Lengua y Sociedad. Revista de Lingüística Teórica y Aplicada,
Vol. 23, n.º 2, julio-diciembre 2024, pp. 963-983,
ISSN-L 1729-9721; eISSN: 2413-2659
https://doi.org/10.15381/lengsoc.v23i2.26074

Speech Recognition Software as a Tool to Enhance EFL Learners' Pronunciation

Software de reconocimiento de voz para mejorar la pronunciación del inglés

Software de reconhecimento de voz para melhorar a pronúncia do inglês

Laura Sánchez
Universidad El Bosque, Bogotá, Colombia
sanchezlaura@unbosque.edu.co
https://orcid.org/0000-0001-9449-5259

Andrés Morales
Universidad El Bosque, Bogotá, Colombia
anmoralesc@unbosque.edu.co
https://orcid.org/0009-0001-7819-7566

Ingrid Rodríguez
Universidad El Bosque, Bogotá, Colombia
rodriguezingrid@unbosque.edu.co
https://orcid.org/0000-0001-8121-4008

Abstract

Through technological advancements, challenges and opportunities have emerged in various fields. Language classrooms are some of the most impacted settings, as students have been able to integrate technological resources into their learning process. One of the areas where technology has proven to be relevant for language learning is pronunciation. Although instruction and feedback in this area are crucial, the available time in the classroom for its development is often limited. For this reason, this study analyzes the impact of using speech recognition software (SRS) as a complementary tool for teaching English pronunciation as a foreign language (EFL). Methodologically, the design is qualitative. The sample consists of 10 basic-level EFL students from a language center at a university in Bogotá, Colombia. Data were collected through field notes, audio and video recordings, as well as student journals. Among the findings, it was discovered that SRS feedback contributes to the improvement of students' pronunciation; moreover, it motivates their commitment to learning and promotes their autonomy.

Keywords: autonomy; feedback; language instruction; pronunciation; speech recognition software.

Resumen

A través de los avances tecnológicos han surgido desafíos y oportunidades en diferentes campos. Las aulas de idiomas son algunos de los escenarios más impactados, puesto que los estudiantes han logrado integrar recursos tecnológicos en su proceso de aprendizaje. Uno de los aspectos en los que la tecnología ha demostrado ser relevante para el aprendizaje de lenguas es la pronunciación. Aunque la instrucción y la retroalimentación en este ámbito son cruciales, el tiempo disponible en el aula para su desarrollo suele ser limitado. Por esta razón, en este estudio se analiza el impacto del uso del software de reconocimiento de voz (SRV) como herramienta complementaria para la enseñanza de la pronunciación del inglés como lengua extranjera (ILE). Metodológicamente, el diseño es cualitativo. La muestra está constituida por 10 estudiantes del nivel básico de ILE de un centro de idiomas de una universidad en Bogotá, Colombia. Los datos se recopilaron por medio de notas de campo, grabaciones de audio y video, así como mediante los diarios de los estudiantes. Entre los resultados, se encontró que la retroalimentación del SVR contribuye a la mejora de la pronunciación de los estudiantes; además, motiva su compromiso con el aprendizaje y promueve su autonomía.

Palabras clave: autonomía; enseñanza de lenguas; pronunciación; retroalimentación; software de reconocimiento de voz.

Resumo

Através dos avanços tecnológicos, surgiram desafios e oportunidades em diversos campos. As salas de aula de idiomas estão entre os cenários mais impactados, pois os alunos conseguiram integrar recursos tecnológicos ao seu processo de aprendizado. Um dos aspectos em que a tecnologia tem se mostrado relevante para o aprendizado de línguas é a pronúncia. Embora a instrução e o feedback nesse âmbito sejam cruciais, o tempo disponível na sala de aula para seu desenvolvimento costuma ser limitado. Por essa razão, este estudo analisa o impacto do uso do software de reconhecimento de voz (SRV) como ferramenta complementar para o ensino da pronúncia do inglês como língua estrangeira (ILE). Metodologicamente, o design é qualitativo. A amostra é constituída por 10 alunos do nível básico de ILE de um centro de idiomas de uma universidade em Bogotá, Colômbia. Os dados foram coletados por meio de notas de campo, gravações de áudio e vídeo, bem como diários dos alunos. Entre os resultados, constatou-se que o feedback do SRV contribui para a melhoria da pronúncia dos alunos; além disso, motiva seu compromisso com a aprendizagem e promove sua autonomia.

Palavras-chave: autonomia; ensino de línguas; feedback; pronúncia; software de reconhecimento de voz.

Received: 08/23/2023 Accepted: 07/29/2024 Published: 12/30/2024

1. Introduction

In the last decade, technology has brought about a significant change in both life and education, and one of the most affected areas has been second language (L2) teaching, where its impact has been clearly evident. Thanks to the great technological development and artificial intelligence IA advances, teachers and students have had the chance to explore new strategies and resources to teach and learn an L2 in didactic and innovative ways, where one of the new possibilities has been related to the oral skill development. In this regard, there are different studies in which the use of speech recognition software (SRS) in L2 classrooms has been explored, and they have shown its impact on aspects such as pronunciation.

Pronunciation is defined as the combination of phonemes and suprasegmental traits (Kelly, 2000) which are key now to use an L2; its learning often requires clear phonetic instructions, as well as constant practice and feedback which nurtures and guides the students' reflection on their performance during the learning process (Brookhart, 2008). However, instruction and feedback has tended to be scarce, due to the lack of resources and time to carry out individual assistance. Nevertheless, the recent boom of technological advances has allowed L2 teachers to find new ways to approach pronunciation and support their students' processes. In this sense, SRS has been considered an innovative tool that can support students' pronunciation enhancement; some of the most popular SRS are Siri by Apple, Google assistant and Amazon's Alexa, among others (Evers and Chen, 2021),

which are within reach of a click and easy to use. Certainly, SRS have been more and more used inside and outside the L2 classrooms (Asratie *et al.*, 2023; Çelik and Baran, 2022).

According to what has been pointed out before, this study was carried out to support a group of basic EFL university students from a language center who presented difficulties regarding their oral skills development, specifically, pronunciation. Instruction they used to receive seemed to be too general and scattered through different moments of a class; hence, it did not really meet the students' needs, nor did it foster their awareness of the necessity to achieve proper pronunciation for effective and clear communication. Then after the first evaluation the learners had to take, which included the four communicative skills, most of the students seemed insecure when taking the speaking test. Recurrent issues such as mispronunciation of individual words and lack of oral fluency were evidenced, which negatively affected their learning process and academic results.

For this reason, the study originated from the necessity to implement a tool to support the instruction on pronunciation that the group of basic EFL learners required; furthermore, due to the necessity of finding one that could be used in and outside the classroom. On this basis, instant feedback that SRS provides in addition to its ease of use made it the chosen tool to support their learning process. Therefore, the research question that guided this study was the following one: What is the effect of using SRS instant feedback as a tool to support instruction on pronunciation in a group of EFL basic learners?

This research is justified due to the impact it could have on a local and broader context. First, EFL students may have the possibility of improving their pronunciation skills using new technologies such as SRS, available in any device, both inside and outside the classroom no matter the language level they have reached. Thus, when providing comments, SRS can empower learners to take charge of their learning process and become autonomous. Secondly, not only EFL teachers, but also any L2 educators could have the opportunity of contributing to the enhancement of students' learning process. It is not a secret that instruction on pronunciation is subjective to the educators when they are teaching and assessing it (Hazen, 2020); therefore, SRS may empower them to carry out innovative practices and meet this type of language needs. Additionally, schools and higher education institutions can also benefit from this project through the implementation of L2 learning strategies and new curricular designs that include this type of technology, which can lead to a better quality of education in this area.

2. Theoretical framework

2.1. Pronunciation

According to Derwing and Munro (2015), pronunciation is the way people use their vocal tract to produce sounds and as a consequence, speech. It encompasses two main features: phonemes and suprasegmentals. Phonemes are the specific sounds characteristic of a language; although they can vary from person to person, or even from a geographical region to another, there are certain traits that are unique to each language (Kelly, 2000).

First, it is important to understand the previously stipulated concepts. Phonemes are the segments referred to the different sounds characteristic of any language. These allow us to differentiate the meanings among words; despite the way speakers from different geographical

areas might slightly change the pronunciation of each sound (Kelly, 2000). On the other hand, the suprasegmental features of a language can be found; according to Lyabode (2011), these traits go beyond phonemes, since they help messages get across accurately since they intervene at the syllabic level. These characteristics are stress, rhythm, intonation, and pitch.

Regarding this area, there have been several studies about the student and teacher's role along the pronunciation teaching and learning processes. To begin with, Plailek and Essien (2021) state that three factors are essential for a successful pronunciation development; first, the importance of an accurate and acceptable non-native teacher's pronunciation, then the accuracy of materials and finally an appropriate teaching method. Additionally, Ahmad (2018) identifies the techniques and activities necessary to help students learn suprasegmental features such as intonation and stress. The author reached three important conclusions. First, the appropriate set of activities, materials and procedures are pivotal for acquiring proper stress and intonation. Second, there are two important traits regarding syllables: adjectives and nouns are generally stressed in the first syllable, and words with prefixes are stressed in the second syllable but never in the prefixes or suffixes. Finally, intonation plays an important role in the accurate interpretation of the message the speaker is trying to convey. For example, there is a falling intonation when the speaker is giving commands, stating ideas or asking wh-questions; when the speaker raises intonation, it is usually to ask yes/no questions or to present a list of items.

Lastly, Vahdany *et al.* (2022) establish that the learning strategies such as cognitive, metacognitive and socioaffective are effective for the learning of pronunciation at some levels, what makes the difference is the quality of implementation. An important conclusion from this study is that textbook materials would be more successful if they contained more communicative-focused activities which could engage students in a more meaningful way. Additionally, it is stated that teachers at basic levels ought to dedicate time to point out the harmfulness of some strategies that can lead to fossilizing bad practices. All factors considered above support the validity of teaching and learning pronunciation.

2.2. Feedback

Formative assessment has proven to be an essential skill to be developed by teachers and Feedback is one of its most important components, thus, it is key to empowering students over their learning process. According to Brookhart (2008), feedback fosters cognitive development and motivation. Effective feedback helps students understand where they are in relation to the learning objectives and what they need to do to ameliorate; as soon as students are able to understand the reasons why they need to enhance a particular aspect and know how they can do it, this increases their motivation. All in all, feedback is pivotal in helping students to improve any area of knowledge.

Another way of providing feedback is through digital formative assessment (Çetin, 2021). Nowadays, learners need to be motivated, so by using online tools, teachers have the possibility to engage them through innovative and current feedback strategies. Hence, technology-mediated learning is a must in a globalized world where these tools prevail in most areas, including education, which cannot be neglected.

Regarding pronunciation, providing effective feedback has two main advantages (Sukiyadi and Purnawarman, 2021). One of them is allowing students to improve their weaknesses and realize their strengths. Through this process teachers give students the opportunity to reinforce and refine their pronunciation skills by means of clear evaluation criteria. Second, learners' L2 could be enhanced due to the new knowledge obtained during the feedback sessions. When feedback is given students can see their mistakes, thus they modify and adapt their fossilized errors according to the new information received.

Additionally, according to Wang *et al*. (2017), teachers need to find support from professional development resources such as seminars, lectures and specialization courses in order to receive proper training regarding feedback. In addition, dialogue with peers is fundamental to nurture novice teachers' feedback practices. Moments to have such conversation should be encouraged and fostered in order to strengthen them.

2.3. Speech recognition software

Nowadays, there are many electronic tools that help students to foster and improve their skills in any language; for example, webpages, apps or devices that are part of our daily life. Among these technologies, it is possible to find speech recognition software (SRS) such as Apple's Siri, Google's assistant, Microsoft's Cortana and their Flip app among others. This software processes the information by decoding and transcribing it. It perceives the voice through a microphone and generates a written script by means of algorithms, patterns or models (Levis and Suvorov, 2012).

According to Yokota *et al*. (2022), this technology provides written transcripts for users to see. Although their accuracy is not 100 %, they are precise enough to provide some form of corrective feedback and thus enhancing their experience practicing their pronunciation. This technology has proven to be effective in detecting pronunciation errors, and as a consequence, standardized international tests such as TOEFL or PTE have been using these resources in order to score the speaking section of their tests (Wung *et al*., 2022). Moreover, SRS could be considered a classroom strategy due to its effectiveness to engage students and consequently its potency to foster autonomy to practice and reinforce pronunciation even when the classes have finished. Additionally, students feel more comfortable while they use this tool, because it lowers their level of anxiety when they are exposed to interaction with their peers (McCrocklin, 2016).

3. Methodology

This research has a qualitative approach since it tries to develop the subject of study from a holistic outlook. This approach requires researchers to examine the data with the assumption that nothing is unimportant; thus, every single detail matters as it could potentially throw some light on the comprehension of the phenomena (Bogdan and Biklen, 1997). Likewise, this study follows an action research design since it is necessary to proceed following the phases of planning, acting, observing and reflecting, that this design proposes in order to foster the enhancement of the ELF students' pronunciation (Zuber-Skerritt and Wood, 2019).

Regarding the previously presented characteristics, the participants of this study were a group of 10 A2 EFL students from a university language center in Bogota, Colombia, who studied on Saturdays

from 8am to 1pm. The language program that they were offered follows the guidelines of the Common European Framework of Reference (CEFR, 2022). It involves six levels from A1 to C1, where students need to achieve B1, B2 or C1 of English proficiency to graduate according to their different majors. Every level is studied during an academic semester, which is divided into three terms that include four weeks of class. After each term, students' language command is assessed during a class session, including their oral production, and they receive feedback. Additionally, students from that institution do not receive formal phonetic instruction; they do not usually know about the phonetic alphabet and how to read phonetic symbols. Convenience sampling was implemented (Stratton, 2021) since, after informing the research purpose to the whole group (17 in all), 10 subjects expressed their willingness to participate and got involved in all the activities that were proposed.

Therefore, the participants and the language center where this study took place were previously informed about the goal and procedures that this research would involve, so that their permission to gather and analyze data besides publishing the results was granted. Additionally, the anonymity and confidentiality of the learners was ensured by assigning them some codes that were established during the processes of data collection and analysis, which are reported in this document.

There were three instruments to collect data: audio and video recordings, field notes and students' journals. The use of recordings provided reliable information regarding students' performance during the pedagogical implementation; they were transcribed and codified to be analyzed. Likewise, field notes from every session enabled the researchers to focus on details that were not possible or easy to record; they were taken in the form of a fluent text as well as tables in order to facilitate the data gathering. Finally, students' journals provided valuable information about students' perceptions, which contributed to recognizing the effect that feedback provided by SRS may have on their pronunciation.

3.1. Pedagogical implementation

This study was carried out during an academic semester in which students worked on their EFL pronunciation. Two cycles of implementation were conducted (see Figure 1.), each one of four class sessions. First, the teacher presented the vocabulary that was going to be used during the academic term and emphasized its pronunciation. Every cycle included about thirty words from the A2 level, based on the CEFR (2022), which were going to be used during the next lessons.

Second, students check the vocabulary meaning and practice its pronunciation. To do so, at the beginning of the implementation, students were asked to set up their mobile phones so that they could start using SRS: Siri, Google Assistant and WhatsApp Dictation; this was done under the teacher's supervision. For the pronunciation practice, students had to command Siri and Google Assistant to perform certain actions; the commands given were "Hey Siri! (Hey Google) show me a picture of…", "Hey Siri! (Hey Google) What's the meaning of…", yet when using WhatsApp, they practiced orally dictating written notes; as they practiced each word, they received feedback from the SRS, which involved carrying out the actions they had commanded, asking them to try saying it once more, telling them it had not been understood, or typing a text that they could review to confirm that the words identified by the SRS were the ones they intended to pronounce. Meanwhile, the teacher moved around the classroom to guide the students. Third, once they had

practiced and used two of the SRS previously mentioned, they recorded the list of words using another SRS called Flip; it provided captions of the oral performance, which worked as feedback; this activity was monitored, and field notes were taken. Fourth, the teacher provided feedback on their pronunciation, and she helped students solve their doubts about the SRS. Once the pronunciation of words was rehearsed, students answered some questions from their journals in order to gather their perception of the activity and their performance.

Fifth, the students worked in pairs to practice a conversation which included the words previously given. Once they felt sure of their pronunciation, they recorded a video with the SRS Flip. The videos were presented in class while the teacher took notes on their performance. Sixth, the SRS provided feedback by means of captions which let the learners notice their successes and mistakes, and then, the teacher provided feedback and solved doubts. And at the end of the activity, the students answered the questions in their journals. Seventh, students were asked to have a spontaneous conversation in situ. They were provided the context and the purpose of the conversation so that they were able to create it in pairs and perform it; the teacher recorded the dialogue by using Flip so that students could receive feedback and identify their mistakes. Eighth, the teacher and the students reflected on the captions provided by Flip, feedback was delivered, and student's journals were filled. At the end, students recorded the list of words for the last time to identify their pronunciation accuracy with the help of the teacher's feedback.

During the data collection procedure, information (see Table 1) was reviewed and codified according to the instruments used, the students involved in the process and the cycle that took place. Then, codes were established as follows:

Table 1
Data codification

Transcript = T1 T4	Field Notes = FN1 - FN4	Students' journals = J1-J4
Students = S1 - S10	Cycles = C1, C2	

Note. The instruments were numbered based on the times they were registered in each cycle.

3.2. Data analysis and methods

To analyze the data collected, the principles of the grounded theory were followed (Bryant and Charmaz, 2007). Accordingly, no categories were pre-established, they emerged from the data in the form of meaningful topics that tended to frequently appear. Likewise, methodological triangulation was implemented, it allowed us to crossed-check the information in order to validate the emerging categories (Bryman, 2012).

4. Analysis

The research question that guided this study was the following: What is the effect of using SRS instant feedback as a tool to support instruction on pronunciation in a group of EFL basic learners? Therefore, to explain the effect of this implementation, two categories emerged; the first category

is performance, which included two subcategories: students' levels of accuracy and Students' perceptions of their own performance; and the second category which is called engagement.

4.1. Performance

Learners' performance, as the accomplishment of a learning activity in any specific situation (Moccozet, 2012), is directly connected to the abilities the participants developed and how they made them evident. Therefore, for the purposes of this study, performance represents the learners' results based on the two cycles of the pedagogical implementation previously described.

To characterize the learners' performance, different instruments were used. Transcripts from students' recordings and teachers' field notes were analyzed in order to identify the language accuracy they had reached during the pronunciation exercises; words and sounds that were correctly pronounced and the mispronounced ones were identified and systematized, so that it was possible to have a clear register of their process. Likewise, students were asked to write their perceptions regarding their process in their journals; their answers included the weaknesses, strengths and achievements they identified after the activities that were proposed. Therefore, by means of the data collected, it was possible to identify the level of performance in terms of pronunciation that students reached along the pedagogical implementation, which allowed us to recognize the evolution that learners presented.

4.1.1. Students' level of accuracy

During the first cycle, the vocabulary students had to work on involved thirty-six words which belonged to the A2 level according to the CEFR (2022). The teacher provided instruction on basic pronunciation, which involved some phonemes that were emphasized in their textbook, and offered guidance on the use of the SRS, so that the learners could practice the vocabulary inside and outside class. At the beginning of each cycle, they practiced the list of words provided (see Table 2) by using SRS Siri or Google assistant and WhatsApp Dictation. Learning how to command the SRS tended to be complicated at the beginning of the first cycle; the students seemed to not trust the feedback since their devices needed to be set properly or they needed to focus on the commands required. Then their level of accuracy was identified based on the number of correctly pronounced items and the most challenging sounds or phonemes that were recorded. However, something that called the attention was the fact that Flip failed to recognize some words without a linguistic context, then in order to include a register of the items that the SRS may not identify, field notes were taken.

Table 2
Sample of field notes that registered word pronunciation accuracy (FN1, C1)

Word	S1	S2	St 8
1. Asia	It was ok	It was ok	She couldn't pronounce the word and she said "ashia" /eɪ//ʒ/ /ə/
2. Egypt	He needs to work on: /iː/	It was ok	She said eject instead of **E**gypt /iː/
3. Egyptian	He needs to work on: /iː/	It was ok	She said eject instead of Egypt /iː/

Note. This information corresponds to part of the field notes that were registered.

As the lessons went by, learners also practiced the words in conversations. The first conversation was prepared, rehearsed and recorded on their own out of the class, and the second was prepared and presented in situ while being recorded by the teacher (see Table 3), so that they could use the vocabulary in context. Before recording the conversations, they used to practice and receive instant feedback about these samples by using different SRS (Siri Google assistant, WhatsApp dictation and Flip), and based on the final recording, the teacher provided more feedback to conclude the activity. Thus, it was possible to identify that the three most recurrent mispronounced phonemes were /i/ (high front, unrounded, tense vowel), /ʃ/ (high, front, unrounded, tense vowel) and /θ/ (voiceless, dental, fricative consonant), which did not have an equivalent in Spanish; the learners tended to require longer practice and constantly required the teacher's modeling of the pronunciation so that they could articulate them properly. The next table shows an example of a learners' conversation transcription; although improvement was evidenced, it was possible to identify mispronunciation of some phonemes.

Table 3
Sample of matrix with transcript of a conversation from Flip (T2, C1)

	T2 C1	Analysis
S1	*Yes, there is a *ship* 1-12 blocks away, but the place is *inside*.", "I think they are open. *just*, Wednesday, *Thursday*,Fridays and during the weekend. But I don't really recommend it", "The place is ugly and the food is not very good", "no problem. Bye, bye."	In general, the pronunciation of these 2 students was good; their speech was understandable and they managed to pronounce most of the words correctly. However, there were some issues with the words "cheap", "Thursday", "nephew". The issues observed in these words are related to the voiceless sound /θ/, the common confusion of the /tʃ/ and /ʃ/ sounds, the long vowel /i:/ as well as the vowel sound /e/ in "nephew". Other than those issues, their pronunciation is clear.
S2	"Excuse me, do you know any Irish restaurants around here?", "Oh, that's great. Do you know *its* schedule?", "Seriously. Why not? I want to take my uncle and my *nephew* there", "??? Your advice? I'll just go somewhere else". "Have a good day. Bye!"	

Note. This information corresponds to part of the matrix that included the transcripts.

Later, the students recorded the list of words one more time and received instant feedback from the SRS and the teacher again; this process was also registered in the teachers' field notes (see Table 4). An important improvement was noticed; mistakes were not as frequent as in the previous sessions and learners tended to speak with more confidence. The next sample exemplifies part of the results of S1 and S2' pronunciation from the final exercise.

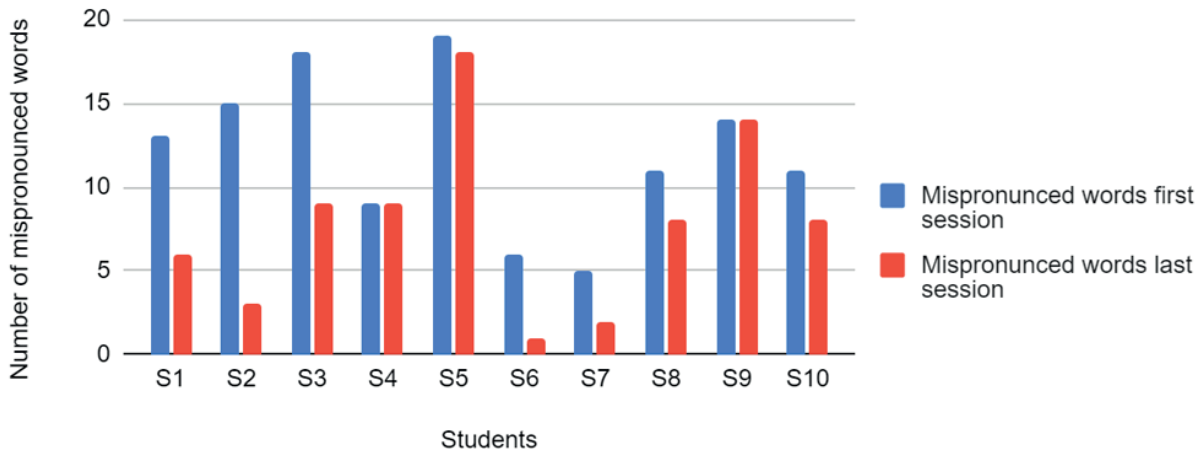
Table 4
Sample of field notes from the last session of the first cycle (FN3, C1)

Word	S1	S2
Birthday	bir**th**d**a**y = birtey /θ/ /eɪ/	It was ok
Comfortable	it was ok	It was ok
Valuable	Val**ua**ble = problem with /æ/ /bəl/	It was ok
Weather	It was ok	It was ok
Bridge	It was ok	bradge /ɪ/

Note. This information corresponds to part of the field notes that were registered.

Thus, the data collected was triangulated to identify the learners' levels of accuracy and the difficulties that they faced when working on pronunciation. Figure 1 visually displays the results of the learners' pronunciation based on the number of mistakes that were identified in the first and the last recording of the vocabulary and the field notes. Most of the students showed pronunciation improvement, and just two students remained at the same level of accuracy; one of them was a student who had a speech pathology, which could be the reason for her not reaching the pronunciation goals.

Figure 1
Level of accuracy at the beginning and at end of the first cycle based on the number of mispronounced words



Note. The data registered includes the number of mispronounced words evidenced from the transcripts and the field notes.

Besides, an analysis made of these words allowed us to identify the phonetic sounds that seemed more difficult to pronounce during this cycle. Vowels /i/ and /I/, among others, were the most challenging ones; it tended to be difficult for them to differentiate from long to short vowel sounds. Additionally, consonant /θ/ remained one of the most difficult consonant sounds; its place and mode of articulation required constant teacher's modeling as well as practice with the SRS. The next tables allow a visual representation of the most difficult phonemes for students to produce at the end of this cycle.

During the second cycle, students worked on a set of thirty words (see Table 5), so they received instruction on pronunciation, practiced the vocabulary by means of SRS and recorded the list of words for the first time. This time some students were able to achieve the correct pronunciation of some words at the beginning of the cycle. Something interesting was the fact that phonemes such as /i/, which were very complicated for the students to pronounce in the first cycle, were correctly pronounced in words such as beach and ceiling. Nevertheless, words which included the phoneme /θ/, such as bathtub, seemed to keep being difficult for most of the students.

Table 5
Sample of field notes from the first session of the second cycle (FN1, C2)

Word	S8	S9	S10
Beach	She needed to practice several times before she got it correctly.	It was ok	It was ok
Dining room	It was ok	It was ok	It was ok
Ceiling	It was ok	It was ok	It was ok
Bathtub	Ba**th**tu**b**= bus stop (sounds /æ/ /θ/ and /ʌ/).	Ba**th** tu**b**= (/æ/ /θ/)	Ba**th** tu**b**= (/æ/ /θ/ /b/)
Stove	sto**ve** = stop problem with /əʊ/; it was pronounced /o/; Spanish interference).	She didn't say the word	sto**ve** = stop problem with /əʊ/; it was pronounced /o/; Spanish interference).
Behind	B**e**hind = /ɪ/	It was ok	It was ok

Note. This information corresponds to part of the matrix that included the transcripts.

As students practiced the pronunciation, their performance evidenced improvement (see Table 6). After rehearsing isolated words, their practice involved the two conversations that were part of the cycle. Students seemed more confident as they had already practiced using the SRS autonomously in and outside class and knew how the activity was. SRS feedback showed them the need for emphasis on the pronunciation of long and short vowels as well as fricative consonants. The following sample illustrates the results of the second conversation.

Table 6
Sample of matrix with transcript of a conversation from Flip (T3, C2)

	T3 C2	Analysis
S8	Hi, I Have told all the three activity. Have we come to you like? Face or we can going to Iceland. Second, we can going to beach and swim and three will come play video games on Maria is birthday. No, I had no book. Book on. OK. Yeah.	*first *island *swim *Birthday *weather 5 out of the 10 words were pronounced correctly. The words with an incorrect pronunciation were related to sounds such as /ɜ:/, /θ/, the mispronunciation of /e/ and the addition of /e/ to the word "swim".
S9	Hi, Annie. Yes, I love it. I want. I wanted to travel for a long time. I I hope we will. And have a good weather. And did you book a flight?	

Note. This information corresponds to part of the matrix that included the transcripts.

Next, in Table 7, after receiving feedback from the SRS and from the teacher, students practiced the sounds they had difficulty with during the week before the last session. Then in class, they were asked to record the list of words once more to get a final register.

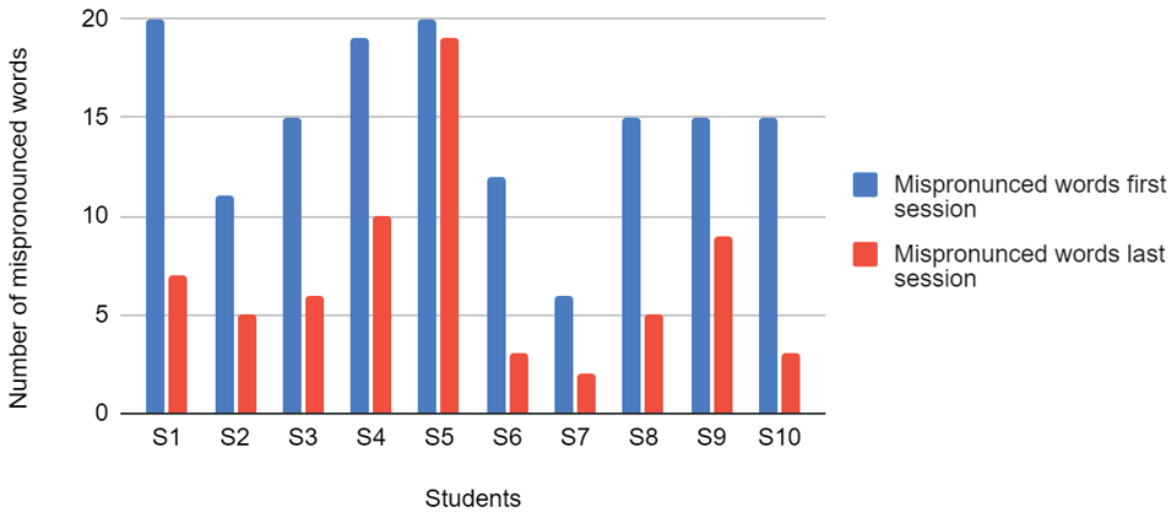
Table 7
Teachers field notes from the last vocabulary recording (FN4, C2)

Word	S8	S9	S10
Beach	It was ok	It was ok	It was ok
Dining room	It was ok	It was ok	It was ok
Ceiling	ceiling: sailing	It was ok	It was ok
Bathtub	It was ok	It was ok	It was ok
Stove	stove = stop problem with /əʊ/; it was pronounced /o/; Spanish interference).	stove = stuff problem with /əʊ/; it was pronounced /o/; Spanish interference).	stove = stop problem with /əʊ/; it was pronounced /o/; Spanish interference).
Behind	It was ok	Behind = /ɪ/	Behind = /ɪ/

Note. This information corresponds to part of the field notes that were registered.

Accordingly, after identifying the mistakes at the beginning and the end of the process, the results showed an improvement. This time all the learners exhibited a decrease in the number of wrongly pronounced words. What is more, the students who had not been able to significantly improve in C1, including the student with a speech pathology, demonstrated a reduction of mispronounced words (see Figure 2).

Figure 2
Performance at the beginning and at end of the second cycle based on the number of mistakes



Note. The data registered includes the number of mispronounced words evidenced from the transcripts and the field notes.

Likewise, after the analysis of the mispronounced words, the identification of the most challenging phonetic sounds during this cycle was carried out. Some vowels, such as /æ/ (low, front, unrounded, lax vowel) and /ə/ (mid, central, unrounded, lax, unstressed - Schwa), among other ones, were the most complicated for students to pronounce. Again, the fact that these sounds

do not have an equivalent in their mother tongue made it difficult for the learners to articulate them incorrectly. However, an evident improvement of the pronunciation of /i/ and /I/, among other vowels, was evidenced as well as for the consonant /θ/. The novelty of these sounds seemed to be the reason for their mispronunciation; hence, more practice was recommended in class and as autonomous work.

All in all, after implementing the two cycles and analyzing the data, a clear improvement of the students' pronunciation was observed, which shows a positive effect of the SRS feedback on the learners' levels of pronunciation accuracy. The constant practice with the help of their devices was key for them to be aware of the correct articulation of the sounds.

4.1.2. Students' perceptions on their performance

Students' perceptions from their learning process have been identified as points of reference to improve the quality of education (Syauqi *et al*., 2020). Therefore, an analysis of the perceptions that the participants had about their performance was made based on their journal entries after having received feedback from the SRS and teachers to verify how accurate they perceived their pronunciation.

Based on their answers there were students who reported having achieved the correct pronunciation of all words, others who commented having achieved the correct pronunciation of the majority of the words, and some others who mentioned they still had difficulties to reach the expected pronunciation. Nevertheless, their perceptions from the beginning to the end of the cycles showed a clear change in the levels of accuracy they perceived.

Based on the journals written during the first sessions of the implementation in C1, students tended to be concerned about their level of achievement. Some of them were very descriptive and analyzed possible reasons why they had not been able to pronounce the list of words properly. In general, they considered they presented a significant lack of knowledge on the subject matter, since they had not been aware of the correct pronunciation of the vocabulary, and also, they realized that the SRS they were using required practice to be mastered.

(1) (La producción de sonidos) se me dificultó debido a que no sabía cómo se pronunciaban [(The production of sounds) was difficult for me because I did not know how to pronounce them]. (S6, J1, C1)

(2) Fue complicado con algunas palabras ya que (el software) entendía otra cosa y otras definitivamente no las entendía [It was complicated with some words because it (the software) understood something else and it definitely didn't understand others]. (S3, J2, C1)

Nevertheless, as the sessions were implemented, the learners seemed to be more confident of their achievement, and they were even able to identify some sounds they needed to work on. Then the majority of them reported having reached the correct pronunciation of most or all the vocabulary. Besides, they recognized the support they had received from the instant feedback provided by the software.

(3) ¡Si! Pude pronunciar los sonidos propuestos en clase [Yes! I was able to pronounce the sounds proposed in class]. (S7, J3, C1)

(4) Las palabras con "th" son un poco más difíciles de pronunciar, aún así después de un tiempo el sistema las reconocía más fácil [Words with "th" are a bit more difficult to pronounce, still after a while the system recognized them easier]. (S2, J1, C2)

(5) Fue fácil, ya que teníamos bastante repaso para mejorar la pronunciación [It was easy, since we had enough revision to improve the pronunciation]. (S7, J3, C2)

In short, according to the students' perceptions, the constant practice and the attention paid to the feedback from the SRS seemed to be very useful to reach their goals, as it is shown in the previous excerpts. Therefore, although at the beginning of the C1 the learners tended to find the pronunciation difficult to achieve and the SRS complex to use, after the pedagogical implementation they realized that they could reach their goals, and that feedback received from the SRS was a useful tool to do it. Data collected from students' journals, the teacher's field notes and the recordings, revealed that, as they constantly practiced and received instant feedback from the SRS, their performance improved. It also let them become aware of their progress session by session, which was meaningful for them.

4.2. Engagement

The second category was engagement. It represents how students get involved and motivated in the class, which are important variables in the L2 classroom and impact the level of participation and interest they develop towards it (Ferrando, 2023; Macklem, 2015). Hence, it was possible to identify the way students became engaged and took part in the activities based on the reflections they wrote in their journals and the notes taken from the class observations.

First, their engagement was evident in the levels of interest that arose after using the software. Some of them recognized their automaticity and how easy it was to receive its instant feedback. Also, they recalled the certainty they had regarding their level of accuracy when using it and how meaningful it was to solve their doubts regarding pronunciation. What is more, some of them mentioned the software they preferred and explained their reasons to use it more than others. This is evident in the following excerpts.

(6) Bien. Es interesante como uno automáticamente va reconociendo cuando una palabra está mal y se corrige [Good. It is interesting how I automatically recognize when a word is wrong and the software corrects it]. (S7, J3, C1)

(7) Con Siri algunas palabras fueron más difíciles, ya que sin contexto no la reconocía. WhatsApp aquí fue más fácil, ya que la misma aplicación corregía la frase mostrando el error [With Siri some words were more difficult since without context it did not recognize them. WhatsApp was easier since the same application corrected the sentence and showed the mistake]. (S2, J1, C2)

(8) He (S6) was aware of the mistakes he had and continued to practice the sounds. (FN3, C2)

Based on the above, the commitment generated by the use of SRS also had an impact on the students' levels of autonomy and self-management. They not only reported recognizing the positive effects of constant practice with the help of the SRS, but also expressed their desire to continue using it in the future for further improvement.

(9) Seguiré mejorando y practicando ciertas palabras que aun presento dificultad con las herramientas que brindó la profesora, ya que es parte fundamental para hablar correctamente el idioma [I will continue to improve and practice certain words that I still find difficult with the tools that the teacher gave me, since it is fundamental to speak the language correctly]. (S9, J4, C2)

On the other hand, the participants' level of engagement was also manifested through the feelings that emerged during the pedagogical implementation. They reported going from frustration to satisfaction, including uncertainty since at the beginning of the process they thought the software did not work correctly. Thus, the teacher's guidance was significant for them to work on their mistakes, trust the SRS feedback, and reach a sense of achievement, which in turn nurtured their engagement. The following excerpt exemplifies their feelings.

(10) First, she (S10) thought that the app was wrong, and she was stressed. But after the teacher's feedback she could recognize her mistakes. (FN1, C1)

After applying SRS in the C1, students were able to use it without noticeable problems in the C2. However, some feelings of frustration arose again from the first sessions of C2 as they were not able to produce sounds properly; then teachers' guidance was needed again. After receiving feedback from the teacher, the constant practice in and outside class allowed students to feel more confident with the use of the software and their performance, which enabled them to reach their goals regarding pronunciation. Some of the students' perceptions in this regard are presented as follows.

(11) Muy efectiva, ayuda muchísimo la retroalimentación de la profe para dar claridad y guiarnos a mejorar. Da tips valiosos [Very effective, the feedback from the teacher helps a lot to give clarity and guide us to improve. She gives valuable tips]. (S5, J2, C2)

(12) Al principio tenía un poco de confusión con la pronunciación de algunas palabras, sin embargo, después de la retroalimentación fue mucho mejor [In the beginning I was confused with the pronunciation of some words; however, after teachers' feedback it was much better]. (S1, J1, C2)

In conclusion, receiving instruction on pronunciation supported by the use of SRS and getting constant feedback from the software and the teacher promoted the students' engagement along their learning process. They not only focused their attention on the vocabulary inside the classroom, but also tended to become more autonomous and self-regulated regarding this practice

outside it. Therefore, based on the student's reflections and class observations, their engagement was evident as well as their satisfaction.

5. Conclusions

Throughout the development of the study, the aim was to identify the effect of implementing pronunciation instruction supported by instant feedback from SRS (Speech Recognition Software) in a group of A2 EFL students. The pedagogical implementation included two cycles during an academic semester. In this regard, the results are presented below, describing the impact and effectiveness of technology-mediated learning in the educational context.

First, SRS was a significant tool to provide feedback on pronunciation since it made them aware of their level of accuracy. No matter the basic training on phonetics, this software helped them realize their mistakes while pronouncing vocabulary. As it recognized most of the words and carried out basic commands when the vocabulary was pronounced properly, learners could try producing the right sounds until they reached their goal. Additionally, SRS like Flip also provide a script, which allows a more concrete analysis of pronunciation mistakes and a more focused practice; nevertheless, it tends to omit certain mistakes or fails at recognizing the right words when there is no context, so practice should include the use of words in full sentences and conversations. Therefore, pronunciation activities which involve communicative tasks are more useful as some SRS requires them to function properly.

Secondly, regarding autonomy, the constant use of SRS as a tool to rehearse the correct pronunciation of vocabulary allowed the learners to become progressively autonomous as they could practice in and outside the classroom, and they only required their devices. Besides, it is clear that the use of AI has emerged as a new tool within the educational context. Language learning wise, the use of this type of technology has open new doors to innovation; the use of structural learning platforms and conversational AI tools, has allowed learners to have a more personalize and adaptative experience. Following the aforementioned, the transformative potential of SRS technology is recognized as it leads to more accurate and personalized learning experiences for students. AI-driven SRS systems can significantly enhance pronunciation improvement, especially in connected speech settings where variations in pronunciation often occur in a continuous stream of speech. These systems now leverage advanced natural language processing (NLP) and machine learning algorithms, enabling them to better understand and process diverse accents, dialects, and speech patterns. This advancement is particularly important in educational contexts, fostering a more inclusive learning environment that provides tailored pronunciation feedback to students from varied linguistic backgrounds. Additionally, AI enhances the adaptability of these systems, allowing them to continuously learn and improve based on student interactions, thereby offering increasingly precise and effective pronunciation training over time. These improvements not only boost the overall accuracy of speech recognition but also contribute to a more engaging and supportive learning experience, making AI-powered SRS a valuable tool in language education.

Thirdly, an equally significant conclusion relates to the human factor. Teachers' guidance is required when using the SRS to practice pronunciation since it needs to be properly set before being used. Besides, teachers' explanation and modeling of the sounds is necessary for students at basic

levels who have not received phonetic instruction and whose lessons do not have an emphasis on this component; thus, their instruction is meaningfully complemented by using instant feedback provided by SRS.

Lastly, this research is expected to serve as a precedent for the execution of more studies that will enrich the resources available in this relatively new area. Thus, it is important to recognize how technology can change the way pronunciation is taught in EFL classrooms. The method carried out in this study can be replicated by teachers in their classrooms so that students can increase their awareness, autonomy and pronunciation accuracy.

References

Ahmad, Y. (2018). Teaching English Pronunciation of Suprasegmental Features on Students of English Education. *SHS Web of Conferences*, (42), 1-5. https://doi.org/10.1051/shsconf/20184200048

Asratie, M., Wale, B., and Aylet, Y. (2023). Effects of using educational technology tools to enhance EFL students' speaking performance. *Education and Information Technologies*, (28), 1-21. https://link.springer.com/article/10.1007/s10639-022-11562-y

Bogdan, R., and Biklen, S. (1997). *Qualitative Research in Educatio*n. Allyn & Bacon.

Brookhart, S. M. (2008). *How to Give Effective Feedback to Your Students*. Association for Supervision & Curriculum Development. https://ebookcentral-proquest-com.ezproxy.unbosque.edu.co/lib/bibliobosque-ebooks/detail.action?docID=361035

Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford University Press.

Bryant A., and Charmaz, K. (2007). Grounded theory research: Methods and practices. In A. Bryant and K. Charmaz (Eds.), *The Sage handbook of Grounded Theory* (pp. 1-28). Sage Publications. https://doi.org/10.4135/9781848607941

Çelik, S., and Baran, E. (2022). Student response system: its impact on EFL students' vocabulary achievement. *Technology, Pedagogy and Education*, *31*(2), 141-158. https://doi.org/10.1080/1475939X.2021.1986125

Çetin, E. (2021). Digital storytelling in teacher education and its effect on the digital literacy of pre-service teachers. *Thinking Skills and Creativity*, (39), 100760. https://www.tandfonline.com/doi/full/10.1080/1475939X.2021.1986125

Council of Europe. (2022). *Common European Framework of Reference for Languages: Learning, teaching, assessment* – Companion volume. Council of Europe Publishing. www.coe.int/lang-cefr

Derwing, T., and Munro, M. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. John Benjamins Publishing Company.

Evers, K., and Chen, S. (2021). Effects of Automatic Speech Recognition Software on Pronunciation for Adults with Different Learning Styles. *Journal of Educational Computing Research*, *59*(4), 669-685. https://journals.sagepub.com/doi/abs/10.1177/0735633120972011

Ferrando, E. (2023). La motivación en la enseñanza de segundas lenguas: aproximación teórico-práctica en el ámbito del enfoque integrado AICLE. *Lengua y Sociedad, 22*(1), 117-137. https://doi.org/10.15381/lengsoc.v22i1.23650

Hazen, H. (2020). Use of oral examinations to assess student learning in the social sciences. *Journal of Geography in Higher Education*, *44*(4), 592-607. https://www.tandfonline.com/doi/abs/10.1080/03098265.2020.1773418?journalCode=cjgh20

Kelly, G. (2000). *How to Teach Pronunciation*. Longman.

Levis, J., and Suvorov, R. (2012). Automatic Speech Recognition. In C. Chapelle (Ed.), *The Concise Encyclopedia of Applied Linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0066.pub2

Macklem, G. (2015). *Boredom in the classroom: Addressing Student Motivation, Self-regulation, and Engagement in Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-13120-7

Moccozet, L. (2012, July). *Introducing Learning Performance in Personal Learning Environments.* 2012 IEEE 12th International Conference on Advanced Learning Technologies. https://doi.org/10.1109/icalt.2012.75

Plailek, T., and Essien, A. (2021). Pronunciation problems and factors affecting English pronunciation of EFL students. *Turkish Journal of Computer and Mathematics Education, 12*(12), 2026-2033. http://ezproxy.unbosque.edu.co:2048/login?url=https://www.proquest.com/scholarly-journals/pronunciation-problems-factors-affecting-english/docview/2622819321/se-2

Meilani, F., Sukiyadi, D., and Purnawarman, P. (29 December 2021). Online Individual Corrective Feedback: English Language Learners' Beliefs in Pronunciation Learning. In *Proceedings of the International Conference on Sustainable Innovation Track Humanities Education and Social Sciences (ICSIHESS 2021)* (pp. 9-14). Atlantis Press. https://www.atlantis-press.com/proceedings/icsihess-21/125967893

Syauqi, K., Munadi, S., and Triyono, M. (2020). Students' Perceptions toward Vocational Education on Online Learning during the COVID-19 Pandemic. *International Journal of Evaluation and Research in Education, 9*(4), 881-886. http://doi.org/10.11591/ijere.v9i4.20766

Stratton, S. (2021). Population Research: Convenience Sampling Strategies. *Prehospital and Disaster Medicine, 36*(4), 373-374. https://doi.org/10.1017/S1049023X 21000649

Vahdany, F., Divsar, H., and Alem, M. (2022). The Interface between Pronunciation Learning Strategies (PLS) and Pronunciation Achievement. *Iranian Journal of English for Academic Purposes*, *11*(2), 56-73. https://journalscmu.sinaweb.net/article_154305_a2bb683213798fd5cf858b86081681c7.pdf

Wang, B., Teo, T., and Yu, S. (2017). Teacher feedback to student oral presentations in EFL classrooms: a case study. *Journal of Education for Teaching*, *43*(2), 262-264. https://doi.org/10.1080/02607476.2016.1257507

Yokota, Y., Iwatsubo, T., Takeuchi, T., Hakoda, A., Nakagawa, Y., Kawabata, K., Inoue, Y., Miyamoto, H., Ikeo, K., Kojima, Y., Miyazaki, J., Abe, T., and Higuchi, K. (2022). Effects of a novel endoscopic reporting system with voice recognition on the endoscopic procedure time and report preparation time: propensity score matching analysis. *Journal of Gastroenterology*, 1-9. https://link.springer.com/article/10.1007/s00535-021-01835-7

Zuber-Skerritt, O., and Wood, L. (Eds.). (2019). Action Learning and Action Research: Genres and Approaches. *Emerald Publishing Limited*. https://doi.org/10.1108/9781787695375

Appendix

Abbreviations used

SRS	Speech Recognition Software
EFL	English as a Foreign Language
L2	Second or foreign language
CEFR	Common European Framework of Reference for Languages
T	Teacher
FN	Field Notes
J	Journal
S	Student
C	Cycle

Authors' contribution

Laura Sánchez and Andrés Morales have participated in the data collection, analysis, and interpretation. Ingrid Rodríguez has contributed to the writing and critical revision of the article. The authors approve the version published in the journal.

Acknowledgments

The authors wish to thank the Language Center of Universidad El Bosque for the support in achieving the objectives of this research.

Funding

The research was conducted without funding.

Conflict of interest

The authors declare no conflicts of interest.

Correspondence: sanchezlaura@unbosque.edu.co

Academic background of the Authors

Laura Sánchez is an associate instructor at Universidad El Bosque in Bogotá, Colombia, with more than 8 years of experience in EFL teaching. She holds a bachelor's degree in Bilingual Education with an emphasis on English teaching from Universidad El Bosque, and a master's degree in Education from the same institution. She has participated in the creation of online EFL learning material at the Language Center of Universidad El Bosque. Her research interests include dynamic assessment, teaching and learning methods in English as a foreign language, and ICT for educational purposes.

Ingrid Rodríguez holds a Bachelor's degree in English and Spanish from Universidad Pedagógica Nacional, a Master's degree in Applied Linguistics for English Language Teaching from Universidad Distrital Francisco José de Caldas, and is currently pursuing a PhD in Social Studies at Universidad Jorge Tadeo Lozano. She has 22 years of certified experience as an English and content teacher in bilingual schools and universities, both in language centers and in undergraduate programs. She currently has more than 11 years of experience as a full-time teacher and Core Faculty in the Bilingualism Bachelor's program at the School of Education at Universidad El Bosque. She also led the Equilateral Knowledge research group in the Bilingualism program from 2018 to 2022 and is currently an active researcher in the Education and Research group at UNBOSQUE.

Andrés Morales is an experienced educator with over 10 years of teaching English in diverse settings, including universities and language academies, and has also worked as an Academic Coach at Santillana in Bogotá, Colombia, where he specialized in implementing educational systems and training teachers. He holds an undergraduate degree in Modern Languages from Universidad Javeriana, a postgraduate degree in University Teaching from Universidad El Bosque, and is currently pursuing a Master of Education in Educational Technology and Learning Design at Simon Fraser University, focusing on integrating technology into teaching and learning.