.

Keywords: EFL students, pictograms, pronunciation, speaking skills, word stress

1. Introduction

Speaking is a central component of English as a Foreign Language (EFL) learning because effective oral communication requires learners to integrate linguistic, discourse, and pronunciation-related resources in meaningful interaction (Goh & Burns, 2012; Hughes & Szczepek Reed, 2017). Within speaking performance, pronunciation and word stress are two important components because they influence how clearly and comprehensibly learners communicate in English. Pronunciation is closely related to the clear production of sounds and words, while word stress contributes to rhythm, emphasis, and comprehensibility in spoken language (Levis, 2018; Uzun, 2024). For this reason, both components are especially relevant in EFL contexts, where learners often have limited opportunities to develop accurate and natural oral production.

Despite its importance, pronunciation instruction continues to receive limited and inconsistent attention in many EFL classrooms. Recent studies show that pronunciation is considered essential for effective communication, but it is not always systematically integrated into classroom practice (Almusharraf, 2024; Metruk, 2025). This issue is particularly relevant for primary EFL students, who are still developing the phonological awareness, oral confidence, and production control needed to communicate clearly in English. In addition, pronunciation teaching should not be restricted to isolated sound production because oral comprehensibility also depends on prosodic features. In this regard, Alghazo et al. (2023) emphasize the importance of addressing both segmental and suprasegmental aspects of pronunciation instruction, including stress placement.

Within this broader view of oral development, word stress deserves explicit pedagogical attention. Inaccurate stress placement can affect rhythm and make spoken messages less comprehensible, especially when learners transfer stress patterns from their first language to English. Therefore, word stress should be treated as a relevant component of speaking performance rather than as a peripheral feature. However, young EFL learners often receive more opportunities to repeat vocabulary than to notice, practice, and produce stress patterns in meaningful oral contexts. This creates the need for instructional strategies that help learners connect spoken forms, meaning, rhythm, and visual support during oral practice.

Considering these issues, multimodal and visual-based approaches offer a relevant alternative for supporting EFL speaking instruction. Multimodal instruction allows teachers to combine verbal, visual, auditory, and gestural resources, making language input more accessible and meaningful for learners. In EFL classrooms, the combination of spoken language with other modes of communication,

such as gesture, gaze, facial expression, and visual cues, can support engagement and understanding during classroom interaction (Qin & Wang, 2021). Panyathikul et al. (2025) also suggest that multimodal teaching can improve EFL learners' pronunciation when audiovisual resources and gestures are used to guide oral production.

Visual materials are particularly relevant because they help learners connect verbal information with concrete representations. Studies on visual and multimedia-supported vocabulary learning have shown that combining verbal input with images, infographics, or audiovisual support can facilitate vocabulary learning, retention, and learner engagement (Alwadei & Mohsen, 2023; Teng, 2023; Suwanasilp & Durongbhandhu, 2023). Although these studies focus mainly on vocabulary, their findings are relevant to speaking instruction because oral production depends partly on learners' ability to retrieve words, recognize their spoken forms, and use them in context. Thus, visual support may help learners move beyond word recognition toward oral use, especially when vocabulary is practiced in meaningful speaking tasks.

Within this group of visual materials, pictograms stand out as simple representations that connect meaning with language forms in an accessible way (Bühler, 2021). Pictograms were selected for this study because, unlike more general picture cards or flashcards, they can be systematically designed to represent both semantic meaning and specific oral cues, such as syllable segmentation, sound duration, mouth position, prominence, or rhythm. They also require fewer technological resources than multimedia materials and can be easily embedded in short oral tasks with young learners. This gives them a pedagogical advantage in primary EFL classrooms, where instruction often needs to be concrete, low-cost, replicable, and cognitively accessible (Mayer, 2021; Paivio, 2007; Sweller et al., 2019).

Previous studies have recognized the value of pictograms and other visual scaffolds in primary EFL contexts, particularly for vocabulary learning, engagement, motivation, reading comprehension, and language production (Abdelaal, 2025; (Pauzhi Coellar & Argudo-Serrano, 2024; Medina & Romero, 2025). However, much of the research on visual support has examined vocabulary or reading outcomes, whereas speaking-focused studies have tended to use broader multimodal resources, such as videos, gestures, or audiovisual input. Consequently, there is limited empirical evidence on pictogram-based instruction as the central pedagogical intervention for improving specific oral components, especially pronunciation and word stress among primary EFL learners.

This reveals an important research gap. While previous studies have provided evidence on the value of pronunciation instruction, multimodal input, and visual support in EFL learning, less is known about how pictograms influence specific dimensions of speaking performance. In particular, few empirical studies have examined whether pictograms contribute differently to pronunciation and word stress development in young learners. This distinction is relevant because these components may respond differently to visual scaffolding. Pictograms may support pronunciation by linking images with oral forms and articulatory cues, whereas their effect on word stress may depend on modelling, repetition, stress highlighting, and corrective feedback.

To address this gap, the present study examines the role of pictogram-based instruction in enhancing speaking skills among primary EFL students, with particular attention to pronunciation and word stress. By focusing on these two components of speaking performance, the study seeks to provide evidence on the pedagogical value of pictograms as visual scaffolds for oral language development. Accordingly, the study was guided by the following research questions and hypotheses:

RQ1. How do pictograms influence the development of pronunciation as an element of speaking skills in young EFL students?

RQ2. How do pictograms contribute to the development of word stress as an element of speaking skills in young EFL students?

RQ3. Which speaking component, pronunciation or word stress, shows greater improvement through the use of pictograms in young EFL students?

H1. Students who receive pictogram-based instruction will show significantly greater gains in pronunciation than students who receive conventional instruction.

H2. Students who receive pictogram-based instruction will show significantly greater gains in word stress than students who receive conventional instruction.

H3. Within the experimental group, word stress will show a significantly greater gain than pronunciation after pictogram-based instruction.

2. Literature Review

2.1 Speaking Skills in EFL: Definition, Importance, and Challenges

Speaking is widely recognized as a foundational pillar of communicative competence in second and foreign language learning. It involves the real-time production of meaningful oral discourse, requiring the integration of linguistic, sociolinguistic, and strategic competencies (Brown & Abeywickrama, 2019). Within English as a Foreign Language (EFL) context, speaking is not merely a productive skill but a highly demanding cognitive process that obliges learners to coordinate pronunciation, vocabulary, grammatical structures, and fluency simultaneously under intense time constraints. According to the Council of Europe (2020), spoken production at lower proficiency levels (A1–A2) is inherently characterized by an emerging and developing control of phonological features, where establishing basic intelligibility serves as the primary criterion for communicative success.

The overall significance of oral proficiency lies in its direct connection to real-life communication, social interaction, and broader academic achievement. As Lightbown & Spada (2021) emphasize, learners who develop strong oral proficiency are better equipped to

engage in collaborative tasks, negotiate meaning, and express complex ideas effectively. Furthermore, speaking competence is closely associated with positive affective variables; as Dörnyei (2020) argues, the development of oral skills fosters increased motivation and self-efficacy, since learners can perceive immediate and tangible outcomes from their ability to interact with others.

Despite its well-established pedagogical value, the development of speaking skills presents multifaceted challenges, particularly in formal EFL contexts. One major constraint is learners' limited exposure to authentic target language input beyond the classroom, which significantly reduces opportunities for meaningful and spontaneous oral practice, as noted by Richards (2015) and Windriya (2025). For young EFL learners, this limitation is further exacerbated by developmental constraints in working memory, making the simultaneous processing of vocabulary and syntax especially demanding. As a result, segmental pronunciation accuracy and suprasegmental features such as word stress often become problematic, directly affecting intelligibility and comprehensibility in early stages of language acquisition, as highlighted by Derwing & Munro (2021).

These cognitive and linguistic challenges are further intensified by affective barriers. In particular, Dewaele & MacIntyre (2019) identify foreign language speaking anxiety, fear of negative evaluation, and reduced self-confidence as key factors that significantly limit learners' willingness to communicate. Within this complex framework, the implementation of cognitively grounded pedagogical strategies that explicitly scaffold pronunciation while reducing affective barriers becomes essential for fostering effective oral communication.

2.2 Elements of Speaking Skills

Speaking ability can be understood as a multidimensional construct that encompasses several interdependent components necessary for effective oral communication, as noted by Brown and Abeywickrama (2019). In this regard, Richards (2015) explains that successful spoken communication relies on the integration of both linguistic and discourse-related elements, such as fluency, grammatical accuracy, lexical range, pronunciation, and suprasegmental features. As Derwing and Munro (2015) point out, fluency involves the capacity to produce speech smoothly and at an appropriate pace, whereas accuracy refers to the correct use of grammatical forms and vocabulary. Additionally, lexical competence allows learners to convey meaning effectively, while accurate segmental pronunciation plays a crucial role in maintaining intelligibility, even in the presence of first-language influence (Flege & Bohn, 2021). Furthermore, suprasegmental features, such as word stress prominence, accurate stress placement, and intonation, play a crucial role in maintaining linguistic rhythm and enhancing overall comprehensibility (Celce-Murcia et al., 2010; Field, 2005). Recent empirical evidence confirms that these phonetic and prosodic dimensions remain primary predictors of communicative success in second language performance (Huensch & Nagle, 2021).

These components do not operate in isolation; rather, they interact dynamically during real-time communication (Bachman & Palmer, 2010). However, at lower levels of proficiency, learners tend to rely more heavily on phonological control to ensure intelligibility, as constraints in grammatical and lexical knowledge can be partially offset when speech is clearly articulated (Council of Europe, 2020). In line with this perspective, Isaacs and Trofimovich (2012) argue that during the early stages of language acquisition, phonological features, particularly suprasegmental aspects, play a central role in shaping comprehensibility, especially when other linguistic resources remain limited.

2.3 Pronunciation in EFL

Pronunciation in English as a Foreign Language (EFL) refers to the ability to produce speech sounds in a way that is intelligible to listeners, encompassing both segmental features and aspects of phonological control that affect overall clarity (Brown, 2019). Rather than aiming for native-like accuracy, contemporary research emphasizes intelligibility and comprehensibility as the primary goals of pronunciation instruction (Derwing & Munro, 2015; Saito, 2021). This perspective is reflected by the Council of Europe (2020), which highlights that learners at the A1–A2 levels demonstrate emerging control of phonological features; a stage where being understood is far more critical than producing error-free speech. Recent literature continues to reinforce the centrality of comprehensibility in pronunciation development, confirming that a listener's ease of understanding is a more realistic and pedagogically relevant goal than native-like acoustic replication (Isaacs & Trofimovich, 2012).

To fully capture the complexity of pronunciation within these learning contexts, contemporary applied linguistics emphasizes the necessity of dissecting oral production into three distinct yet interrelated dimensions:

2.3.1 Sound Articulation Accuracy

According to Brown (2019), this dimension concerns the learner's physical ability to manipulate speech organs, such as the lips, tongue, and jaw, to accurately form individual English phonemes, including both vowel contrasts and consonants. While modern pedagogical frameworks no longer demand native-like perfection, assessing articulation accuracy remains vital, as systematic distortions or omissions of target sounds create a cumulative cognitive burden on oral clarity, making it more difficult for interlocutors to decode the message (Kenworthy, 1987; Suzukida & Saito, 2021).

2.3.2 Intelligibility

Intelligibility is defined as the degree to which a listener can understand a speaker's intended message (Derwing & Munro, 2015). Contemporary EFL teaching models treat intelligibility as the primary benchmark of communicative success. Even when learners exhibit limitations in vocabulary or grammar, establishing a baseline of phonological intelligibility serves as a critical communicative threshold that prevents breakdowns in interaction (Saito, 2021).

2.3.3 First Language Interference

L1 interference refers to deviations in L2 pronunciation caused by the transfer of phonological categories and articulatory habits from the native language into English (Flege, 1995). Learners frequently substitute unfamiliar English phonemes such as /θ/ or /ð/ with the closest sounds available in their L1 inventory. This phenomenon is central to understanding persistent pronunciation difficulties and their effects on comprehensibility (Suzukida & Saito, 2019).

The overarching importance of these dimensions lies in their direct impact on successful communication. Even when learners possess adequate vocabulary and grammatical knowledge, inadequate pronunciation can result in misunderstandings and communication breakdowns (Saito, 2021). In this sense, pronunciation functions as a gateway skill that ensures spoken messages are accessible to interlocutors. Furthermore, improved pronunciation has been linked to increased learner confidence and a greater willingness to communicate, both of which are key factors in second language acquisition (Derwing & Munro, 2015; MacIntyre et al., 1998; Dörnyei, 2005).

However, pronunciation remains one of the most challenging areas for EFL learners. These difficulties often stem from structural differences between the phonological systems of the first language and English, which lead to persistent sound substitutions, omissions, and distortions (Kenworthy, 1987; Derwing & Munro, 2015). In addition, limited exposure to authentic spoken input and insufficient opportunities for practice in classroom settings further constrain phonological development (Richards, 2022). Taken together, these challenges underscore the need for explicit instructional approaches that provide clear articulatory models alongside meaningful communicative practice.

2.4 Word Stress in EFL

Word stress, a fundamental suprasegmental feature of English, refers to the relative emphasis placed on a syllable within a word, making it more prominent than neighboring syllables through a combination of increased loudness, duration, and pitch variation (Roach, 2009). Within EFL context, mastering word stress is critical, as prosodic features significantly influence speech comprehensibility and listener processing efficiency (Yenkimaleki & van Heuven, 2021). This perspective aligns with the Council of Europe (2020), which classifies stress under phonological control, noting that although basic-level learners may demonstrate emerging awareness, accurate stress placement remains highly developmental. To examine word stress in EFL oral production, contemporary research distinguishes among stress placement accuracy, acoustic prominence, and its effects on rhythm and comprehensibility:

2.4.1 Stress Placement Accuracy

This dimension evaluates whether learners assign primary lexical stress to the correct syllable according to English phonological norms. Stress placement in English is highly variable, and incorrect placement can significantly hinder word recognition by disrupting the listener's lexical access, even when segmental pronunciation is accurate (Derwing & Munro, 2015).

2.4.2 Stress Prominence

Stress prominence refers to the acoustic salience of the stressed syllable, achieved through pitch movement, vowel lengthening, and intensity (Roach, 2009). A common difficulty for EFL learners is insufficient vowel reduction in unstressed syllables, which weakens prominence contrasts and reduces perceptual clarity.

2.4.3 Effect on Rhythm and Comprehensibility

Word stress contributes to the stress-timed rhythm of English, characterized by regular recurrence of stressed syllables (Celce-Murcia et al., 2010). When learners assign equal prominence to all syllables, the rhythm becomes syllable-timed, increasing listener processing effort and reducing overall comprehensibility.

Despite its communicative importance, word stress remains one of the most difficult suprasegmental features for EFL learners. Difficulties are often attributed to L1 interference, where learners transfer native stress rules and rhythmic patterns into English (Kenworthy, 1987; Derwing & Munro, 2015). Additionally, traditional instruction often prioritizes grammar and vocabulary over prosodic features, limiting learners' opportunities to develop awareness of stress patterns (Kenworthy, 1987; Celce-Murcia et al., 2010). This highlights the need for explicit pedagogical intervention focused on both the perception and production of word stress (Celce-Murcia et al., 2010; Saito, 2021).

2.5 The Role of Pictograms in L2 Pronunciation Development

Pictograms are simplified visual representations that facilitate language comprehension and production by linking abstract linguistic input with immediate visual cues (Pauzhi & Argudo-Serrano, 2024; Wang et al., 2022; Bühler, (2021). Within language education, the pedagogical efficacy of these visual tools is anchored in multimodal learning principles, which emphasize the integration of verbal and non-verbal modes to optimize information processing. In this regard, Mayer (2021) demonstrates through his Cognitive Theory of Multimedia Learning that learners construct more robust mental representations when information is presented simultaneously through visual and auditory channels, rather than a single modality. This multimodal synergy fundamentally reduces extraneous cognitive load and enhances long-term retention. From a semiotic perspective, Kress and van Leeuwen (2020) further argue that these visual representations serve as dynamic resources through which learners actively construct and internalize meaning in foreign language environments.

Beyond general comprehension, the role of pictograms in targeted pronunciation instruction can be understood through Dual Coding

Theory. As Paivio (2007) argues, information encoded through parallel and independent verbal and non-verbal systems generates additive memory traces that enhance cognitive recall. In EFL contexts, this dual activation functions as essential scaffolding, strengthening form–meaning connections during oral production tasks (Schmitt & Schmitt, 2020). When learners encounter a target word accompanied by a pictogram, the visual representation acts as an anchor that stabilizes conceptual meaning. This stabilization, in turn, reduces demands on working memory, enabling learners to focus more effectively on the physical processes of speech production, including accurate sound articulation and the reduction of first-language (L1) phonological interference (Derwing & Munro, 2015).

Drawing from Cognitive Load Theory, Sweller (2024) emphasizes that instructional design must minimize split-attention effects to maximize learning efficiency. Integrating pictograms directly mitigates the real-time cognitive demands of spontaneous speech processing. By providing an instantaneous, non-linguistic representation of meaning, pictograms eliminate mental translation from the L1, bypass cross-linguistic lexical competition, and reduce phonetic distortions. Therefore, pictograms function as cognitively grounded resources that systematically support L2 phonological development.

2.6 Pictograms in EFL Speaking: Implications for Pronunciation and Word Stress

Beyond their general pedagogical value, pictograms hold explicit potential for supporting the development of productive speaking skills, particularly within the domains of segmental pronunciation and suprasegmental word stress. In pronunciation acquisition, targeted visual representations serve as critical cognitive anchors, enabling learners to map abstract phonetic categories onto salient imagery, thereby reinforcing the encoding of stable phonological forms. Derwing and Munro (2015) emphasize that this multimodal scaffolding is highly valuable at lower proficiency levels, where developing learners rely heavily on concrete visual representations to physically decode, process, and accurately articulate unfamiliar target sounds.

Regarding suprasegmental development, pictograms can be strategically engineered to heighten learners' phonological awareness of syllable prominence by visually segmenting or formatting linguistic strings. According to Zúñiga Vargas (2021), by manipulating the visual salience of a pictogram, such as its scale, positioning, or color weight, instructors can provide an instantaneous cue that mirrors the acoustic dimensions of loudness, pitch, and duration. This explicit visual scaffolding directly facilitates stress placement accuracy and accentuates stress prominence, both of which are recognized by the Council of Europe (2020) as foundational prerequisites for maintaining the characteristic rhythm and general intelligibility of spoken English. Given that English lexical stress patterns are notoriously unpredictable, utilizing pictograms as dual-coded anchors supports both the real-time recognition and long-term cognitive retention of correct stress mapping (Ivanova, 2024).

Furthermore, the integration of pictograms during oral production tasks directly enhances learners' affective variables, specifically boosting confidence and willingness to communicate (WTC). By providing immediate, non-linguistic semantic cues, pictograms minimize the real-time cognitive demands of lexical retrieval during speech (Mayer, 2021; Schmitt & Schmitt, 2020). As Dewaele and MacIntyre (2019) demonstrate, L2 learners are significantly more likely to actively engage in oral production tasks when instructional materials alleviate communicative anxiety and lower the psychological barriers associated with spontaneous speech. Consequently, pictograms transcend basic comprehension support; they function as a dynamic pedagogical mechanism that actively drives L2 phonological development, offering a highly viable tool for optimizing both sound articulation and prosodic accuracy in EFL environments.

The following framework links the intervention to learning mechanisms and expected results

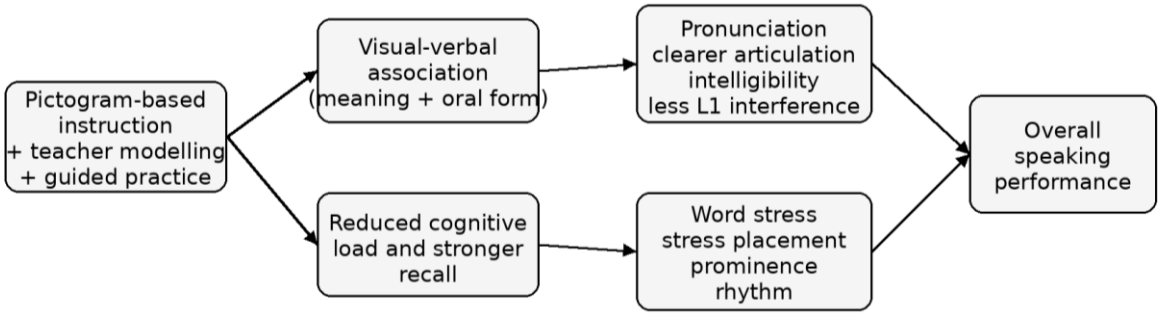

Figure 1. Conceptual framework of pictogram-based speaking instruction

This conceptual framework assumes that pictogram-based instruction influences speaking performance through two main mechanisms: visual-verbal association and reduced cognitive load. Visual-verbal association helps learners connect meanings with oral forms, while reduced cognitive load allows them to focus attention on articulation, stress placement, and rhythm during guided speaking practice. These mechanisms are expected to support pronunciation, word stress, and overall speaking performance.

3. Method

3.1 Research Design

This study employed a quasi-experimental pre-test/post-test design embedded in a quantitative-dominant mixed-methods approach. The

quantitative component was used to determine changes in pronunciation, word stress, and total speaking performance, whereas the qualitative component was used to explain students' perceptions of how pictograms supported these changes. The design was quasi-experimental because students remained in their existing class sections for institutional, ethical, and organizational reasons, moving minors across classrooms would have disrupted regular school schedules and instructional continuity. This decision is consistent with educational intervention research in natural classroom settings (Shadish et al., 2002). The quasi-experimental design was therefore used to test the three research hypotheses by comparing pronunciation and word-stress gains between the control and experimental groups, as well as by comparing both speaking components within the experimental group.

Because students were not randomly assigned, possible threats to internal validity included selection effects, maturation, testing effects, and teacher-related variability. Several procedures were used to reduce these risks: both groups studied the same curricular topics during the same period, completed parallel pre-test and post-test tasks, were assessed with the same analytic rubric, and were statistically compared at baseline before the intervention. The absence of random assignment is acknowledged as a design limitation, but the use of pre-test equivalence checks and gain-score comparisons strengthened the interpretation of the intervention effects.

3.2 Participants

The study included 203 primary students aged 11–12 years enrolled in seventh grade in public schools in Ecuador. The schools followed the national English curriculum and offered three 40-minute English classes per week. Participants were native Spanish speakers and were classified at the A1 level according to the Common European Framework of Reference for Languages (Council of Europe, 2020). Inclusion criteria were enrolment in the participating classes, parental consent, student assent, regular participation in the intervention period, and completion of both the oral pre-test and post-test.

Participants belonged to two pre-existing class sections, referred to as intact groups because the researcher did not create new groups or reorganize students for the purpose of the study. One section was assigned as the control group (n = 99; 52 females and 47 males), and the other as the experimental group (n = 104; 59 females and 45 males). The sample was selected through convenience sampling based on school access and institutional authorization. Although this sampling strategy may limit generalizability, it was appropriate for conducting a classroom-based intervention under authentic educational conditions (Etikan et al., 2016; Creswell & Creswell, 2018).

Baseline equivalence was statistically verified before examining intervention effects. No significant pre-test differences were found between the control and experimental groups in pronunciation, word stress, or overall speaking performance, indicating that the groups were comparable at the start of the study.

Ethical considerations were followed throughout the study. Informed consent was obtained from parents or legal guardians, and assent was obtained from the students. Participation was voluntary, and students were informed that they could withdraw from the study without academic penalty or negative consequences for their regular English class. All data were collected and stored anonymously to protect confidentiality.

3.3 Instruments

Oral pre-test and post-test were administered to all participants in the control and experimental groups (N = 203) to assess students' speaking performance before and after the intervention. The oral tests were developed as parallel forms with the same format, number of prompts, scoring criteria, and a comparable level of difficulty, ensuring consistency between pre- and post-test assessment. They elicited oral responses related to the curricular topics taught during the intervention: time, prices, and daily activities. These topics were selected because they were age-appropriate, aligned with the national curriculum, and suitable for eliciting target vocabulary in short communicative responses for assessing the two phonological variables.

Speaking performance was evaluated using an analytic rubric with two dimensions: pronunciation and word stress. Pronunciation was assessed through sound articulation, intelligibility, and first-language interference. Word stress was assessed through stress placement accuracy, stress prominence, and the effect of stress on rhythm and comprehensibility. These indicators were selected because they represent key components of comprehensible speech in EFL contexts (Derwing and Munro 2015; Council of Europe, 2020). Each dimension was rated on a 0–5 scale, producing a maximum total speaking score of 10 points. The rubric is summarized in Appendix A and can be included as supplementary material in its full version.

Instrument validity was supported through expert review and pilot testing. Four EFL specialists reviewed the oral prompts and rubric for clarity, relevance, age appropriateness, and alignment with the research objectives. Their feedback was used to refine wording, scoring criteria, and administration procedures. A pilot study with 11 students outside the main sample was then conducted to test task clarity and timing. Reliability analysis yielded Cronbach's alpha = .80, indicating acceptable internal consistency for the rubric-based scores. Formal inter-rater reliability statistics could not be calculated because the original scoring procedure did not include independent double scoring by two raters; this has been acknowledged in the limitations, and future work should include independent raters and report agreement indices such as ICC.

To complement the quantitative data, face-to-face interviews with six open-ended questions were conducted with a randomly selected subgroup of students from the experimental group (n = 20). The questions explored how students perceived the contribution of pictograms to pronunciation, word stress, and overall oral production. Interviews were conducted in Spanish, the students' first language, using age-appropriate wording. Selected excerpts were later translated into English for reporting.

3.4 Intervention and Procedure

The intervention lasted five months and was implemented within the regular English class schedule. Both groups studied the same curricular topics and completed comparable oral practice tasks. The control group followed conventional instruction, which consisted of teacher explanation, vocabulary presentation, oral repetition, textbook-based practice, and short teacher-led speaking activities without systematic use of pictograms. The experimental group received the same curricular content, but pictograms were systematically integrated as visual scaffolds during modelling, repetition, oral drills, feedback, and contextualized speaking tasks.

Each pictogram-based lesson followed a consistent sequence during the 40-minute class period. First, the teacher introduced target words or short expressions using pictograms linked to the lesson topic. Second, the teacher modelled pronunciation and word stress while drawing attention to visual cues, such as mouth images for articulation, divided words for syllable structure, stars or color marks for stressed syllables, and size contrasts for prominence. Third, students completed guided repetition, word identification, short oral drills, and controlled speaking practice. Finally, they used the target vocabulary in pair or individual oral tasks in meaningful contexts, such as asking and answering about time, prices, or daily routines.

The teacher used the same instructional sequence across lessons to improve implementation consistency. No separate teacher-training program was implemented; instead, the lesson sequence, target vocabulary, pictogram set, feedback focus, and scoring criteria were reviewed before the intervention so that classroom delivery remained aligned with the study objectives. Corrective feedback was provided during practice, particularly on sound articulation, intelligibility, stress placement, rhythm, and comprehensibility.

Table 1. Examples of pictogram-based instructional procedures used in the experimental group

Instructional focus	Example pictogram cue	Sample classroom activity	Expected oral outcome
Pronunciation	Mouth picture; divided word (ap-ple).	Students repeat the word, identify the sound, and produce it in a short phrase.	Clearer articulation and intelligibility.
Word stress	Star, red syllable, big/small circles, clapping hand.	Students mark the stronger syllable, clap the rhythm, and say the word in a short response.	More accurate stress placement and rhythm.
Contextualized speaking	Pictograms for time, prices, and daily activities.	Students use visual prompts to ask and answer short questions in pairs.	Oral production beyond isolated word repetition.

Pictogram cues were used as visual scaffolds to support pronunciation, word stress, and contextualized speaking activities.

Prior to the intervention, an oral pre-test was administered to both groups. A group of ten evaluators with experience as EFL teachers and in listening and speaking assessment supported the administration of the oral tests, and students' responses were audio-recorded for subsequent scoring. After the intervention, an oral post-test was administered under the same conditions. The same analytic rubric was used for both testing phases, making it possible to compare students' pre-test and post-test performance in pronunciation, word stress, and total speaking performance.

3.5 Data Analysis

The quantitative data obtained from the pre-test and post-test were analyzed using IBM SPSS Statistics v25. Descriptive statistics were calculated for pronunciation, word stress, and total speaking scores. Independent-samples t-tests were used to verify baseline equivalence between the control and experimental groups at pre-test. Paired-samples t-tests were used to examine within-group pre-test/post-test changes, and Welch's t-tests on gain scores were used to compare improvement between groups because this procedure is robust when group sizes or variances are not identical. Effect sizes were reported using Cohen's dz for within-group changes and Hedges' g for between-group gain comparisons. Confidence intervals were included to support interpretation, and the significance level was set at $\alpha = .05$.

Assumption testing was conducted before interpreting parametric comparisons. Shapiro-Wilk tests were used to examine gain-score normality, and Levene's tests were used to examine homogeneity of variance. Although some Shapiro-Wilk tests indicated deviations from normality, parametric procedures were retained because the group sizes were large and Welch's t-tests are robust under moderate deviations from normality. Levene's tests did not indicate significant heterogeneity of gain-score variance for the main outcomes.

Interview data were first examined through reflexive thematic analysis following Braun and Clarke's (2021) approach in order to identify patterns in students' explanations of how pictograms supported pronunciation and word stress. The qualitative findings were then integrated with the quantitative results through an explanatory mixed-methods joint display (Guetterman et al., 2015). This procedure allowed students' perceptions to be connected with the observed pre-test and post-test gains rather than being presented as isolated thematic categories.

4. Results

Before addressing the research questions, baseline equivalence between the control and experimental groups was examined. The results showed no significant pre-test differences between groups in pronunciation, $t(200.92) = 1.08$, $p = .283$, word stress, $t(200.76) = -1.27$, $p = .206$, or overall speaking score, $t(189.49) = -0.10$, $p = .918$. These results indicate that both groups started from comparable levels of oral performance.

Assumption checks were also examined before interpreting the inferential analyses. As shown in Table 2, gain-score normality was not

fully met for all outcomes; however, Levene's tests showed no significant heterogeneity of variance. Given the sample size and the use of Welch's t-tests for between-group comparisons, the analyses were considered appropriate for evaluating intervention effects.

Table 2. Preliminary statistical checks for baseline equivalence and gain-score analyses

Check	Outcome	Statistic	p	Interpretation
Baseline equivalence	Pronunciation pre-test	t(200.92) = 1.08	.283	Groups comparable
Baseline equivalence	Word stress pre-test	t(200.76) = -1.27	.206	Groups comparable
Baseline equivalence	Total speaking pre-test	t(189.49) = -0.10	.918	Groups comparable
Normality	Pronunciation gain (control/experimental)	W = .904 / .948	< .001 / < .001	Deviations noted; robust procedures retained
Normality	Word stress gain (control/experimental)	W = .917 / .979	< .001 / .100	Mostly acceptable
Normality	Total speaking gain (control/experimental)	W = .947 / .973	< .001 / .034	Deviations noted; robust procedures retained
Homogeneity	Pronunciation gain	Levene F = 1.761	.186	Variance assumption met
Homogeneity	Word stress gain	Levene F = 1.747	.188	Variance assumption met
Homogeneity	Total speaking gain	Levene F = 3.415	.066	Variance assumption met

Note. Baseline equivalence was examined using independent-samples comparisons at pre-test. W = Shapiro-Wilk statistics. Levene's tests were used for homogeneity of variance.

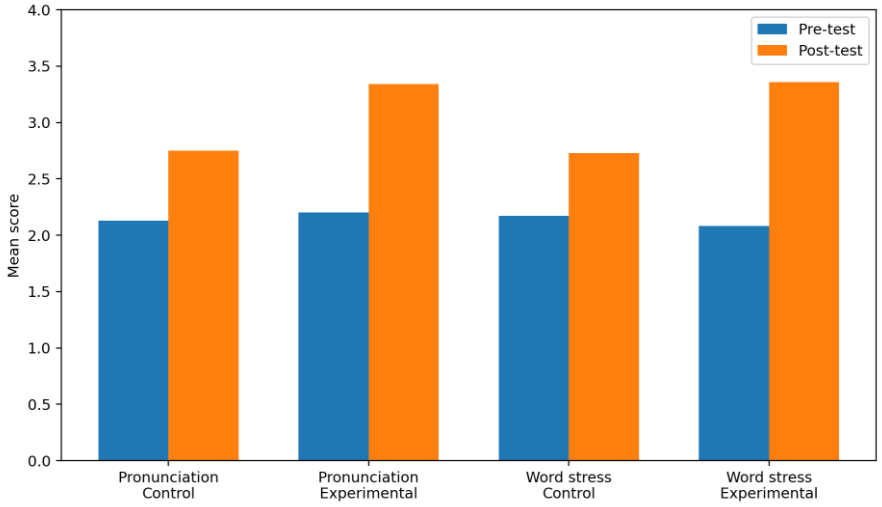

Figure 2. Pre-test and post-test mean scores by group and speaking component

RQ1. How do pictograms influence the development of pronunciation as an element of speaking skills in primary EFL students?

To answer the first research question, pronunciation scores were compared from pre-test to post-test in both groups. Table 3 shows that both groups improved after the instructional period; however, the experimental group showed a substantially larger gain than the control group.

Table 3. Pre-test, post-test, and gain scores for pronunciation

Group	Pre-test M (SD)	Post-test M (SD)	Mean gain	Gain 95% CI	Gain %
Control	2.13 (0.49)	2.75 (0.45)	0.62	[0.56, 0.69]	29.3%
Experimental	2.20 (0.51)	3.34 (0.58)	1.14	[1.06, 1.22]	51.8%

Note. Mean gain was calculated as post-test minus pre-test. CI = confidence interval.

The gain-score comparison confirmed that students who received pictogram-based instruction improved their pronunciation significantly more than those who received conventional instruction, Δgain = 0.52, 95% CI [0.41, 0.62], t(197.21) = 9.90, p < .001, Hedges' g = 1.38. This large effect indicates that pictogram-based instruction had a strong positive influence on pronunciation development. Therefore, H1 was supported.

RQ2. How do pictograms contribute to the development of word stress as an element of speaking skills in primary EFL students?

To answer the second research question, word-stress scores were analyzed across the control and experimental groups. As shown in Table 4, both groups improved, but the experimental group showed a markedly greater increase.

Table 4. Pre-test, post-test, and gain scores for word stress.

Group	Pre-test M (SD)	Post-test M (SD)	Mean gain	Gain 95% CI	Gain %
Control	2.17 (0.47)	2.73 (0.46)	0.57	[0.51, 0.63]	26.2%
Experimental	2.08 (0.51)	3.36 (0.59)	1.28	[1.22, 1.35]	61.8%

Note. Mean gain was calculated as post-test minus pre-test. CI = confidence interval.

The comparison of gain scores confirmed that the improvement in the experimental group was significantly greater than that of the control group, Δgain = 0.72, 95% CI [0.63, 0.80], t(201.00) = 16.49, p < .001, Hedges' g = 2.30. This very large effect indicates that pictograms made a strong contribution to the development of word stress, particularly by helping students recognize, internalize, and produce stress patterns during oral tasks. Therefore, H2 was supported.

To provide a broader view of the intervention effect, Table 5 summarizes the between-group gain-score comparisons for pronunciation, word stress, and total speaking performance.

Table 5. Between-group comparison of gain scores

Outcome	Experimental gain M	Control gain M	Δgain	95% CI	t(df)	p	Hedges' g
Pronunciation	1.14	0.62	0.52	[0.41, 0.62]	9.90 (197.21)	< .001	1.38
Word stress	1.28	0.57	0.72	[0.63, 0.80]	16.49 (201.00)	< .001	2.30
Total speaking score	2.42	1.19	1.23	[1.10, 1.37]	18.23 (182.52)	< .001	2.53

Note. Positive Δgain values indicate greater improvement in the experimental group. CI = confidence interval.

RQ3. Which speaking component, pronunciation or word stress, showed greater improvement through the use of pictograms?

To answer the third research question, pronunciation and word-stress gains were compared within the experimental group, since this group received the pictogram-based intervention. Table 6 shows that both components improved, but word stress showed a greater gain than pronunciation.

Table 6. Comparison of pronunciation and word-stress gains in the experimental group.

Comparison	Pronunciation gain M	Word stress gain M	Mean difference	t	df	p	Cohen's dz
Word stress vs. pronunciation	1.14	1.28	0.14	3.15	103	.002	0.31

Note. A positive mean difference indicates greater improvement in word stress than in pronunciation.

The experimental group obtained a mean gain of 1.28 points in word stress and 1.14 points in pronunciation, resulting in a mean difference of 0.14 points. This difference was statistically significant, t(103) = 3.15, p = .002, Cohen's dz = 0.31. Although the effect size was small to moderate, the result indicates that pictogram-based instruction had a slightly stronger effect on word stress than on pronunciation. Therefore, H3 was supported.

Integrated qualitative evidence

To address the mixed-methods component and clarify how the interview data complemented the quantitative results, Table 7 presents a joint display integrating the main statistical findings, representative student responses, and the resulting interpretation. The qualitative evidence explains how students perceived the role of pictograms in pronunciation and word-stress development, rather than treating the interviews as isolated thematic categories.

Table 7. Joint display integrating quantitative and qualitative evidence

Quantitative finding	Representative interview evidence	Integrated interpretation
Pronunciation improved significantly more in the experimental group.	Students explained that mouth pictures helped them position the tongue for sounds such as /th/, that divided words such as ap-ple helped them pronounce each part, and that visual markers helped them distinguish long and short sounds.	Pictograms supported the connection between visual meaning, sound form, and articulatory awareness, which helps explain the pronunciation gains.
Word stress showed the largest between-group effect.	Students reported that symbols such as a star, red syllable, hammer, clapping hand, or big circle helped them identify the stronger part of the word.	Visual cues made stress placement perceptually salient and easier to reproduce, which helps explain the larger word-stress gains.
Word stress improved more than pronunciation within the experimental group.	Several students perceived stress as easier because the visual cue clearly showed which syllable was stronger, whereas some sounds such as /th/ and /r/ were still challenging.	Stress patterns may be more directly represented through visual contrast and rhythm, while sound articulation requires finer motor control and repeated practice.

Overall, the quantitative and qualitative results converge in showing that pictogram-based instruction supported both pronunciation and word stress. The strongest effect was observed in word stress, and students' interview responses suggest that this advantage was related to the visibility of stress cues and the ease with which they could be applied during oral production.

5. Discussion

The findings of this study provide empirical evidence that pictogram-based instruction enhances the development of EFL oral performance, particularly pronunciation and word stress. Although both groups improved from pre-test to post-test, the greater gains observed in the experimental group indicate that pictograms functioned as meaningful pedagogical scaffolds rather than as supplementary

classroom illustrations. This interpretation is strengthened by the integration of interview evidence, which showed that students used pictograms to notice, remember, and produce oral features more clearly.

Regarding the first research question, pictogram-supported instruction positively influenced pronunciation development, particularly in sound articulation, intelligibility, and the reduction of first-language interference. This finding aligns with research showing that pronunciation instruction enhances intelligibility and comprehensibility when systematically integrated into classroom practice (Levis, 2018; Saito & Plonsky, 2019). However, the present study extends this literature by showing that a simple visual scaffold can support pronunciation practice among primary learners, not only through repetition but also through concrete cues for articulation and sound discrimination.

The qualitative evidence clarifies this mechanism. Students reported that mouth images supported tongue positioning, segmented words helped them pronounce syllables more clearly, and visual markers helped them distinguish long and short sounds. These responses suggest that pictograms facilitated the integration of meaning, phonological form, and articulatory awareness. This finding is consistent with Mayer's Cognitive Theory of Multimedia Learning and Paivio's Dual Coding Theory, which explain that learning is strengthened when verbal and visual information are meaningfully connected (Mayer, 2021; Paivio, 2007).

In relation to the second research question, pictograms had an even stronger effect on word stress. This result is important because previous work has shown that suprasegmental features, including lexical stress, strongly affect comprehensibility but are often underemphasized in EFL classrooms (Celce-Murcia et al., 2010; Field, 2005). The present study contributes to this discussion by showing that pictograms can make stress placement and prominence visible through symbols such as stars, colors, circles, hammers, and clapping hands. In this sense, pictograms may be particularly suitable for word stress because stress is relational and contrastive: learners must identify which syllable is stronger in relation to others.

This finding is broadly consistent with Ghaemi and Rafi (2018), who reported that visual aids supported the retention of English word stress patterns, and with Olson (2022), who highlighted the usefulness of visual feedback for refining stress production. Nevertheless, the present study differs from this previous work because it examined pictograms as the central instructional tool within classroom-based speaking tasks. Therefore, it offers more specific evidence on how low-tech, symbolic visual cues can support prosodic development in primary EFL learners.

The third research question showed that word stress improved more than pronunciation within the experimental group. One possible explanation is that word stress can be represented through visual contrast, such as size, color, or position, whereas pronunciation accuracy requires more fine-grained articulatory control and auditory discrimination (Derwing & Munro, 2015). Cognitive Load Theory also helps explain this difference. Because oral production requires learners to process meaning, articulation, rhythm, and fluency simultaneously, pictograms may have reduced cognitive demands by providing immediate visual references (Sweller, 1988; Sweller et al., 2019). This support appears especially beneficial for word stress, where a visual cue can directly mark prominence.

Some classroom factors may also have contributed to the observed gains. The intervention was implemented during regular English lessons, which may have supported continuity and familiarity. The repeated use of the same lesson sequence may have helped students understand what to do with the pictograms, while corrective feedback and modelling likely strengthened the effect of the visual cues. At the same time, learner motivation, teacher expertise, and possible exposure to English outside school may have influenced students' progress. These factors should be considered when interpreting the findings and when replicating the intervention in other contexts.

From a pedagogical perspective, the findings suggest that EFL teachers can use pictograms to move learners from recognition to oral production. For pronunciation-focused activities, pictograms can be used to represent mouth position, syllable segmentation, or sound length before students produce words in short phrases. For word stress, teachers can use stars, circles, colors, clapping gestures, or size contrasts to mark the stressed syllable and then guide students through repetition, rhythm practice, and short communicative responses. Importantly, pictograms should be integrated into modelling, feedback, and meaningful oral tasks rather than used as isolated visual decorations.

The study also has limitations that should be acknowledged. The intervention was conducted in a restricted geographical and institutional context, so the findings should be generalized with caution. In addition, the five-month intervention captured pre-test/post-test gains but did not examine long-term retention. Teacher-related variables may also have influenced the results, and formal inter-rater reliability statistics were not available for the rubric-based scores. Future studies should include delayed post-tests, independent raters, acoustic analyses, and comparisons between pictograms and other visual instructional approaches such as flashcards, picture cards, gesture-based instruction, or multimedia resources.

Taken together, the quantitative and qualitative findings indicate that pictogram-based instruction supported primary EFL students' oral development by linking visual meaning with speaking practice. The study contributes to EFL pedagogy by showing that pictograms can support not only vocabulary or reading-related learning but also key aspects of oral performance, particularly pronunciation and word stress. It also extends multimodal and cognitive accounts of language learning by showing how simple visual cues can mediate the transition from receptive recognition to productive speaking performance.

6. Conclusions

The findings of this study indicate that pictogram-based instruction had a positive effect on primary EFL students' speaking development. Students who received instruction supported by pictograms showed greater improvement than those in the control group in both

pronunciation and word stress, suggesting that visual scaffolds can facilitate the development of oral language features that are often challenging for primary learners.

More specifically, pictograms supported pronunciation by helping students connect visual representations with oral production. However, their strongest effect was observed in word stress, where students in the experimental group achieved the greatest gains. This suggests that pictograms may be particularly useful for teaching suprasegmental aspects of speaking, such as stress placement, rhythm, and comprehensibility, because these features can be made more concrete through visual support.

From a pedagogical perspective, the results support the integration of pictograms into EFL speaking instruction, especially when they are used as part of guided oral practice, teacher modelling, repetition, corrective feedback, and contextualized speaking tasks. Therefore, pictograms should not be viewed only as vocabulary or visual recognition aids, but as instructional scaffolds that can help learners notice, remember, and produce relevant oral features more effectively.

At the same time, the findings should be interpreted within the scope of the study. The intervention was conducted in a specific educational context with primary EFL learners, and the assessment was based on rubric scores. Future research could examine pictogram-based instruction in different educational levels and EFL settings, include delayed post-tests, incorporate acoustic analyses or inter-rater reliability procedures, and compare pictograms with other visual scaffolding strategies such as flashcards, picture cards, gestures, or multimedia resources.

Acknowledgments

We would like to express our gratitude to the authorities of Universidad Técnica Particular de Loja for their support of this research through the EFL Learning, Teaching and Technology Research group. We also extend our sincere thanks to all participants who generously completed the questionnaires.

Author contributions

All authors contributed equally to this work. Each of us participated in all stages of the manuscript, including:

- Conceptualization and study design
- Data collection and analysis
- Interpretation of results
- Writing and revision of the manuscript

Therefore, all authors share equal responsibility and credit for the entire work.

Funding

Not applicable

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Abdelaal, F. (2025). The impact of a pictogram-based instructional module on oral communication and academic engagement among pre-service FFL teachers: a quasi-experimental study. *Discover Education*, *4*(1), 487. https://doi.org/10.1007/s44217-025-00928-2

Alghazo, S., Jarrah, M., & Al Salem, M. N. (2023). The efficacy of the type of instruction on second language pronunciation acquisition. *Frontiers in Education, 8*, 1182285. https://doi.org/10.3389/feduc.2023.1182285

Almusharraf, A. (2024). Pronunciation instruction in the context of world Englishes: Exploring university EFL instructors' perceptions and practices. *Humanities and Social Sciences Communications*, *11*, 888 https://doi.org/10.1057/s41599-024-03365-y

Alwadei, A. M., & Mohsen, M. A. (2023). Investigation of the use of infographics to aid second language vocabulary learning. *Humanities and Social Sciences Communications, 10*, Article 108. https://doi.org/10.1057/s41599-023-01569-2

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE. https://doi.org/10.4135/9781036232078.n9

Brown, H. D. (2019). *Principles of language learning and teaching* (6th ed.). Pearson.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Bühler, D. (2021). *Universal, intuitive, and permanent pictograms*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32310-3

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages*: *Learning, teaching, assessment: Companion volume.* Council of Europe Publishing.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.42

Derwing, T. M., & Munro, M. J. (2021). *Pronunciation fundamentals.* John Benjamins.

Dewaele, J. M., & MacIntyre, P. D. (2019). The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. *Evidence-based second language pedagogy*. 263-286). Routledge. https://doi.org/10.4324/9781351190558-12

Dörnyei, Z. (2020). *Innovations and Challenges in Language Learning Motivation* (1st ed.). Routledge. https://doi.org/10.4324/9780429485893-1

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4. https://doi.org10.11648/j.ajtas.20160501.11

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, *39*(3), 399-423. https://doi.org/10.2307/3588487

Flege, J. E. (1995). *Second language speech learning theory*. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233-277). York Press. https://doi.org/10.1017/9781108886901.003

Flege, J. E., Aoyama, K., & Bohn, O. S. (2021). The Revised Speech Learning Model (SLM-r) Applied. In: Wayland R, ed. *Second Language Speech Learning: Theoretical and Empirical Progress*. 84-118. Cambridge University Press.

Ghaemi, F., & Rafi, F. (2018). The impact of visual aids on the retention of English word stress patterns. *International Journal of Applied Linguistics and English Literature*, *7*(2), 225-231. https://doi.org10.7575/aiac.ijalel.v.7n.2p.225

Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach.* Cambridge University Press.

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine, 13*(6), 554-561. https://doi.org/10.1370/afm.1865

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*(3), 626-668. https://doi.org/10.1111/lang.12451

Hughes, R., & Szczepek Reed, B. (2017). *Teaching and researching speaking* (3rd ed.). Routledge. https://doi.org/10.4324/9781315692395

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on second language intelligibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475-505. https://doi.org/10.1017/S0272263112000150

Ivanova, M. (2024). Training English Word Stress Perception and Production with Technology. *The EuroCALL Review*, *31*(2), 45-56. https://doi.org/10.4995/eurocall.2024.20400

Kenworthy, J. (1987). *Teaching English pronunciation*. Longman.

Kress, G., & Van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge. https://doi.org/10.4324/9781003099857

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Lightbown, P. M., & Spada, N. (2021). *How languages are learned* (5th ed.). Oxford University Press.

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781108894333.003

Medina, C. M. S., & Romero, I. C. (2025). Using pictograms to enhance reading comprehension in English young learners. *Polo del Conocimiento*, *10*(1), 2237-2255.

Metruk, R. (2025). EFL pronunciation instruction: Czech vs Slovak EFL students' perceptions. *International Journal of Instruction*, *18*(4), 713-728. https://doi.org/10.29333/iji.2025.18438a

Olson, D. J. (2022). Visual feedback and relative vowel duration in L2 pronunciation: The curious case of stressed and unstressed vowels. In J. Levis & A. Guskaroska (Eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference* (pp. xxx–xxx). Iowa State University. https://doi.org/10.31274/psllt.13353

Paivio, A. (1990). *Mental representations: A dual coding approach.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Lawrence Erlbaum Associate Psychology Press.

Panyathikul, W., Poopatwiboon, S., & Phusawisot, P. (2025). Improving EFL secondary learners' pronunciation through multimodal teaching. *Journal of Education and Learning*, *14*(1), 148-160. https://doi.org/10.5539/jel.v14n1p148

Pauzhi, P., & Argudo-Serrano, J. (2024). The effects of using pictograms during the English teaching-learning process in primary school. *Religación*, *9*(39). https://doi.org/10.46652/rgn.v9i39.1141

Qin, Y., & Wang, P. (2021). How EFL teachers engage students: A multimodal analysis of pedagogic discourse during classroom lead-ins. *Frontiers in Psychology*, *12*, 793495. https://doi.org/10.3389/fpsyg.2021.793495

Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge University Press.

Richards, J. C. (2015). *Key Issues in Language Teaching*. Cambridge University Press. https://doi.org/10.1017/9781009024600

Richards, J. C. (2022). *Teaching speaking in a second language classroom*. Cambridge University Press.

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly*, *46*(4), 842-854. https://doi.org/10.1002/tesq.67

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among L2 learners? *Language Teaching Research*, *25*(3), 403-425. https://doi.org/10.1177/1362168820938317

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3), 652-708. https://doi.org 10.1111/lang.12345

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108569057

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Suwanasilp, D., & Durongbhandhu, N. (2023). English vocabulary with a picture application for enhancing Thai EFL students' daily English vocabulary memorization. *rEFLections*, *30*(3), 712-734. https://doi.org/10.61508/refl.v30i3.268078

Suzukida, Y., & Saito, K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, *25*(3), 431-450. https://doi.org/10.1177/1362168819858246

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, 110, 1-5, 102423. https://doi.org10.1016/j.lindif.2024.102423

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and*

Teaching, *17*(3), 738-754. https://doi.org/10.1080/17501229.2022.2131791

Uzun, T. (2024). Perception of prosodic speech features: Final intonation and word stress for EFL learners. *Uluslararası Dil Edebiyat ve Kültür Araştırmaları Dergisi*, *7*(2), 212-223. https://doi.org/10.37999/udekad.1449612

Wang, Q., Jun, Y., & Lee, H. (2022). *Nonverbal communication of symptoms: Visualization and cognition of symptom pictograms on user comprehension*. *Archives of Design Research*, 35(2), 7-25. http://dx.doi.org/10.15187/adr.2022.05.35.2.7

Windriya, S. D. R., Wajdi, M., Dizon Jr, R. P., Hasan, M. K., & Idri, N. (2025). Exploring the role of speaking in second language acquisition: Key challenges learners face and effective strategies applied. *Journal of Language and Pragmatics Studies*, *4*(1), 26-37. https://doi.org/10.58881/jlps.v4i1.49

Yenkimaleki, M., & van Heuven, V. J. (2021). Effects of attention to segmental vs. suprasegmental features on the speech intelligibility and comprehensibility of the EFL learners targeting the perception or production-focused practice. *System*, *100*, 102557, 1-12. https://doi.org/10.1016/j.system.2021.102557

Zúñiga Vargas, J. P. (2021). Promoting the Pedagogical Use of Pictograms: A Case Study of Five In-Service Costa Rican English Teachers. *BELT - Brazilian English Language Teaching Journal*, *12*(1), 1-16. https://doi.org/10.15448/2178-3640.2021.1.41142

Appendix

Analytic speaking rubric

Dimension	Indicator	High performance	Developing performance	Limited performance
Pronunciation	Sound articulation	Produces target sounds clearly and consistently.	Produces most sounds clearly, with occasional distortion.	Frequent sound distortions reduce clarity.
	Intelligibility	Message is easily understood by listeners.	Message is generally understandable with some effort.	Message is difficult to understand.
	First-language interference	L1 influence does not interfere with comprehensibility.	Some L1 influence is present but meaning remains clear.	L1 influence frequently affects comprehensibility.
Word stress	Stress placement accuracy	Places stress on the correct syllable in target words.	Sometimes places stress accurately.	Stress placement is often inaccurate.
	Stress prominence	Stressed syllables are clearly marked through emphasis.	Prominence is partially marked.	Prominence is weak or absent.
	Rhythm and comprehensibility	Stress contributes to rhythm and clear oral production.	Rhythm is uneven but generally comprehensible.	Incorrect stress disrupts rhythm and understanding.

Note. The full rubric used a 0–5 scale for each dimension, with pronunciation and word stress contributing to a total speaking score of 10 points.