e-ISSN 1135-9250

EDUTEC - Revista Electrónica de Tecnología Educativa.

Issue 93 – September 2025

AI-Enhanced Multisensory Learning Intervention to Improve Pronunciation in EFL Students

Intervención de aprendizaje multisensorial apoyada por IA para mejorar la pronunciación en estudiantes de inglés como lengua extranjera

Diana Carrascal Tris (D.C.T.). Universidad de Valladolid (España)

Qianting Yuan (Q.Y.) Universidad de Valladolid (España)

ABSTRACT

Despite early exposure to English as a foreign language (EFL), Spanish learners often struggle with its pronunciation due to fundamental differences between the transparent orthography in Spanish and the deep orthography in English. This study examines the effectiveness of a multisensory learning (MSL) methodology enhanced by artificial intelligence (AI) tools and based on the dual-route model of reading, which distinguishes between sublexical decoding and lexical word recognition, in improving English pronunciation accuracy. 94 university students were divided into a target group and a control group. The intervention involved a pre-test and a post-test and a set of reading-aloud tasks, all conducted via the Wooflash platform and focused on five difficult English vowel phonemes: /ə/, /uː/, /ʌɪ/, /eɪ/, and /ɔː/. The objective was to compare the effectiveness of MSL and traditional training (TT), and to assess participants' experiences of the MSL approach. Quantitative results showed that MSL was as effective as TT in improving pronunciation accuracy, while qualitative feedback from the post-activity questionnaire revealed emotional, cognitive and metacognitive benefits of the MSL intervention, including increased confidence and awareness of pronunciation challenges. These findings suggest the potential of MSL as a learner-centered and engaging tool for improving pronunciation in EFL contexts.

RESUMEN

Pese a la exposición temprana al inglés como lengua extranjera (EFL), los estudiantes españoles suelen tener dificultades con la pronunciación debido a las diferencias entre la ortografía transparente del español y la ortografía opaca del inglés. Este estudio analiza la eficacia de una metodología de aprendizaje multisensorial (MSL), apoyada por la IA y basada en el modelo de doble ruta de lectura, que distingue entre la descodificación subléxica y el reconocimiento léxico, para mejorar la pronunciación en inglés. Participaron 94 estudiantes universitarios, divididos en un grupo experimental y uno de control. La intervención incluyó una prueba previa y otra posterior, junto con tareas de lectura en voz alta mediante la plataforma Wooflash, centradas en cinco fonemas vocálicos del inglés: /ə/, /uː/, /ʌɪ/, /eɪ/ y /ɔː/. El objetivo fue comparar la eficacia del MSL con la formación tradicional (TT) y evaluar la experiencia del alumnado con MSL. Los resultados cuantitativos mostraron que MSL fue tan eficaz como la TT para mejorar la pronunciación, mientras que las respuestas cualitativas reflejaron beneficios emocionales, cognitivos y metacognitivos, como mayor confianza y conciencia fonológica. Estos hallazgos apuntan al potencial del MSL como herramienta motivadora y centrada en el alumnado para mejorar la pronunciación en EFL.

KEYWORDS - *PALABRAS CLAVE*

Second language learning, phonetics, language learning, educational technology, teaching methods
Aprendizaje de segunda lengua, fonética, aprendizaje del lenguaje, tecnología educativa, métodos de enseñanza

Received: 17-06-2025
Accepted: 03-08-2025

1. INTRODUCTION

1.1. Background and Context of English Pronunciation in EFL

Pronunciation is an underestimated critical component of English as a Foreign Language (EFL) learning, significantly influencing communicative competence and learner confidence (Rallo, 2022; Saito, 2021; Pennington & Rogerson-Revell, 2019). As English has become the global lingua franca between speakers with different first languages since the late 1980s, clear and intelligible pronunciation is essential for effective communication, particularly in contexts like Spain, where English is a widespread requirement (Gómez González & García Muras, 2025; Seidlhofer, 2011; Veiga Pérez, 2017). Accurate pronunciation enhances comprehension, minimizes misunderstandings, and fosters more fluid and meaningful conversations; moreover, it can also boost learners' confidence, self-esteem, motivation, receptive skills, and even cultural sensitivity (Gómez González & García Muras, 2025; Derwing & Munro, 2015).

Despite its importance, pronunciation has often been marginalized in EFL instruction during recent years (Fouz-González, 2020). It frequently gives way to methodologies such as Grammar-Translation and Reading-based approaches (Levis, 2005; Nikbakht, 2011). Such traditional training (TT) often lacks timely feedback and personalized instruction, which are vital for improving pronunciation skills and avoiding fossilized errors in students (Nikbakht, 2011). Other challenges include cross-linguistic influence from learners' first language (L1) (e.g., Spanish in this study) and insufficient exposure to authentic input. This discrepancy between the importance of pronunciation and its limited instructional emphasis calls for innovative, learner-centered approaches to address the pedagogical gap.

1.1.1. Pronunciation Challenges for Spanish EFL Learners

English and Spanish differ significantly in orthographic and phonological systems. Spanish has transparent orthography with a near one-to-one grapheme-phoneme correspondence (GPC), whereas English has a deep orthography where one grapheme can correspond to multiple phonemes (e.g., the grapheme "a" in *cat* /æ/, *car* /ɑː/, *grape* /eɪ/, *sofa* /ə/) (García & Froud, 2018; Sammour-Shehadeh et al., 2025). These inconsistencies contribute to cross-linguistic transfer (Fadillah, 2020; Odlin, 2012; Pennington & Rogerson-Revell, 2019), which leads Spanish speakers to misapply Spanish phonological rules to English pronunciation.

English vowels are particularly challenging for Spanish learners, whose native inventory includes only five pure vowel sounds (i.e., /i/, /e/, /a/, /o/, /u/) (García & Froud, 2018). In contrast, English possesses a much wider range of monophthong and diphthongs, some without Spanish equivalents (e.g., /ɜː/, /ə/, /æ/, /ʌ/, /ɑː/) (Pennington & Rogerson-Revell, 2019). These unfamiliar sounds are often substituted with the closest Spanish vowel sound or produced with undue emphasis spelling (e.g., *today* as /tu'deɪ/ instead of /tə'deɪ/). While contrasts like /i/-/ɪ/ and /ɑ/-/ʌ/ have been well-researched (e.g., Casillas, 2015; Yang et al., 2016), diphthongs such as /ʌɪ/, /eɪ/ and /ɔː/ remain underexplored despite their high difficulty for Spanish learners of EFL.

1.1.2. Current Situation of Pronunciation Instruction in EFL Context, particularly in Spain

The role of pronunciation in EFL instruction has fluctuated. Initially emphasized in methods such as the Direct Method and the Audiolingual Approach, it was later sidelined by grammar-focused methods like the Cognitive Approach (Nikbakht, 2011). Since the early 2000s, interest has resurged, particularly in suprasegmental features such as stress, rhythm, and intonation (Pennington & Rogerson-Revell, 2019), though pronunciation remains marginal in many contemporary curricula and teaching materials (Gómez González & García Muras, 2025).

In Spain, despite various education reforms, pronunciation continues to be undervalued and under-addressed in EFL classrooms (Gómez González & García Muras, 2025; Rallo, 2022; Saito, 2021). This gap is reflected by persistent pronunciation difficulties across educational stages. Recent research highlights a significant discrepancy between curriculum objectives and classroom realities. Their findings align with a broader European trend, calling for comprehensive revisions to current pronunciation teaching methodologies (Gómez González & García Muras, 2025).

In contrast, recent research increasingly supports explicit phonetic instruction as a crucial strategy to address EFL pronunciation challenges (Calvo Benzies, 2016; Gómez González & García Muras, 2025; Pennington & Rogerson-Revell, 2019). Effective practices include teaching cross-linguistic phonetic inventories, grapheme-phoneme mappings, and L1-L2 contrasts to raise phonological awareness (Chapuis & Berthele, 2024; Pennington & Rogerson-Revell, 2019). However, these findings are rarely implemented systematically in current teaching practices, which indicates a clear gap between research-proven recommendations and classroom execution.

1.2. Theoretical Framework: Pseudoword Reading and EFL Pronunciation

Combining listening with reading is a widely endorsed pedagogical practice to improve EFL pronunciation, as the visual representation of words supports learners in mapping phonemes to graphemes (Saito & Plonsky, 2019). This section outlines the cognitive mechanisms of word reading, focusing on the dual-route processing model (section 1.2.1), and explores how sublexical decoding through pseudoword training contributes to pronunciation development (section 1.2.2).

1.2.1. Dual-Route Model: Lexical and Sublexical Processing

Reading a word requires processing visual, orthographic, phonological, and semantic information. In recent decades, several models have been developed to explain the word recognition process, including the Dual-Route Cascade Model (DRC) (Coltheart et al., 2001), Dual Process Connectionist Model (PDC), and Triangular Connectionist Model (Coltheart et al., 2001; Seidenberg & McClelland, 1989). All these models involve phonological, orthographic, and semantic modules, with their engagement depending on the lexical status (word or pseudoword), the reader's familiarity with the word, and their reading experience (novice or experienced reader) (Coltheart et al., 2001).

The DRC model, one of the most widely accepted, distinguishes two processing routes: the sublexical route and the lexical route (Coltheart et al., 2001). The sublexical route uses

grapheme-phoneme conversion rules to decode unfamiliar or pseudowords letter by letter. The lexical route, conversely, is used to recognize words that have a representation in the orthographic lexicon. While the sublexical route is primarily used at the beginning of reading acquisition, it gradually gives way to a more prominent role for the lexical route as reading proficiency develops.

In Spanish, grapheme-phoneme conversion is straightforward due to its consistent GPCs. However, English, with its deep orthography—a spelling often corresponds to different phonemes, as exemplified in section 1.1.1—poses significant difficulty to the EFL reading process for Spanish speakers (Łockiewicz et al., 2020; Ziegler & Goswami, 2005).

EFL learners must draw on phonological, orthographic, and morphological knowledge to decode unfamiliar items. Lower-proficiency students tend to rely more on bottom-up processing, focusing on individual words and their meanings (Cao et al., 2014; Coltheart, 2006; Rayner & Reichle, 2010; Snowling et al., 2022). During this process, sublexical processing is essential for reading new or unfamiliar words, while lexical processing allows for the rapid identification of familiar words. Reading pseudowords, which are not meaningful words but follow the phonetic patterns of the language, supports the sublexical route and trains learners to apply grapheme-phoneme rules and eventually enhance their ability to decode unfamiliar real words (Snowling et al., 2022).

1.2.2. Sublexical Decoding and Pseudoword Training in EFL Pronunciation Development

Pseudowords, also referred to as invented words or wug words, are carefully constructed letter strings that visually and phonologically resemble real words in a specific language but inherently possess no actual meaning (Coltheart, 2006). They rigorously follow the phonetic and orthographic rules of the target language to ensure their pronounceability while excluding semantic cues (e.g., *shum, laip,* and *cigbet* in English).

Integrating pseudoword training supports the sublexical route by reinforcing GPCs. Learners practice decoding novel, unfamiliar strings, thereby strengthening foundational phonological skills critical for pronunciation (Ernesto Macaro, 2012; Łockiewicz et al., 2020). Without semantic cues, learners must systematically apply decoding rules, promoting explicit phonological awareness which directly enhances pronunciation accuracy (Saito & Plonsky, 2019).

Research shows that pseudoword training improves decoding competence and phonological representation, yielding lasting improvements in both reading pseudowords and correctly pronouncing similar real words (Mlakar et al., 2025; Zhang & Peng, 2022). Consequently, embedding pseudoword activities in EFL instruction holds significant promise for improving pronunciation proficiency, particularly for learners whose L1 provides insufficient training in sublexical decoding (Mora et al., 2022).

1.3. AI-Enhanced Multisensory Learning (MSL) in Foreign Language (FL) Education

1.3.1. Effectiveness of MSL Approaches in FL Learning

MSL is a pedagogical approach that engages various senses—sight, hearing, touch, and movement—simultaneously to enhance understanding and retention. MSL has been widely proven to be effective in students with specific learning needs (Kast et al., 2007; Kelly & Phillips, 2022). However, MSL in FL learning has still to prove its potential in enhancing language learning, motivation, and retention across FL learners through empirical research. Hence, MSL offers several advantages for FL learning, especially complex skills such as pronunciation. By engaging multiple senses (auditory, visual, kinesthetic), MSL enhances the encoding and retrieval of information, leading to improved retention and a deeper understanding of the material (Shams & Seitz, 2008; Zlatkova-Doncheva, 2022). Furthermore, interactive activities that incorporate various sensory modalities, such as watching videos, listening to music, or engaging in role-playing scenarios, are significantly more engaging than passive learning methods. This heightened engagement naturally boosts student motivation and fosters a greater desire to learn (Zlatkova-Doncheva, 2022). MSL also promotes the development of a broader range of skills, including not only language proficiency but also communication abilities, critical thinking, and creativity, through collaborative group work and creative tasks.

For pronunciation specifically, multimodal approaches that integrate visual, auditory, and kinesthetic elements have proven effective in reinforcing articulatory features and prosodic structures (Shams & Seitz, 2008). Visual aids, such as diagrams of the vocal tract and spectrograms, can clarify the mechanics of sound production, while physical actions like hand gestures can reinforce specific articulatory movements (Li et al., 2023). Phonetic training itself has evolved to embrace multimodal approaches, moving beyond TT phoneme-focused methods to integrate a variety of sensory elements for more comprehensive learning. MSL acts as a bio-cognitive amplifier for the acquisition of complex language skills like pronunciation (Nikbakht, 2011; Pennington & Rogerson-Revell, 2019). By mimicking natural learning processes and engaging diverse neural pathways, it strengthens memory traces and facilitates the integration of motor, auditory, and visual information necessary for accurate speech production. This approach moves beyond purely intellectual understanding to embody the learning, making it more robust and accessible, especially for skills that require fine motor control and perception.

1.3.2. Integrating Artificial Intelligence (AI) with MSL for Pronunciation Improvement

AI is transforming education, offering innovative solutions that address many inherent limitations of traditional FL instruction (Shams & Seitz, 2008). AI enables personalized learning, dynamic environments, and adaptive progress tracking (Hsiao & Chang, 2023; Selim, 2024; Setyaningsih et al., 2024)—features that align naturally with the principles of MSL. AI tools can enhance pronunciation training by providing instant correction, gamified practice and interactive feedback, and thereby support learner motivation and reduce anxiety through self-paced, private learning contexts (Ai et al., 2025; Song & Song, 2023; Yuan & Liu, 2025). These benefits are especially relevant in lowering the affective filter, which is a key barrier to FL learning (Krashen, 1982).

Empirical evidence further supports the integration of AI and MSL. For example, the ELSA Speak app—powered by AI-driven speech recognition—has demonstrated to improve EFL students' pronunciation skills (Permatasari, 2024). Advanced AI models, such as NSE-CATNet, apply Transformer-based models to enhance speech intelligibility by capturing long-term contextual dependencies. These models also incorporate phonetic class information to reduce misclassification errors, which yields more accurate pronunciation feedback (Permatasari, 2024).

Hence, the combination of the adaptive, data-driven capacities of AI with the embodies, engaging nature of MSL may significantly improve students' motivation, enrich the personalization of the EFL experience, and demonstrably enhance pronunciation skills. This integrated approach thus represents a promising pedagogical framework for EFL pronunciation instruction.

1.4. Rationale, Objectives and Research Questions

This last section of the Introduction presents the rationale behind the study, identifies the research gap and outlines the objectives and research questions.

1.4.1 Research Gap and Justification of the Intervention

Despite years of EFL instruction, Spanish university students continue to face persistent challenges with English vowel pronunciation largely due to L1 interference and the limitations of TT (Calvo Benzies, 2017; Pennington & Rogerson-Revell, 2019). Existing studies have explored AI and MSL approaches separately, but few have integrated them within the DRC model to target sublexical decoding of problematic vowel phonemes in this population.

The DRC model of reading highlights the importance of sublexical decoding, particularly for processing unfamiliar words and deep orthographies like English. Pseudoword training directly targets this decoding route by reinforcing GPCs (Mlakar et al., 2025). The integration of AI and MSL offers a promising approach to strengthen this mechanism by delivering adaptive, multisensory input and real-time feedback.

This study aims to move beyond general observations of pronunciation difficulty or comparative method efficacy. Instead, it seeks to determine how targeted training can improve sublexical processing in EFL learners and reduce fossilized pronunciation errors. This will offer practical insights for curriculum design and replication in EFL contexts.

1.4.2. Objectives and Research Questions

The primary objective of this study is to assess the effectiveness of an AI-enhanced MSL intervention, theoretically framed on the dual-route model of reading, for improving EFL vowel pronunciation among Spanish university students. Furthermore, the study also explores learners' experiences of the intervention.

To address these objectives, the following research questions have been formulated:

RQ1. Is the MSL intervention as effective as the TT for improving EFL pronunciation accuracy?

RQ2. Is the MSL intervention particularly effective for pseudoword pronunciation compared to real words?
RQ3. What are participants' experiences of the MSL intervention?

2. METHOD

This research project is part of a teaching experience developed during the 2024/2025 academic year in different Specialized/Technical English courses offered to undergraduate students in Nursing, Medicine and Journalism Degrees at a public university in Spain. A total of 94 second-year students participated voluntarily in the study; none of the tasks contributed to their course grades. The students were divided into a target group (50 Nursing students) and a control group (44 Medicine and Journalism students). All the participants were native Spanish speakers and learners of EFL.

2.1. Ethical Statement

This study involved human participants and the use of technological tools for data collection. The following ethical measures were adopted to ensure participants' rights, data protection, and responsible technology use.

Before they started the activity, all the participants were given a written informed consent which specified the procedure, the purpose of the study, and the confidentiality commitment to protect their privacy and identities. The participants who agreed to take part in the study were asked to sign the consent form and were given a hard copy for their records. Participation was entirely voluntary, and students were informed that none of the tasks was assessed as part of their course grade and that they could withdraw at any time without consequences.

The research protocol was reviewed and approved by the institutional ethics committee on June 5, 2025.

Personal data: All personal data were handled in accordance with the General Data Protection Regulation (GDPR). Data collection was limited to what was strictly necessary for the research objectives. Participant data were anonymized and stored securely to guarantee confidentiality. Only the research team had access to the data, and all the identifying information was removed prior to analysis.

The study used Wooflash, which is an online interactive platform with different activities, including audio recording. To address potential ethical implications related to its use, the following measures were taken:

Regarding bias and accuracy, the audio recordings were analyzed manually by the same well-trained researcher to avoid potential bias in speech recognition technologies and individual differences.

In terms of accessibility, access to materials was provided via Campus Virtual or through a QR code distributed in class. Participants used their own devices, and instructions were provided

to ensure equal access to the platform regardless of technical proficiency. A supervisor offered on-site support for those who encountered technical issues.

Regarding privacy, the platform's recording function was used exclusively for the purposes of the study, and participants were informed that their recordings would not be shared outside the research context. No biometric or sensitive data were collected, and all recordings were stored securely in compliance with data protection laws.

2.2. Activity design

All the activities in this project were developed on Wooflash. The participants completed an English proficiency test at the beginning of the course. The results indicated that their English level ranged from intermediate to advanced. All the rest of the tasks, which included a pre-test, a training session, a post-test, and a post-activity questionnaire in the case of the target group, took place in a 50-minute class session.

On the day of the main tasks, the participants first took the pre-test, which involved reading aloud ten pseudowords and ten real words, together with five filler items. The target items focused on five specific phonemes, each represented by two graphemes. These phoneme-grapheme correspondences were selected because their pronunciations in English differ greatly from the Spanish pronunciation rules; therefore, they can be challenging for native Spanish speakers. The words were randomized and shown on slides one by one without accompanying images. The post-test followed the same procedure but used a different set of words. Table 1 shows a summary of the items in the pre- and post-tests.

Table 1

Pre- and post-test Items

Phonemes	Graphemes	Pseudowords	Real words
	Pre-test		
/ə/	ar, or	pellor, gralatar	cellar, translator
/uː/	oo, ew	sewer, soovenir	gloomy, brewery
/ʌɪ/	ie, igh	detie, resignment	mighty, replier
/eɪ/	ai, ey	ebey, detainor	daisy, disobey
/ɔː/	aw, au	seesau, desawer	lawful, somersault
Fillers	rumble, saddle, lobbin, baffle, grumbly		
	Post-test		
/ə/	ar, or	lensor, vinagar	altar, adaptor
/uː/	oo, ew	mootless, screwdary	chewing, gloomily
/ʌɪ/	ie, igh	unlie, preghtening	brighten, decrier
/eɪ/	ai, ey	braiser, disabey	survey, acquaintance
/ɔː/	aw, au	assaut, recauder	outlaw, tarpaulin
fillers	trellis, flare, crystal, crispy, jingle		

Notes. Read-aloud and audio-recorded items in the pre- and post-tests.

In both the pre- and post-tests, one word was displayed at a time on the screen of each participant's device. Participants were instructed to read the word aloud after pressing the "record" button on the screen. Audio recordings were automatically captured by the Wooflash platform. Once a recording was completed, participants clicked the "next" button to proceed to the following item. They were allowed to listen to their recordings and re-record their responses as many times as they wished.

Regarding the training session, for the target group, it included a series of interactive exercises incorporating both auditory and visual elements. Participants were presented with a total of 20 sentences, each of which contained a pseudoword and a real word that rhymed, as shown in example 1.

(1) She drank some cold *dratar* instead of *water*.

Each sentence was paired with a set of related images generated by Bing Image Creator, as shown in Figure 1, and read aloud by SpeechGen.IO. All audio files were reviewed to ensure that the practice items were pronounced correctly. In addition, both female and male voices were used equally, and both British and American English accents were included to guarantee gender balance and accent variety.

Figure 1

An image of the training session generated by Bing Image Creator

The training tasks required participants to match each sentence with its corresponding image, audio clip, or transcription through a variety of activity formats, such as multiple-choice questions and pairing exercises. There were no objectively correct answers for the questions so as to increase student motivation and engagement. The training for the target group employed a MSL approach, involving visual, auditory, and tactile modalities.

For the control group, the same practice items were presented, but without accompanying images. Instead, sentences were displayed on the screen, and audio clips were played to the

participants. Participants were not given any interactive tasks to complete, which reflected a TT format.

Participants in both the target group and the control group were allowed to listen to the audio clips as many times as they wanted and were encouraged to practice reading the sentences aloud.

In order to assess participants' experiences with the MSL training session, the target group completed a questionnaire at the end of the class. It consisted of eight questions focusing on their feelings during the activities, the aspects they found most enjoyable or challenging, and the perceived impact of the practice on their understanding and pronunciation of the text. Participants rated their responses as "positive," "neutral" or "negative" and were invited to provide open-ended comments if they wished.

2.3. Data collection, codification and analysis

In order to address the research questions, the participants' audio recordings from the pre- and post-tests were collected using the recording feature on Wooflash. Each participant produced 25 audio recordings per test: 10 corresponding to pseudowords, 10 to real words and 5 to fillers. The researchers then coded the audio recordings of the target items in an Excel spreadsheet.

Coding was done dichotomously as either "correct" (1) or incorrect (0), focusing only on the pronunciation of the target grapheme in each word. For instance, if a participant mispronounced the first syllable of the pseudoword "pellor" but correctly articulated the grapheme "-or," the item was coded as correct. In the case of self-correction, only the last pronunciation of each recording was taken into account.

Unintelligible recordings were discarded, and participants with fewer than 10 codable audio samples per test were excluded from the analysis. The final analysis included data from 41 participants in the target group and 22 participants in the control group.

The statistical analysis was conducted using a generalized linear mixed model (GLMM), which is suite to binary outcome variables, as is the case of pronunciation accuracy in our study. Additionally, a Difference-in-Differences (DID) analysis was conducted to compare improvement over time between the target and control groups by isolating the effect of the MSL intervention relative to the control condition. To ensure robust estimation, multiply imputed datasets were used to handle missing values, and the DID estimates were pooled following Rubin´s rules.

Qualitative data obtained through a post-activity questionnaire were analyzed for sentiment distribution (i.e., numbers of "positive," "neutral" and "negative" responses) and for common reflections in the open-ended comments on different aspects (e.g., engagement, confidence building, pronunciation awareness) to offer complementary insights into learners' experiences with the MSL training session.

All data and analysis scripts for this study are available via the Open Science Framework at https://osf.io/su8rx/?view_only=e185d939be81424aa44d64ac1fcd8f58.

3. RESULTS

This study compared MSL pronunciation training with the TT approach. It aimed to explore whether the former was as effective as the latter, as well as the participants' experiences of the MSL training method.

3.1. Effectiveness of the intervention

A summary of the pooled fixed effects from the GLMM is presented in Table 2. The model examines the effects of test time (pre-test vs. post-test), group (target vs. control) and word type (real word vs. pseudoword) on the correctness of responses. The model included a three-way interaction term (time × group × word type) and random intercepts for student and item.

Table 2

Fixed effects from the GLMM model predicting correct production

Term	Estimate	Std. error	*p* value	Conf. low	Conf. high	OR	OR_low	OR_high
Time (post) × Group (target)	-0.28	0.32	0.373	-0.90	0.34	0.75	0.40	1.41
Time (post) × Word Type (real)	-1.54	0.58	**0.008**	-2.68	-0.41	0.21	0.07	0.66
Group (target) × Word Type (real)	0.22	0.30	0.455	-0.36	0.81	1.25	0.69	2.26
Time × Group × Word Type	0.06	0.44	0.895	-0.81	0.93	1.06	0.44	2.53

Notes. Significant effects are in **bold**.

In the pre-test, as seen in Table 3 below, the target group showed lower accuracy for pseudowords than the control group. For real words, the target group and the control group were more comparable. While there were baseline differences in accuracy, particularly for pseudowords, a GLMM revealed no significant interaction between group and word type (β = 0.22, SE = 0.30, p = 0.455), as shown in Table 2, which suggests similar proficiency patterns across the two groups before training.

Table 3

Accuracy by group and word type in the pre- and post-tests

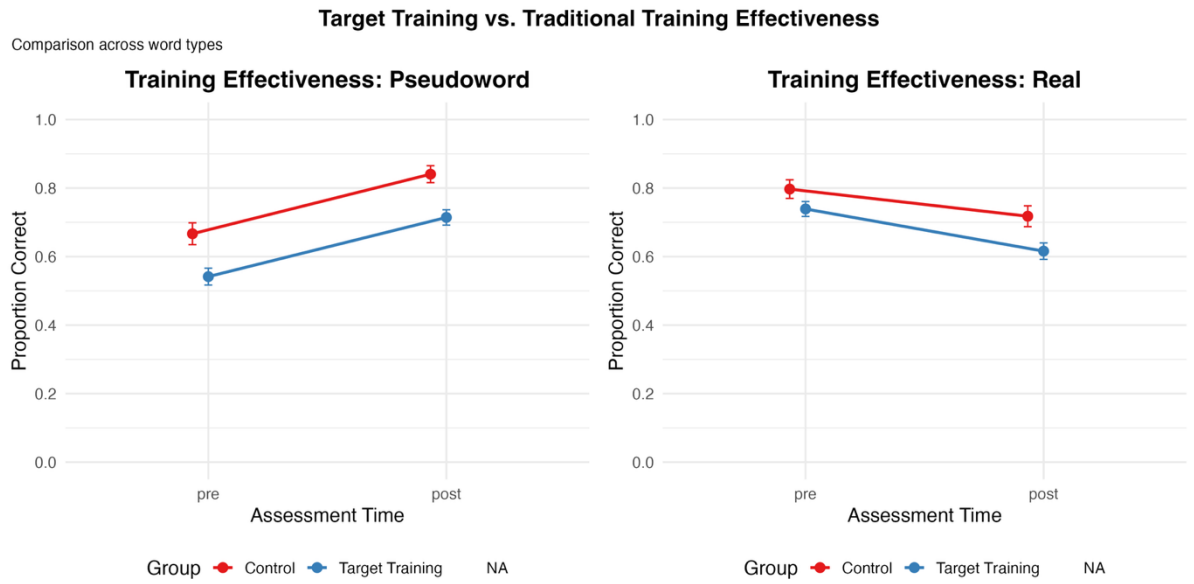
Word type	Target group (pre-test)	Control group (pre-test)	Target group (post-test)	Control group (post-test)
Pseudowords	0.54 (0.02)	0.67 (0.03)	0.71 (0.02)	0.84 (0.02)
Real words	0.74 (0.02)	0.80 (0.03)	0.62 (0.02)	0.72 (0.03)

Notes. Values represent proportion correct (SE).

In the post-test, as seen in Figure 2, both groups showed improvements in pronouncing pseudowords, though real word accuracy declined slightly.

Figure 2

Effectiveness of MSL training and TT training across word type



The GLMM model revealed a significant baseline difference between the target and the control groups (β = -0.70, SE = 0.27, p = 0.011), as shown in Figure 2, which indicates lower pre-test accuracy in the target group. However, the interaction between group and time was non-significant (β = -0.28, SE = 0.32, p = 0.373), as seen in Table 2. This indicates that both groups showed similar improvement from pre- to post-test, although the initial differences between the groups persist in the post-test.

In terms of the training effect, in general, both groups improved significantly from pre-test to post-test (β = 1.25, SE = 0.42, p < .010), which showed a general learning effect. A DID analysis was conducted based on the group × time × word type interaction terms in the GLMM; the results were pooled across multiply imputed datasets, as shown in Table 4. The results showed no statistically significant difference in improvement between the two groups.

Table 4

DID estimates of training effect by word type

Word type	Estimate	SE	z	*p*	OR	95% CI (OR)	*p* adj
Pseudowords	0.28	0.32	0.89	.373	1.33	[0.71, 2.48]	.373
Real words	0.22	0.30	0.73	.464	1.25	[0.69, 2.28]	.464

Notes. Estimates are based on log-odds from the GLMM. Confidence intervals are on the OR scale.

In terms of the effect of word type, the results in Table 4 show that the odds of improvement were slightly higher in the target group for both pseudowords (OR = 1.33) and real words (OR = 1.25). However, these differences were not statistically significant, and the confidence intervals included 1.00. Adjusted p-values for multiple comparisons also remained non-significant after correcting for multiple comparison. These findings suggest that MSL training was as effective as TT.

Last but not least, the three-way interaction (group × time × word type) which tested whether training effects differed by word type was not statistically significant (β = 0.06, SE = 0.44, *p* = 0.895), as shown in Table 2.

In sum, the quantitative results of the study indicate that the effectiveness of the MSL approach is comparable to that of the TT approach (RQ1), regardless of word type (RQ2).

3.2. Participants' experiences of the MSL Intervention

While the quantitative analysis in section 3.1 has shown comparable effectiveness between the MSL and the TT approaches, the qualitative data gathered from the post-activity questionnaire offer us valuable insights into the participants' emotional engagement, perceived learning outcomes and self-awareness of language improvement.

Overall, the participants expressed a predominantly positive impression of the MSL intervention. Out of the total responses to the eight questions, 173 were positive, 139 neutral and only 20 negative. This indicates general acceptance and appreciation of the approach. The participants valued the structured yet interactive nature of the read-aloud tasks, the opportunity for self-monitoring and improvement, and the low-pressure environment that encouraged experimentation and learning.

In terms of emotional engagement and confidence-building, most of the participants described the MSL training as enjoyable: 17 reported positive feelings while reading aloud, 17 responded neutrally, and only 7 expressed negative feelings. According to the results, initial embarrassment or anxiety often gave way to increased comfort and even enjoyment as the activity progressed. One participant noted, "At first a little bit embarrassing but when you start… you forget about other people listening." This transition from anxiety to confidence was echoed in final reflections, where 29 participants expressed positive sentiments about completing the tasks, often citing increased confidence and a sense of accomplishment.

Pronunciation emerged as both a challenge and a perceived area of improvement. Many participants acknowledged that repeating words or sentences aloud contributed to more accurate pronunciation and improved fluency. As one student commented, "Reading aloud helps improving my pronunciation and being more fluid and confident." These reflections are in line with the pedagogical goals of the MSL approach, which emphasizes phonological awareness and multisensory engagement.

Regarding text pronunciation and comprehension, responses were more varied. While 24 participants reported that the reading-aloud activities improved their understanding of the text, 16 considered its impact neutral. Some participants noted that invented or unfamiliar words were particularly difficult to pronounce and to understand. This suggests that the cognitive load involved in pronunciation may sometimes interfere with semantic processing. However, many participants still appreciated the novelty of the tasks and found them engaging. In particular, sentence-level exercises, which provided contextualized practice that supported both syntactic and prosodic development, were considered by the participants as especially useful.

Last but not least, the MSL training appeared to promote critical reflection on individual language learning processes. Responses to the question about self-discovery revealed increased awareness of the participants' individual pronunciation challenges, emotional reactions to speaking aloud, and renewed motivation to improve. One participant reflected, "I learned that I have to put aside the embarrassment," while another acknowledged gaps, "I don't have any idea of how to pronounce by my own."

To sum up, the participants' feedback reflects the affective, cognitive and metacognitive benefits of the MSL intervention and shows its potential as a meaningful and learner-centered tool in EFL pronunciation training (RQ3).

4. DISCUSSION AND CONCLUSIONS

The primary purpose of the present study was to investigate the effectiveness of an AI-enhanced MSL intervention for EFL pronunciation in Spanish university students, comparing it with TT, focusing on accuracy and learner experience. Following a pre-test, a training session, and a post-test, quantitative results demonstrated comparable effectiveness between MSL and TT in terms of training effect (RQ1), with no statistically significant difference observed regardless of the word type (RQ2). In contrast, qualitative results highlighted significantly positive learner experiences with the MSL intervention.

The MSL intervention proved as effective as the TT approach in improving EFL pronunciation accuracy across the five targeted vowel and diphthong phonemes (i.e., /ə/, /uː/, /ʌɪ/, /eɪ/, and /ɔː/). The absence of a statistically significant difference indicates that both methods facilitated learning, reinforcing that explicit pronunciation training is indeed effective. However, this finding, especially when viewed alongside the qualitative data, suggests that MSL presents a highly preferable approach due to its potential to notably increase student motivation and engagement while achieving comparable learning outcomes (RQ3).

The lack of a statistically significant difference in improvement between the two groups based on word type might be attributed to the brevity of the training session (RQ1 & RQ2). Deeper changes in cognitive processes, such as those related to sublexical and lexical pathways, could require a longer period of intervention to manifest measurable differences. Nevertheless, this intervention directly targeted the mechanisms highlighted by the DRC model, emphasizing the importance of strengthening the sublexical route—a crucial aspect for Spanish speakers navigating English's opaque orthography.

The MSL intervention was exceptionally well-received by students, as reflected by the 173 positive responses (RQ3). The qualitative data obtained through the questionnaire provided crucial context for the quantitative findings. Students consistently reported enjoying the training and noted its effectiveness in reducing their embarrassment or anxiety. These results strongly suggest that the affective filter (Krashen, 1982) was significantly lowered, thereby facilitating learning. Participants perceived tangible learning gains, which in turn boosted their confidence and enhanced their overall learning experience. This process also promoted critical self-reflection and increased awareness of individual pronunciation challenges.

In summary, beyond its demonstrated effectiveness in EFL pronunciation improvement, a key strength of this intervention lies in its enjoyable nature for Spanish students. Its engaging and motivational character undeniably contributed to a lowered affective filter after the MSL session. One of the main limitations of the present study was its constrained duration, restricted to a single 50-minute session. Another limitation pertains to the sample size, which was limited to university students from specific fields. However, this intervention can be replicated with larger and more diverse learner populations (e.g., different age groups, proficiency levels, academic fields).

This innovative project, therefore, serves as a foundational step. Looking forward, the aim is to replicate this study across different age groups and transform it into multi-week training sessions, allowing for a more thorough investigation into the long-term impact on cognitive processing, pronunciation accuracy, and overall communicative confidence. By further refining and expanding such AI-enhanced MSL approaches, this research contributes to shaping the future of EFL education, providing evidence-based strategies to foster more effective and confident English students globally.

5. FUNDING

This study was supported by funding from the University of Valladolid through the Innovative Teaching Project "Mejora de la lectura en inglés mediante IA y Metodologías Multisensoriales en Estudiantes con y sin Dificultades de Aprendizaje," during the 2024-2025 academic year.

6. AUTHORS' CONTRIBUTIONS

Conceptualization, D. C. T. and Q. Y.; data curation, D. C. T. and Q. Y.; formal analysis, D. C. T. and Q. Y.; funding acquisition, D. C. T.; investigation, D. C. T. and Q. Y.; methodology, D. C. T. and Q. Y.; project administration, D. C. T.; resources, D. C. T. and Q. Y.; validation, D. C. T. and Q. Y.; visualization, Q. Y.; writing—original draft preparation, D. C. T. and Q. Y.; writing—review and editing, D. C. T. and Q. Y..

7. REFERENCES

Ai, Y., Hu, Y., & Zhao, W. (2025). Interactive learning: harnessing technology to alleviate foreign language anxiety in foreign literature education. *Interactive Learning Environments*, *33*(3), 2440–2459. https://doi.org/10.1080/10494820.2024.2409990

Calvo Benzies, Y. J. (2016). *The teaching and learning of English pronunciation in Spain: An analysis and appraisal of students' and teachers' views and teaching materials*. [Doctoral dissertation, Universidade de Santiago de Compostela]. MINERVA. http://hdl.handle.net/10347/13862

Calvo Benzies, Y. J. (2017). Advanced Spanish university students problems with the pronunciation of English vowels. Identification, analysis and teaching implications. *ODISEA, 14*, 37–56. https://doi.org/10.25115/odisea.v0i14.256

Cao, F., Kim, S. Y., Liu, Y., & Liu, L. (2014). Similarities and differences in brain activation and functional connectivity in first and second language reading: Evidence from Chinese learners

of English. *Neuropsychologia*, *63*, 275–284. https://doi.org/10.1016/j.neuropsychologia.2014.09.001

Casillas J. (2015). Production and perception of the /i/-/I/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica, 72*(2-3), 182–205. https://doi.org/10.1159/000431101

Chapuis, N. D., & Berthele, R. (2024). Teaching methods emphasizing phonological forms enhance L2 vocabulary learning. *Language Teaching Research*. https://doi.org/10.1177/13621688241270803

Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, *4*(1), 5–17. https://doi.org/10.1080/13603110600574322

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

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

Ernesto Macaro, L. E. (2012). Decoding ability in French as a foreign language and language learning Motivation. *The Modern Language Journal, 95*(4), 496–518. https://doi.org/10.1111/j.1540-4781.2011.01238.x

Fadillah, A. C. (2020). Pronunciation difficulties of EFL learners in Indonesia: A literature study. *Jurnal Budaya FIB UB, 1*(2), 7–13. https://jurnalbudaya.ub.ac.id/

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology, 24*(1), 62–85. https://doi.org/10125/44709

García, P. B., & Froud, K. (2018). Perception of American English vowels by sequential Spanish–English bilinguals. *Bilingualism: Language and Cognition, 21*(1), 80–103. https://doi.org/10.1017/S1366728916000808

Gómez González, M. de los Á., & García Muras, R. (2025). EFL pronunciation instruction in Spanish primary schools: From prescribed curriculum to classroom practice. *Education Sciences*, *15*(4), 427. https://doi.org/10.3390/educsci15040427

Hsiao, J. C., & Chang, J. S. (2023). Enhancing EFL reading and writing through AI-powered tools: design, implementation, and evaluation of an online course. *Interactive Learning Environments, 32*(9), 4934–4949. https://doi.org/10.1080/10494820.2023.2207187

Kast, M., Meyer, M., Vögeli, C., Gross, M., & Jäncke, L. (2007). Computer-based multisensory learning in children with developmental dyslexia. *Restorative Neurology and Neuroscience*, *25*(3–4), 355–369. https://doi.org/10.3233/RNN-2007-253417

Kelly, K., & Phillips, S. (2022). *Teaching literacy to learners with dyslexia: A multisensory approach* (3rd ed.). Corwin UK.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Pergamon Press.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–388. https://doi.org/10.2307/3588485

Li, P., Baills, F., Alazard-Guiu, C., Baqué, L. & Prieto, P. (2023). A pedagogical note on teaching L2 prosody and speech sounds using hand gestures. *Journal of Second Language Pronunciation, 9*(3), 340–349. https://doi.org/10.1075/jslp.23043.li

Łockiewicz, M., Jaskulska, M., & Fawcett, A. (2020). Decoding and word recognition in English as a native and a foreign language in students with and without dyslexia (English vs. Polish students). *Dyslexia, 26*(1), 18–35. https://doi.org/10.1002/dys.1648

Mlakar, H., Hirst-Plein, J. & Koch, M. J. (2025). Cognitive and linguistic influences on EFL real word and pseudoword spelling: Predictors and error analysis. *Languages, 10*(5), 93. https://doi.org/10.3390/languages10050093

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, *79*(1), 1–43. https://doi.org/10.1515/PHON-2022-2018,

Nikbakht, H. (2011). EFL pronunciation teaching: A theoretical review. *The Journal of Applied Linguistics*, *4*(1), 146–176.

Odlin, T. (2012). *Language transfer.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524537

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives.* Palgrave Macmillan.

Permatasari, S. (2024). Enhancing pronunciation skills in EFL students through the ELSA speak application. *Indonesian EFL Journal, 10*(2), 205–212. https://doi.org/10.25134/ieflj.v10i2.10137

Rallo, L. (2022). Factors affecting pronunciation accuracy in English as a foreign language: The case of Spanish Catalan intermediate learners. *Atlantis. Journal of the Spanish Association for Anglo-American Studies, 44*(2), 45–64. https://doi.org/10.28914/Atlantis-2022-44.2.03

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 787–799. https://doi.org/10.1002/wcs.68

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly, 55*(3), 866–900. https://doi.org/10.1002/tesq.3027

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

Sammour-Shehadeh, R., Prior, A., & Kahn-Horwitz, J. (2025). Spelling challenges in English as a foreign language: vowels, digraphs, and novel phonemes. *Reading and Writing*, 1–24. https://doi.org/10.1007/s11145-025-10659-3

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Selim, A. S. M. (2024). The transformative impact of AI-powered tools on academic writing: Perspectives of EFL University Students. *International Journal of English Linguistics*, *14*(1), 14. https://doi.org/10.5539/IJEL.V14N1P14

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Setyaningsih, E., Asrori, M., Ngadiso, Sumardi, Zainnuri, H., & Hariyanti, Y. (2024). Exploring high school EFL teachers' experiences with magic school AI in lesson planning: Benefits and insights. *Voices of English Language Education Society*, *8*(3). https://doi.org/10.29408/VELES.V8I3.27700

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Snowling, M. J., Hulme, C., & Nation, K. (Eds.). (2022). *The science of reading*. Wiley Blackwell. https://doi.org/10.1002/9781119705116

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1260843

Veiga Pérez, C. (2017). Spanish teenager's pronunciation of English as a second language. *Revista de Educación de La Universidad de Granada, 24*, 273–292.

Yang X., Shi F., Liu X., Zhao Y. (2016). Learning styles and perceptual patterns for English /i/ and /l/ among Chinese college students. *Applied Psycholinguistics, 37*(3), 673–701. https://doi.org/10.1017/S014271641500020X

Yuan, L., & Liu, X. (2025). The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. *Computers in Human Behavior*, *162*, 108474. https://doi.org/10.1016/j.chb.2024.108474

Zhang, Z., & Peng, P. (2022). Reading real words versus pseudowords: A meta-analysis of research in developmental dyslexia. *Developmental Psychology*, *58*(6), 1035–1050. https://doi.org/10.1037/dev0001340

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Zlatkova-Doncheva, K. (2022). A multisensory approach to teaching students with dyslexia. *Pedagogical Almanac*, *30*(1), 70–78. https://doi.org/10.54664/xuav9711

Cite this work:

Carrascal Tris, D., & Yuan, Q. (2025). AI-Enhanced Multisensory Learning Intervention to Improve Pronunciation in EFL Students . *Edutec, Revista Electrónica de Tecnología Educativa*, (93), 233-250. https://doi.org/10.21556/edutec.2025.93.4197