



**UNIVERSIDAD BOLIVARIANA DEL ECUADOR**  
**MAESTRÍA EN PEDAGOGÍA DEL INGLÉS COMO LENGUA**  
**EXTRANJERA.**

**TRABAJO DE TITULACIÓN**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN**  
**PEDAGOGÍA DEL INGLÉS COMO LENGUA EXTRANJERA**

**TEMA:**

**Development of a Text-To-Speech Interaction Platform to Improve**  
**Pronunciation**

**Autores:**

**Daniel Fabricio Contreras Moscol**

**Vicente Javier Coello Vásquez**

**Tutora:**

**BA. Carmen Andrea Letamendi Lazo, Msc.**

**ECUADOR**

**2024**

**DEDICATION**

Dedico este trabajo a mis familiares, colegas, profesores y alumnos que cada día me inspiran a seguir adelante.

Daniel Contreras

**DEDICATION**

Dedico esta tesis a mi familia, a mi esposa y todas aquellas personas que han formado parte de este arduo camino.

Javier Coello

## **Development of a Text-To-Speech Interaction Platform to Improve Pronunciation**

### **Abstract**

This study aimed to explore the effectiveness of Text-to-Speech (TTS) technology in enhancing pronunciation and overall language proficiency among A2-level English as a Foreign Language (EFL) learners at the University). It was conducted over a 30-day intervention period, the study involved 16 randomly selected students who practiced daily using a TTS reader. A Pre-test and Post-test design measured the impact, revealing significant improvements in pronunciation, as evidenced by the increase in mean scores from the Pre-test to the Post-test and the statistical significance confirmed by a paired sample t-test ( $t(15) = -2.304, p = 0.036$ ). Survey data showed positive student perceptions, with high satisfaction regarding the user interface, speech quality, and engagement levels. While the study considers the potential of TTS technology as a beneficial tool in EFL education, it acknowledges limitations related to sample size and intervention duration. Recommendations include integrating TTS into curricula, providing educator training, and conducting further research to examine long-term effects and broader language skill impacts.

**Keywords:** Text-to-Speech (TTS), pronunciation improvement, language proficiency, language acquisition.

## Resumen

Este estudio tuvo como objetivo explorar la efectividad de la tecnología de Texto a Voz (TTS) en la mejora de la pronunciación y la competencia lingüística general entre los estudiantes de inglés como lengua extranjera (EFL) de nivel A2 en la Universidad de Babahoyo. Realizado durante un período de intervención de 30 días, el estudio involucró a 16 estudiantes seleccionados al azar que practicaron diariamente utilizando un lector TTS. Se utilizó un diseño de pre-test/post-test para medir el impacto, revelando mejoras significativas en la pronunciación, evidenciadas por el aumento en las puntuaciones medias del pre-test al post-test y la significancia estadística confirmada por una prueba t de muestras emparejadas ( $t(15) = -2.304$ ,  $p = 0.036$ ). Los datos de la encuesta mostraron percepciones positivas de los estudiantes, con una alta satisfacción respecto a la interfaz de usuario, la calidad del habla y los niveles de compromiso. Si bien el estudio destaca el potencial de la tecnología TTS como una herramienta beneficiosa en la educación EFL, reconoce las limitaciones relacionadas con el tamaño de la muestra y la duración de la intervención. Las recomendaciones incluyen la integración de TTS en los planes de estudio, la capacitación de los educadores y la realización de investigaciones adicionales para examinar los efectos a largo plazo y el impacto en habilidades lingüísticas más amplias.

**Palabras claves:** Texto a Voz (TTS), mejora de la pronunciación, competencia lingüística, adquisición de lenguaje.

## Table of Contents

Introduction.....	1
Chapter 1: Literature Review .....	9
Chapter 2: Research Methodology.....	27
2.1. Statement and justification of the research approach.....	27
2.2. Description of the research methods and their purposes in the context of the proposal .....	28
2.3. Conceptualization and operationalization of the main categories .....	29
2.4. Description of the instruments based on the research approach. ....	31
2.5. Description of the validation of Instruments.....	33
2.6. Population, Sample, And Sampling .....	34
2.7. Data Analysis and Ethical Considerations .....	35
2.8. Initial Diagnose .....	36
Chapter 3: Research Proposal .....	41
3.1. Validation of the proposal.....	41
3.2. Theoretical foundations .....	43
3.3. Characteristics of the proposal.....	45
3.4. Structure, and dynamics of the components of the proposal .....	47
3.4.1. Pre-test before intervention.....	47

3.4.2.	30-Day Intervention Period.....	48
3.4.3.	Post-Test activity after intervention.....	49
3.4.4.	Analysis and Results .....	49
3.5.	Description of the methodological and/or technological requirements	50
3.6.	Results and Discussion .....	52
3.6.1.	Intervention results.....	52
3.6.1.1.	Post-test results .....	52
3.6.1.2.	Paired T-Test.....	55
3.6.2.	Descriptive statistics based on students' perceptions. ....	58
3.6.3.	Discussion .....	67
3.7.	Limitations in the study .....	69
	Conclusions.....	71
	Recommendations.....	73
	BIBLIOGRAPHICAL REFERENCES .....	74
	APPENDICES .....	81

## Index of Tables and Figures

<b>Table 1</b>	Scores of 16 students of Level 3	37
<b>Table 2</b>	The main statistic of Scores of Level 3	39
<b>Table 3</b>	Post-test scores of 16 students	52
<b>Table 4</b>	Post-test main statistic of Scores of Level 3 students	54
<b>Table 5</b>	Paired T-test results	57
<b>Table 6</b>	Demographic statistics of survey	58
<b>Table 7</b>	Usage of the TTS reader rated by students	61
<b>Table 8</b>	Students' perception of the intervention	63
<b>Figure 1</b>	Students obtained scores in the Pre-test activity	38
<b>Figure 2</b>	Pre- and Post-tests variations	53
<b>Figure 3</b>	The perspective of students about the platform over time	60
<b>Figure 4</b>	Aspects of improvement of using the platform	65

## **Introduction**

In recent years, language education has been changing due to the integration of advanced technology into pedagogical practices. Among those challenges faced by language learners, one of the most important is pronunciation accuracy. However, for non-native speakers, achieving a high level of accuracy remains difficult, the use of text-to-speech (TTS) technology has presented a solution to address these challenges. This TTS technology converts written text into spoken language, bridging the gap between written content and auditory input.

The discourse and empirical findings on this topic have included the significant impact of TTS technology on language acquisition. Al-Jarf (2022) study underscores the potential of TTS software to elevate the decoding skills and pronunciation accuracy of EFL freshman students. Bonces (2019) has explored the implementation of TTS technology as a means to increase fluency in a second language (L2). Meanwhile, (Y. Chen et al., 2018) have ventured into the development of a corpus-based online platform, leveraging TTS to facilitate pronunciation learning of Mandarin.

On the other hand, Gottardi et al. (2022) have offered good insights into the affordances of automatic speech recognition and TTS technologies in enhancing L2 pronunciation. These referenced studies highlight the potential of TTS technology in enhancing language learners' phonetic decoding, fluency, and pronunciation accuracy. Thus, this proposed platform seeks to synthesize and expand upon these insights to create an innovative learning tool that not only refines reading accuracy

and pronunciation but also improves learners' self-assurance in their language learning journey.

While technology has shown promise in aiding language learning, the integration of resources to enhance these specific skills remains an ongoing concern. Despite its potential to bridge the gap between written and spoken language, there is a scarcity of platforms that target the improvement of pronunciation skills. The existing literature highlights the potential of TTS technology to improve aspects of language learning, but a comprehensive platform customized explicitly to L2 learners' needs is lacking.

Additionally, integrating TTS technology with pedagogical strategies and the assessment of its impact on learners' pronunciation warrant further exploration. Therefore, a critical inquiry is needed to develop an innovative text-to-speech interaction platform that addresses these shortcomings and provides learners with an engaging tool for improving pronunciation skills. Then, the question that guides this research is: How can a text-to-speech interaction designed platform effectively enhance L2 learners' pronunciation?

This research aims to address the gap in language education by using a text-to-speech interaction platform designed to enhance pronunciation among second language learners. Despite the recognition of the importance of these skills in language proficiency, learners often face obstacles in attaining the desired level of accuracy and fluency.

Then, the incorporation of TTS technology in a user-friendly and interactive platform holds promise for integrating learners' diverse needs and learning preferences. Such a platform has the potential to provide learners with an immersive experience that supports pronunciation through repeated exposure to spoken language. By developing and working in the use of the platform, this research aims to contribute to the enhancement of pronunciation in a single integrated solution.

The selected thesis topic aligns with the research directions at the Universidad Bolivariana del Ecuador, specifically in the domain of educational methodologies and technology innovation. This proposition conforms to the university's dedication to driving instructional techniques and improving the learning atmosphere through state-of-the-art approaches. Furthermore, the pedagogical implications of this platform extend beyond its immediate impact on learners' language skills. The proposed research aligns with contemporary educational trends that prioritize personalized and technology-enhanced learning experiences.

On the other hand, this research is led by the pressing need to address the challenges faced by L2 learners in achieving pronunciation accuracy. The main evidenced problem in the students relates to speaking and pronunciation, they are afraid to talk and according to observations, they are afraid to make mistakes, mostly because their pronunciation is not good enough. Thus, the main and supporting aims are detailed:

***Research aim***

Analyze the use of an innovative text-to-speech interaction platform that effectively enhances second language (L2) learners' pronunciation accuracy, thereby contributing to their overall language proficiency and confidence.

***Supporting aims***

- Analyze the pedagogical implications and potential applications of the developed platform within language learning curricula.
- Explore the advantage of a user-friendly text-to-speech interaction platform that integrates advanced natural language processing and speech synthesis technologies through an intervention with A2-Level students.
- Evaluate the students' perception of the TTS's effectiveness in improving pronunciation accuracy among L2 learners.

Then, the independent variable of this study is the Text-to-Speech Interaction Platform which refers to a software tool developed to convert written text into spoken language. This platform is designed to enhance the pronunciation accuracy of second language (L2) learners, by engaging learners with textual content through auditory means, facilitating the development of pronunciation abilities.

The operationalization of this variable involves creating a user-friendly online platform that integrates advanced natural language processing and speech synthesis technologies. This platform contains user interface design, speech

synthesis capabilities, synchronized text highlighting during speech, and interactive exercises.

On the other hand, the dependent variable is Language Proficiency Improvement which signifies the overall enhancement of various language skills, including pronunciation. This concept represents the advancement of learners' linguistic competencies, leading to improved comprehension, effective communication, and heightened confidence in the target language. The operationalization of this variable involves measuring the quantifiable enhancement of pronunciation accuracy among L2 learners. The term "pronunciation improvement" refers to their skill in articulating sounds, stress patterns, and intonation.

The research approach will primarily involve a combination of descriptive statistics and a survey-based analysis to comprehensively explore the platform's impact and learners' perceptions. The research will employ a descriptive statistical approach to analyze the measurable outcomes resulting from the platform's utilization. Pre- and post-assessment tests will be administered to a selected group of L2 learners who engage with the Text-to-Speech Interaction Platform. Then, descriptive statistics such as means, standard deviations, and percentage changes will be calculated to provide a clear overview of the extent of improvement achieved through the platform.

To gain a comprehensive understanding of student's perceptions and experiences, a survey will be administered to participants using the Text-to-Speech

Interaction Platform. The survey will include questions addressing aspects such as user satisfaction, ease of use, perceived effectiveness, motivational factors, and challenges encountered. Likert scale and open-ended questions will be employed to capture both quantitative and qualitative feedback.

Descriptive statistical methods will be utilized to summarize and interpret the quantitative survey responses. Moreover, thematic analysis will be conducted on qualitative responses to uncover recurring themes and sentiments expressed by the participants. Initially, it will evaluate the impact of the Text-to-Speech Interaction Platform on language proficiency, and pre- and post-assessment tests will be designed and administered.

These tests will include reading exercises to assess learners' ability to decode and comprehend written text and pronunciation assessment exercises to evaluate their articulation of sounds, stress patterns, and intonation. The pre-assessment will establish a baseline, while the post-assessment will measure improvements after platform engagement. The scores from these assessments will provide quantitative data for descriptive statistical analysis, offering insights into learners' proficiency enhancement.

A survey questionnaire will be developed to gather students' perceptions and experiences with the Text-to-Speech Interaction Platform, consisting of both closed-ended and open-ended questions. The closed-ended questions will utilize Likert scale responses to quantify participants' satisfaction levels, perceived effectiveness, ease of use, and motivations for platform usage. Additionally, open-

ended questions will allow participants to provide qualitative feedback, enabling them to express their opinions, challenges, and suggestions.

This research will be conducted at the University of Babahoyo, a Higher Education Institution, located in Babahoyo, Los Ríos. This university operates from morning to evening, the whole week, where around 9000 students are educated. In the language center called CENID (from Spanish: Centro de Idiomas), there are 25 teachers who virtually present their classes daily. At this current semester, June-September, 6000 students were approximately enrolled, from A1 to B1 CEFR levels. Students need to complete 5 English courses where they reach the English proficiency necessary for completing their studies and graduating.

Students are from 18 to 30 years old and come from different backgrounds and places. Students are from different parts of Los Rios province, and most of them live in rural areas. The group is highly diverse, culturally and linguistically, for that reason, CENID authorities try hard to get a supportive and collaborative learning environment, encouraging active engagement in meaningful tasks and authentic communication.

The research outcomes will provide educators, curriculum developers, and instructional designers with empirical insights into the platform's efficacy, guiding them in making informed decisions about incorporating technology-enhanced learning tools into language education. For this reason, the proposed research aligns with a growing societal need for effective language education tools that guide these challenges.

The social need for technology-enhanced language learning solutions that offer personalized and interactive experiences further underscores the relevance and timeliness of this research. Moreover, the study's outcomes have the potential to influence pedagogical practices, enhance language education curricula, and contribute to learners' overall language competence, thus fulfilling its role in addressing the contemporary challenges of language education and meeting the demands of a rapidly changing linguistic area.

This thesis is divided into three chapters: i. Literature review where through a rigorous revision of academic works will validate the initial ideas and propose a theoretical framework to be a base; ii. The methodology chapter will include the overall process to gather and analyze data pertinently and iii. A research proposal was developed to analyze the best path to help students improve their pronunciation skills.

## **Chapter 1: Literature Review**

Through this chapter it is discussed the use of TTS technology in language learning and its ability to assist learners in achieving accurate pronunciation when learners have difficulties with unfamiliar phonemes, stress patterns, and intonation. This chapter then describes previous important research developed and discusses its importance and implications for Language Learning.

### **1.1. Research on Technology-Enhanced Language Learning**

TTS technology offers learners opportunities for repeated exposure to authentic spoken language. By engaging with TTS-generated audio, learners can fine-tune their listening skills, adapting to different accents, speech rates, and linguistic variations.

#### **1.1.1. Global Research Trends about TTS and Language Learning**

Technology-enhanced language learning has gained attention on the global stage, driven by advancements in Text-to-Speech (TTS) technology, across various countries, scholars have explored the potential of TTS into language education. Previous research referred to the evolution of TTS technology, as evidenced by studies by (Y. Chen et al., 2018) and Nachmani and Wolf (2019) where deep learning and neural networks have enhanced speech synthesis, resulting in more natural-sounding voices.

Additionally, researchers have conducted investigations into the use of TTS in education, highlighting its role as a tool to facilitate language learning and pronunciation improvement (Afonso et al., 2022; Yang et al., 2022).

Moreover, global research has consistently emphasized the positive impact of TTS on pronunciation skills, Wood et al. (2018) have reviewed Meta-analyses and have underscored the effectiveness of technology-enhanced approaches, including TTS and read-aloud tools, in improving reading comprehension among students with reading disabilities. These findings have contributed to the growing recognition of technology's potential in addressing pronunciation challenges and increasing language proficiency (Wood et al., 2018).

Furthermore, the integration of technology into language curricula has been a prominent theme in global research. Studies have explored innovative pedagogical strategies, such as authentic digital storytelling, which leverage technology to enhance speaking skills in English as a Foreign Language (EFL) classes (Arroba & Acosta, 2021).

### **1.1.2. Local Research on Technology-Enhanced Language Learning**

In Ecuador, the educational landscape has also witnessed a growing interest in technology-enhanced language learning. Research conducted by (Acosta & Cajas, 2018) has investigated the use of teaching resources in EFL classes in Ecuadorian universities. This local perspective highlights the need to understand how technology is employed within the specific context of Ecuador's higher education institutions (Acosta & Cajas, 2018).

Furthermore, some authors have explored strategies to improve pronunciation and speaking skills among EFL learners in Ecuador, indicating a local awareness of the challenges faced by Ecuadorian students in language

learning, particularly in the domains of pronunciation and speaking (Benitez-Correa et al., 2020; Encalada & Sarmiento, 2019). This research underscores the importance of addressing these challenges through innovative approaches that often involve technology integration. Overall, both global and local research efforts in Ecuador emphasize in the potential of TTS and related technologies to bridge the gap between written and spoken language.

## **1.2. Text-to-Speech (TTS) Technology and Language Learning**

Text-to-speech (TTS) technology appeared as a tool in language learning, offering solutions to long-standing challenges faced by learners. This technology, which converts written text into spoken language, holds the potential to improve aspects of language acquisition, with a particular focus on pronunciation and listening skills.

### **1.2.1. Evolution of TTS technology**

The evolution of Text-to-Speech (TTS) technology has been important, marked by significant advancements in both capabilities and applications, starting from its rudimentary origins, TTS has come a long way, revolutionizing diverse sectors, including language learning, accessibility, and artificial intelligence (Kumar et al., 2023).

In the mid-20th century, TTS technology took its first steps as researchers began exploring methods to synthesize human-like speech from text. Early TTS systems relied on rule-based algorithms and simple concatenation of pre-recorded speech units to generate speech (Lu, 2019). While these systems were limited in

linguistic accuracy and naturalness, they served as the foundation for future innovations.

In the 1970s witnessed a significant breakthrough with the introduction of formant synthesis, where frequencies of the vocal tract were manipulated to produce more natural-sounding speech. This approach allowed for greater control over the speech generation process and improved intelligibility (Nachmani & Wolf, 2019).

As the late 20th century presented advancements in concatenative synthesis, took center stage. In this method, entire units of recorded speech, known as phonemes or diphones, were combined to form coherent speech. This approach led to substantial improvements in the naturalness and expressiveness of synthesized speech (Valentini-Botinhao & Yamagishi, 2018).

Moreover, Delić et al. (2019) discussed a statistical parametric synthesis and Hidden Markov Models (HMMs) proposed techniques such as neural networks that revolutionized the field. These models allowed for the detailed modeling of speech characteristics by learning from extensive datasets of human speech, resulting in more natural and human-like voices (Delić et al., 2019).

In recent years, neural TTS has been used in TTS technology, where deep learning architectures, which include WaveNet and Tacotron, have pushed the boundaries of what TTS can achieve. Neural TTS systems can generate speech with remarkable naturalness, including intonation and emotion, making it challenging

for listeners to distinguish between synthesized and human speech (Nachmani & Wolf, 2019).

Furthermore, Danielescu et al. (2023) explored customization and voice cloning, which have become prominent features of modern TTS technology. These capabilities allow for the creation of unique, personalized voices and have practical applications in industries like entertainment, virtual assistants, and accessibility tools.

### **1.2.2. Types of TTS systems and their applications in education**

Text-to-Speech (TTS) systems, have been potentiated into various types of specific applications (Adamopoulou & Moussiades, 2020). In the field of education, these diverse TTS systems have opened up exciting possibilities for enhancing the teaching and learning experience (Sečujski et al., 2019), such as:

- **Rule-Based TTS Systems:** Rule-based TTS, pioneered during the early stages of TTS technology, focusing on predefined linguistic rules and phonetic models, as noted by Lu (2019). These systems are well-suited for educational applications, such as assisting students with disabilities, as they offer control over speech synthesis parameters. In education, rule-based TTS systems are often used to provide audio support for digital learning materials, making content accessible to students with varying needs.
- **Concatenative TTS Systems:** Salas-Pilco and Yang (2022) discussed assembling speech from a database of recorded phonemes or diphones. This approach excels in capturing natural speech prosody. In educational

contexts, concatenative TTS is employed to create engaging audiobooks, e-learning modules, and language-learning resources.

- **Statistical Parametric TTS Systems:** Delić et al. (2019) highlighted the leverages of statistical models, such as Hidden Markov Models (HMMs) and neural networks, to synthesize speech. These systems are known for their adaptability and can produce a wide range of voices. In education, statistical parametric TTS is utilized to generate personalized instructional materials, educators can customize content to students' learning styles and preferences, producing engagement and comprehension.
- **Neural TTS Systems:** Nachmani and Wolf (2019) discussed the representation of cutting edge of TTS technology. Based on deep learning architectures like WaveNet and Tacotron, these systems produce natural and expressive speech. In the educational sphere, neural TTS is a game-changer. It enables the creation of virtual tutors, interactive language learning platforms, and accessible content for students with disabilities. The lifelike voices generated by neural TTS systems create immersive learning experiences.
- **Customizable TTS Systems:** Danielescu et al. (2023) explained that it could empower users to create unique, personalized voices. This type of TTS has found applications in education by allowing educators to develop custom narrations for instructional materials. Additionally, it facilitates the creation of diverse voices for language learning platforms, accommodating students' cultural and linguistic preferences.

As a result, TTS systems have diversified applications in education to offer educators versatile tools for enhancing the learning process. From making content accessible to students with disabilities to creating great language learning experiences.

### **1.2.3. Advantages and limitations of TTS technology**

One of the most prominent advantages of TTS technology is its role in making digital content more accessible. TTS enables individuals with visual impairments, learning disabilities, or language barriers to access written information (Gottardi et al., 2022). It promotes inclusivity in education by providing equal opportunities for all students to engage with educational materials (Wood et al., 2018).

Thus, TTS technology can significantly aid language learners. By converting text into spoken language, TTS helps learners improve pronunciation, listening comprehension, and fluency. It provides learners with an authentic model for correct pronunciation, making it a valuable tool for language educators (Chen et al., 2023).

Moreover, TTS systems allow for voice personalization, which can create a more relatable experience and customizable voices can support to individual preferences and regional accents, integrating a sense of familiarity and cultural relevance in educational materials (Yu et al., 2023). On the other hand, TTS technology can significantly boost efficiency, it enables the rapid conversion of written content into speech, saving time for both educators and students. This

efficiency is particularly useful in scenarios where large volumes of text need to be processed or when content needs to be adapted for different learning styles (Reddy et al., 2023).

While TTS technology has made progress, there are still limitations in achieving completely natural and human-like speech. Synthetic voices may lack the nuances, intonation, and emotional expression found in human speech, potentially impacting the overall user experience (Reddy et al., 2023). However, TTS systems may struggle with certain accents, dialects, or languages, leading to pronunciation inaccuracies, this limitation can hinder language learners looking for precise pronunciation models (Dongmei, 2021). TTS technology may also struggle with understanding context, which can result in mispronunciations or improper emphasis on words, especially in complex or ambiguous sentences, affecting the comprehension of content (Frissell, 2019).

Developing high-quality TTS systems requires substantial computational resources, including large datasets and complex machine-learning models. Implementing and maintaining such systems can be resource-intensive for educational institutions (Sairanen, 2023). Although TTS technology is dependent on devices and infrastructure, which may not be universally available, especially in underserved regions there is a lack of access to suitable devices or the internet can limit its usefulness for some learners (Danielescu et al., 2023).

### **1.3. Language Pronunciation in English as a Foreign Language Learning**

Pronunciation has an important role in language learning, particularly in the context of English as a foreign language (EFL). Accurate pronunciation is not only an essential component of effective communication but also a marker of language proficiency (Saito & Plonsky, 2019). In the EFL classroom, achieving correct pronunciation is often a challenging task for learners, and this challenge has been the subject of extensive research worldwide.

Research has highlighted different aspects influencing pronunciation in EFL learning, including individual learner differences, teaching methodologies, and technological interventions. For instance, (Fadillah, 2020) conducted a literature study on the pronunciation difficulties faced by EFL learners in Indonesia, this study underscored the significance of addressing these difficulties early in the language learning process to prevent fossilization of errors (Jahara & Abdelrady, 2021).

In the Ecuadorian context, where English is taught as a foreign language, research has examined the specific challenges faced by learners. Encalada and Sarmiento (2019) explored learners' perceptions of self-recording videos as a means to develop speaking skills. This local study demonstrated the importance of innovative teaching strategies to improve pronunciation and overall speaking proficiency.

Moreover, Benitez-Correa et al. (2020) conducted a case study focusing on improving the past tense pronunciation of regular verbs among EFL undergraduate students in Ecuador. The research utilized Audacity software, highlighting the potential of technology-enhanced tools in addressing pronunciation challenges. Then, it is important to recognize that pronunciation improvement is not only the responsibility of the learners.

Teachers and teaching methodologies shape learners' pronunciation skills. Huensch (2019) emphasized the significance of teacher training and classroom practices in supporting effective pronunciation instruction. Authentic and engaging teaching strategies, such as authentic digital storytelling, can enhance learners' speaking skills and motivate them to overcome pronunciation challenges (Arroba & Acosta, 2021).

Furthermore, the integration of technology, including text-to-speech (TTS) technology, has opened new avenues for pronunciation improvement. The study by Chen et al. (2023) examined college EFL learners' perceptions of using Google Assistant for foreign language learning, highlighting the potential of intelligent technology in language pronunciation instruction. TTS technology, as a component of this, can serve as a tool for providing learners with accurate pronunciation models (Mirza, 2015).

### **1.3.1. Pronunciation and pedagogical approaches**

Some pedagogical approaches have been developed to address pronunciation in language teaching, expressing the diverse needs and goals of

learners (Saito & Plonsky, 2019). First of all, traditional pedagogical approaches to pronunciation instruction often involve the use of phonetic symbols and auditory discrimination exercises. These approaches emphasize explicit instruction and repetition, focusing on individual phonemes and their articulation (Pennington, 2021). Such methods help learners develop a foundational understanding of the phonetic components of a language, aiding in the correction of specific pronunciation errors.

In recent years, communicative and integrative approaches to language teaching have gained prominence, which prioritize the integration of pronunciation into meaningful communicative contexts. Learners engage in authentic, real-world speaking tasks, emphasizing fluency and effective communication rather than just accuracy (Saito & Plonsky, 2019). This shift towards a more holistic approach recognizes that pronunciation is a dynamic aspect of language learning best developed through active use in context (Huensch, 2019).

Moreover, task-based and experiential learning approaches encourage learners to engage in pronunciation practice through experiential activities and real-world tasks. Learners may participate in role-plays, debates, or interactive dialogues, where clear pronunciation is essential for successful communication (Arroba and Acosta, 2021). As a result, advancements in technology have opened up new possibilities for pronunciation instruction.

For instance, computer-assisted pronunciation training (CAPT) programs and applications often utilize speech recognition technology and visual feedback to

provide learners with real-time pronunciation assessment and correction (Stroparo and Floriani, 2023). Additionally, text-to-speech (TTS) technology has been integrated into language teaching platforms, offering learners access to accurate pronunciation models (Chen et al., 2023b).

Many language educators emphasize the importance of integrating pronunciation instruction into the broader language curriculum rather than treating it as an isolated component. By including pronunciation practice in lessons, teachers can help learners see the connection between pronunciation and overall language skills (Huensch, 2019).

### **1.3.2. Integration of Technology in Language Education**

Modern technology offers different tools and platforms that enhance language acquisition, making the learning process more engaging, interactive, and accessible. Several key aspects characterize the integration of technology in language education:

- Digital learning platforms and learning management systems (LMS) have become central to language education. These platforms provide a virtual space for teachers and learners to interact, access course materials, submit assignments, and engage in collaborative activities (Encalada and Sarmiento, 2019). They also offer features like automated grading and analytics to monitor student progress.
- The proliferation of online language courses, often delivered through Massive Open Online Courses (MOOCs) or specialized language

learning apps, has increased language education. Learners can access high-quality language instruction from anywhere, at their own pace (Adamopoulou and Moussiades, 2020).

- Language learning apps have gained popularity due to their convenience and user-friendly interfaces. Apps like Duolingo, Babbel, and Rosetta Stone offer interactive lessons, gamification elements, and personalized learning experiences, to diverse language learners (Adamopoulou and Moussiades, 2020).
- VR and AR technologies provide immersive language learning experiences. Learners can practice language skills in simulated real-world scenarios, enhancing their ability to communicate effectively in practical situations (Alemi and Khatoony, 2020). One important platform that delivers these types of activities is Nearpod, with some free access templates in empower students to practice with authentic materials.
- TTS technology has been integrated into language education to improve pronunciation and listening skills. TTS platforms convert text into spoken language, allowing learners to hear correct pronunciation models and practice listening comprehension (Yang et al., 2022).
- Speech recognition software provides real-time feedback on pronunciation and speaking skills. Learners can receive immediate guidance and corrections, enabling them to refine their pronunciation (Yang et al., 2022).

- Gamification techniques, such as leaderboards, badges, and rewards, are used to motivate language learners. Language learning games make the process enjoyable and help reinforce vocabulary and grammar (Kumar et al., 2023).
- Social media platforms facilitate language exchange between learners and native speakers. Learners can practice speaking and writing skills through online conversations and collaborations (Encalada & Sarmiento, 2019).
- AI-powered language assistants, like chatbots and virtual language tutors, offer personalized language learning experiences. They can provide instant answers to language-related questions and engage learners in interactive dialogues (Yang et al., 2022).
- Online language communities and forums allow learners to connect with others who share their language learning goals. Learners can seek advice, share resources, and practice their language skills in a supportive online environment (Encalada & Sarmiento, 2019).

The integration of technology in language education also presents challenges, which include issues related to access, equity, data privacy, and the need for effective teacher training (Kumar et al., 2023). Balancing technology with effective pedagogy is essential to ensure that learners receive a well-rounded language education.

### **1.3.3. Integration of Technology in EFL Education**

In EFL contexts, technology integration takes various forms, each contributing to a more dynamic and effective learning environment. Digital learning platforms, including learning management systems and online courses, have become instrumental in delivering EFL instruction (Encalada & Sarmiento, 2019). These platforms offer a central hub for course materials, assignments, and communication, streamlining the learning process and enabling asynchronous learning.

Language learning apps and software have revolutionized EFL instruction by providing learners with interactive content accessible on multiple devices (Adamopoulou & Moussiades, 2020). Apps like Duolingo or Nearpod offer personalized lessons, language exercises, and gamified elements that motivate learners to engage consistently. Additionally, Virtual Reality (VR) and Augmented Reality (AR) have opened up immersive language learning experiences. EFL learners can step into virtual environments where they interact with native speakers, practice real-life scenarios, and improve their communicative skills (Alemi & Khatoony, 2020).

Text-to-Speech (TTS) technology supports EFL pronunciation enhancement (Yang et al., 2022). By converting text into spoken language, TTS platforms help learners hear and imitate correct pronunciation models, aiding in accent reduction and listening comprehension. Then, speech recognition software provides real-time feedback on learners' pronunciation and speaking skills, offering

guidance for improvement (Yang et al., 2022). These tools help EFL learners to improve their spoken language skills effectively.

Thus, gamification elements, such as leaderboards and rewards, have been employed to make EFL learning engaging and enjoyable (Kumar et al., 2023). Language learning games reinforce vocabulary, grammar, and speaking skills while motivating learners to achieve milestones. Artificial Intelligence (AI)-powered language assistants, such as chatbots and virtual tutors, bring personalized learning experiences to EFL learners. These AI systems engage learners in dialogues, answer language-related queries, and provide instant assistance (Yang et al., 2022).

#### **1.3.4. Teachers' experiences and perception**

EFL educators are on the frontlines, implementing technology-driven pedagogical approaches and witnessing firsthand the impact on their students (Kumar et al., 2023). Many EFL teachers acknowledge the benefits of technology in the classroom. They find that digital resources, such as interactive apps and online language platforms, provide valuable supplementary materials that can integrate to different learning styles and abilities (Lameras and Arnab, 2021; Lindner et al., 2019; Wu et al., 2023).

Moreover, EFL teachers have observed that gamification elements, language apps, and virtual reality simulations captivate students' interest and motivate them to practice more consistently (Kumar et al., 2023). Language learning becomes less of a chore and more of an enjoyable experience. However, teachers also face challenges in the integration of technology. Many educators

grapple with the need for digital literacy and training to effectively use technology in their classrooms (Chen et al., 2023). They must adapt to new tools and methodologies, which can be daunting, especially for those with limited prior experience.

Another concern among EFL teachers is the potential for technology to replace the human element of teaching. While digital assistants and chatbots can provide immediate responses to language queries, some educators worry that these tools may reduce the personal connection between teachers and students (Chen et al., 2023). Looking for the right balance between technology and human interaction is a challenge.

Teachers' perceptions also extend to the assessment of technology-driven language learning. They observe that technology allows for a more comprehensive and immediate assessment of students' progress, as speech recognition software and language apps can provide instant feedback on pronunciation and language proficiency (Yang et al., 2022).

### **1.3.5. TTS and Pronunciation Improvement**

EFL learners often have some challenges in accurately articulating sounds, stress patterns, and intonation, which can hinder their overall language proficiency (Alemi & Khatoony, 2020). Then, TTS technology bridges the gap between written and spoken language, offering an alternative for learners to develop their pronunciation skills (Pourhosein Gilakjani et al., 2020). One key advantage of TTS

technology in pronunciation improvement lies in its ability to provide learners with repeated exposure to spoken language.

Learners can listen to authentic pronunciation models as the TTS system converts written text into clear and natural-sounding speech (Gilakjani, 2017). Moreover, learners can compare their own pronunciation to the TTS model, identifying areas that need improvement (Kumar et al., 2023). This real-time feedback allows for targeted practice and correction, accelerating the learning process.

Additionally, TTS technology can support learners' diverse needs, it can be integrated into various language learning platforms, offering learners flexibility in how they engage with pronunciation practice (Adamopoulou & Moussiades, 2020). Learners can choose when and how to utilize TTS tools, aligning with their individual learning styles. However, it's essential to acknowledge the limitations of TTS in pronunciation improvement. While TTS technology can provide valuable support, it should not replace the guidance of skilled language instructors (Saito & Plonsky, 2019).

## **Chapter 2: Research Methodology**

This chapter outlines the key methodologies used in the development of the proposed research. It provides an in-depth explanation of essential components including the study design, the operationalization of variables, the instruments employed, the characteristics of the sample studied, and the process of the initial diagnosis.

### **2.1. Statement and justification of the research approach**

The selected research approach was an explanatory sequential approach within a mixed-methods design, chosen to offer a multifaceted and in-depth exploration of the phenomenon under investigation (Rodríguez-Fuentes & Calle-Díaz, 2023). In the initial quantitative phase, the study focuses on analyzing previously collected reading fluency measures, specifically speech rate (words correct per minute, wcpm), using the digitally supported reading intervention program TTS Reader. This approach allows for an examination of objective, numerical data, providing insights into the participants' fluency development over the 4-week intervention period (Dongmei, 2021).

The subsequent qualitative phase, informed by the quantitative findings, looked for the reflections of the participants by incorporating participants' self-perceptions and experiences, the study aimed to contextualize and enrich the quantitative results. This two-phase design enables the research to move beyond statistical analysis through a more comprehensive understanding of the impact of the intervention (Bonces, 2019). The qualitative data, in the form of semi-structured

reflections, explain the patterns and provide insights into factors influencing fluency development that might not be captured by quantitative metrics alone.

The explanatory sequential approach is particularly valuable in this context as it allows the researchers to go beyond the 'what' (quantitative findings) to explore the 'why' and 'how' through qualitative exploration. This iterative process ensures a more holistic interpretation of the research question, contributing to the overall validity and reliability of the study. The combination of quantitative and qualitative methods in this approach strengthens the study's rigor, providing a more robust foundation for creating meaningful conclusions about the effectiveness of technology-enhanced language learning interventions on language fluency development in an EFL context (Dongmei, 2021).

## **2.2. Description of the research methods and their purposes in the context of the proposal**

The research employs a mixed-methods design, integrating both quantitative and qualitative methods to comprehensively investigate the impact of technology-enhanced language learning interventions on language fluency development in an English as a Foreign Language (EFL) context. In the initial quantitative phase, the study relies on speech rate which involves examining the fluency development of lower intermediate EFL students over 4 weeks (Rodríguez & Calle, 2023). The analysis includes a detailed exploration of participants' speech rates at different stages of the intervention, providing quantitative measures of fluency development.

The subsequent qualitative phase involves collecting and analyzing reflections from the participants who engaged in the reading-aloud protocols. Through semi-structured reflections, the study seeks to capture the participants' experiences, perceptions, and insights into their fluency development (Bonces, 2019). This qualitative component adds depth and context to the quantitative findings, offering a more holistic understanding of the multifaceted aspects influencing language fluency enhancement. The integration of these two research methods allows for a comprehensive exploration of the research question, providing a perspective on the effectiveness of technology-enhanced language learning interventions. The mixed-methods approach ensures a more robust and well-rounded investigation, as it combines the strengths of both quantitative precision and qualitative depth, contributing to a better interpretation of the research findings (Al-Jarf, 2022).

### **2.3. Conceptualization and operationalization of the main categories**

The matrix operationalization explained the independent and dependent variables and their dimensions and indicators (See Annex 1). First, the independent variable, the Text-to-Speech Interaction Platform, serves as an innovation in language learning technology where its user interface dimension goes beyond emphasizing the importance of a learner-friendly design that facilitates navigation (Bonces, 2019). The platform's Speech Synthesis dimension supports language improvement, demanding an evaluation of synthesized speech quality. This is not only technical precision but also the fluency of the spoken language, essential for an authentic language learning experience. Text Highlighting, as another

dimension, seeks to align visual cues with auditory stimulation, aiming for a synchronized and effective aid in pronunciation. Interactive Exercises, the fourth dimension, explain the platform's practical impact by assessing the effectiveness of exercises in improving learners' pronunciation skills.

On the other side, the dependent variable, Language Proficiency Improvement, covers the enhancement of language skills important for effective communication. Reading Accuracy, the first dimension explains not only the percentage of correctly pronounced words but also evaluates how well learners comprehend and accurately interpret sentences. Pronunciation Skills, the second dimension, is a multifaceted evaluation involving the percentage of accurately pronounced sounds, the intonation patterns, and the stress patterns. User Experience and Perceptions, the third dimension, introduces a subjective element, incorporating learners' satisfaction ratings, task completion time, ease of navigation, and perceived effectiveness, providing insights into the emotional and practical dimensions of the learning experience. Lastly, Pedagogical Integration, the fourth dimension, assesses the platform's integration into the educational context, considering factors like the frequency of usage in lessons and the extent to which exercises can be customized to meet the diverse needs of individual learners.

The study's overarching goal is to develop a user-centric text-to-speech interaction platform that not only leverages advanced technologies but also integrates into the pedagogical landscape (Al-Jarf, 2022; Bonces, 2019). By operationalizing these variables, the research not only aims to measure the platform's effectiveness but also to intricate dynamics between its features and

learners' language proficiency development. This methodical approach ensures a comprehensive understanding of the platform's impact, facilitating informed enhancements and contributing insights to the discourse on technology-enhanced language learning.

#### **2.4. Description of the instruments based on the research approach.**

In this study, multiple instruments were employed to gather quantitative and qualitative data, providing an understanding of the impact of technology-enhanced language learning on EFL students. These instruments include reading fluency measures obtained from the TTS Reader platform, independent raters' rankings, and a survey designed to capture students' perspectives.

The data used in this study originates from participants using TTS Reader, a digitally supported reading intervention program available online. The platform incorporates various tasks, such as Key Words, Pre-test, Read Along, Practice, Quiz, and Passing Task. The Passing Task, specifically, involves students reading the text aloud to the instructor, who assesses speech rate and assigns an expression score. To complement the instructor-assigned expression scores, two more independent raters, non-linguists but also non-native English speakers, were engaged to rank students' overall fluency in terms of intelligibility, listener effort, and comprehensibility. Raters were presented with de-identified audio recordings at different stages of the intervention. A survey was administered to the EFL students to gain insights into various aspects of their experience. The survey covered the following key variables:

- **First Impressions:** Students were asked to describe their initial impressions of the TTS reader platform and if those impressions changed over time.
- **Anxiety:** Participants were questioned about any feelings of anxiety or nervousness before or during the read-aloud protocols and the perceived influence on their performance.
- **Self-Perceived Improvement:** Students were prompted to reflect on their perceived improvement in English fluency and identify specific areas of improvement.
- **Language Areas Affected:** Participants were asked to identify which language areas (e.g., pronunciation, prosody, comprehension) were most affected by the read-aloud protocols.
- **Discipline of Study:** Questions assessed students' discipline and commitment to completing the protocols and how well these aligned with their study routine.
- **Personal Satisfaction:** Participants were invited to rate their overall satisfaction with the read-aloud protocols and provide insights into contributing factors.

This mixed-methods approach, combining quantitative data from the platform and raters with qualitative insights from the survey, contributes to the understanding of the impact of technology-enhanced language learning on EFL students.

## **2.5. Description of the validation of Instruments**

Ensuring the reliability and validity of instruments is paramount in any research study. In this research, the validation process primarily involved expert judgment, a recognized method to assess the relevance and clarity of instruments. Expert judgment was chosen due to its suitability for evaluating the appropriateness of the survey questions and the robustness of the TTS reader platform and independent raters' rankings. The survey instrument passed through a rigorous validation process through expert judgment. Those experts, consisting of experienced educators, language learning specialists, and professionals familiar with technology-enhanced language learning, were convened. The experts critically evaluated each survey question for clarity, relevance, and alignment with the study's objectives.

Feedback from the expert panel was instrumental in refining the survey questions to ensure that they accurately captured the intended variables. Additionally, modifications were made to enhance the comprehensibility of questions, reducing the likelihood of misinterpretation by participants. This iterative process of consultation with experts contributed to the face validity and content validity of the survey instrument. For the TTS Reader platform and independent raters' rankings, expert judgment focused on the alignment of these instruments with the research goals. The experts, analyzed the language learning methodologies and assessment techniques, and evaluated the appropriateness of the platform tasks and the reliability of the ranking criteria.

The experts considered the alignment of the platform tasks with established language learning principles and assessed the consistency of the criteria used by independent raters. Their judgments contributed to refining the scoring system and ensuring that the data collected accurately reflected the fluency development of EFL students. The validation process, rooted in expert judgment, enhances the credibility and trustworthiness of the instruments used in this study. The iterative nature of validation allowed for continuous refinement, addressing potential sources of bias or ambiguity.

## **2.6. Population, Sample, And Sampling**

The population for this study comprises students at the University of Babahoyo enrolled in Level 3 of the English program. Specifically, Level 3 students are chosen as the primary sample for their relevance to the research focus on pronunciation skills. Five courses of Level 3 were selected to be studied based on their willingness to participate in the Text-to-Speech (TTS) Interaction Platform. Initially, the sampling procedure involved a systematic approach consisting of using the platform with the whole class, however, only 16 students were selected randomly to be evaluated as the focal group.

Secondly, for the perception survey, we're applying it to the total sample size, determined based on the number of students enrolled in the 5 courses of Level 3 during the time of the study which is 385. Then, using a simple random sampling for a finite population, having 95% confidence and 5% errors, it was obtained 193 randomly applying this instrument to 193 students. Finally, the inclusion criteria

involve students actively enrolled in Level 3 English courses during the study period, ensuring that participants are currently engaged with the curriculum and exposed to the TTS Interaction Platform. This approach ensures that the sample represents the target population adequately and provides insights into the experiences and perceptions of students at a similar language proficiency level.

### **2.7. Data Analysis and Ethical Considerations**

The data analysis for this study involves a two-fold process to examine the impact of the Text-to-Speech (TTS) Interaction Platform. Firstly, for the repetition part on the TTS Interaction Platform, a quantitative approach was employed. The data collected during students' interactions with the platform, such as speech synthesis quality, alignment of text highlighting with speech, and the effectiveness of interactive exercises, was subjected to statistical analysis. Descriptive statistics, including means and standard deviations, was used to summarize the quantitative data. Additionally, inferential statistics, such as t-tests, was applied to assess the significance of differences between pre- and post-assessment scores.

Secondly, for the perception analysis of a larger group of students, qualitative methods were employed. Open-ended survey responses and reflections collected from students went thematic analysis. This qualitative approach allows for an exploration of students' experiences, perceptions, and suggestions regarding the TTS Interaction Platform. Several themes and patterns were identified, providing rich, context-specific insights into the effectiveness and potential improvements of the platform.

On the other hand, ethical considerations in educational research are important, especially when involving human participants. In this study, several ethical principles were considered. First of all, informed consent was obtained from all participants, clarifying the purpose of the study, the nature of their involvement, and their right to withdraw at any stage without consequences. This is because, confidentiality needed to be ensured by anonymizing participants, assigning unique identifiers, and securely storing any personal data. The data collected were used for research purposes, and the results will be presented in an aggregated, non-identifiable format.

The study also prioritized the well-being of participants. Any potential risks or discomfort associated with using the TTS Interaction Platform were minimized, and participants had the opportunity to express concerns. Additionally, the study results were shared with participants, allowing them to benefit from the insights gained. Finally, the research design and procedures have adhered to institutional ethical guidelines and standards.

### **2.8. Initial Diagnose**

The pre-test results provided an initial diagnostic overview of the pronunciation accuracy of students in Level 3 (CEFR A2) across courses F, G, H, I, and J. The scores were based on a rubric, with each student receiving a score out of 10 and it is presented in Table 1:

**Table 1**

*Scores of 16 students of Level 3 which are the intervention group*

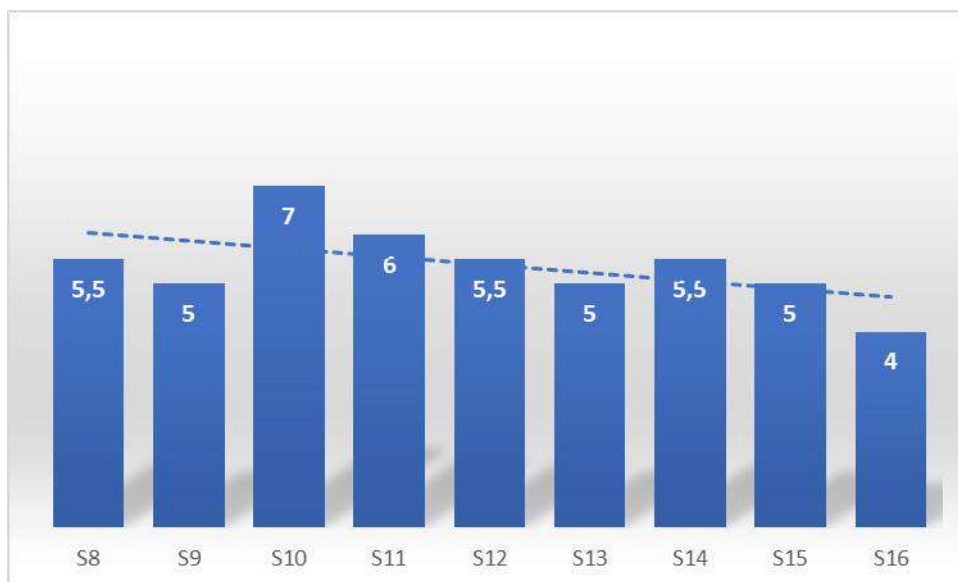
<b>No</b>	<b>Pre-Test Score</b>
S1	3,5
S2	6
S3	5
S4	6
S5	6
S6	3
S7	5
S8	5,5
S9	5
S10	7
S11	6
S12	5,5
S13	5
S14	5,5
S15	5
S16	4

*Note.* This table was elaborated by the Authors.

Initially, the scores exhibited variability, ranging from 3 to 7. This suggested that there is a diversity of proficiency levels within the group. Then, the mean pre-test score was around 5.25, indicating a mid-range level of pronunciation accuracy among the students. While some students, such as S3 and S7, had had relatively higher scores (7/10), indicating a better grasp of pronunciation, there were students, like S1 and S14, with lower scores (3.5/10 and 3/10, respectively), indicating potential challenges in pronunciation. Figure 1 also presents this information.

**Figure 1**

*Students obtained scores in the Pre-test activity*



*Note.* This Figure was elaborated by the Authors.

Continuing the data review, a significant number of students fell within the range of 5 to 6.5, suggesting a common proficiency level within this range. The Pre-test results highlighted areas where students may need improvement, whether it's in specific sounds, stress patterns, or overall pronunciation fluency. Then, these scores could serve as a baseline for comparison with Post-test results, helping to measure the effectiveness of the TTS platform in enhancing pronunciation accuracy.

For that reason, it's essential to consider these results as a starting point, acknowledging the individual strengths and areas of improvement for each student. The TTS platform's impact on these initial scores will be assessed through the post-test, providing valuable insights into the platform's efficacy in improving

pronunciation skills. Thus, Table 2 presents the main statistics obtained from these results:

**Table 2**

*The main statistic of Scores of Level 3 students of the intervention group*

		Intervention
N	Valid	16
	Lost	0
Mean		5,3125
Standard error of mean		0,2447
St. Deviation		1,1147
Variance		1,2431

*Note.* This table was elaborated by the Authors.

The mean pre-test score was 5.3125, which indicated the average performance of the students slightly across the given set of scores. Then, standard error of the mean was 0.2447 which represented the precision of the mean estimate. In this context, it suggests that if we were to take multiple samples and calculate the mean, we would expect the mean to fall within 0.2447 points of the calculated mean of 5.3125.

The standard deviation is 1.1147, providing a measure of how spread out the individual scores are around the mean. A higher standard deviation indicated greater variability in the scores. In this case, the scores had a moderate amount of variability. Moreover, the variance is 1.2431 which values helped understand the

extent to which individual scores deviate from the mean. A higher variance indicates a greater dispersion of scores.

Finally, the mean score suggests an average performance around the midpoint of the scoring range. The standard error indicates the precision of this estimate. The standard deviation and variance highlight the variability and spread of individual scores, providing insights into the distribution of performance among the students.

### **Chapter 3: Research Proposal**

In this chapter, the designed methodology was implemented, beginning with the proposal to provide students at the University of Babahoyo with the opportunity to use the TTS reader app. The chapter then analyzes the results obtained from this process and the designed instruments, followed by a thorough discussion of these results.

#### **3.1. Validation of the proposal**

The validation of the proposed Text-to-Speech (TTS) interaction platform is rooted in both empirical and theoretical foundations, ensuring its viability and feasibility. Empirically, previous research has demonstrated the efficacy of TTS technology in enhancing language learning outcomes, particularly in improving pronunciation accuracy and overall fluency (Rodríguez-Fuentes & Calle-Díaz, 2023). By integrating TTS into the curriculum, the platform leverages proven methodologies to support students in achieving better pronunciation and fluency.

The pilot studies conducted at the University of Babahoyo with Level 3 English students (A2 according to CEFR) provide preliminary evidence that students experience notable improvements in their pronunciation and fluency after consistent use of the TTS platform. These findings are aligned with established theories of language acquisition that emphasize the importance of auditory input and repetition in language learning (Krashen, 1982).

Theoretically, the platform aligns with Vygotsky's socio-cultural theory, which posits that learning is a socially mediated process, and technology can serve

as a valuable tool in this mediation (Vygotsky and Cole, 1978). The interactive exercises and feedback mechanisms of the TTS platform create a dynamic learning environment that supports both autonomous learning and guided practice. Furthermore, the platform's design is informed by the principles of Universal Design for Learning (Griful-Freixenet et al., 2021).

The viability and feasibility of the TTS interaction platform are supported by both technical and pedagogical considerations. Technically, advancements in natural language processing and speech synthesis have made it possible to develop high-quality TTS systems that are capable of producing clear, natural-sounding speech. These systems are increasingly accessible and can be integrated into existing educational infrastructures with minimal disruption. Pedagogically, the platform is designed to complement traditional language instruction by providing students with additional opportunities for practice and immediate feedback.

The benefits of the TTS interaction platform are multifaceted. For students, the platform offers personalized learning experiences that cater to their individual needs and learning paces, thereby enhancing motivation and engagement. The immediate feedback provided by the TTS system helps students to self-correct and improve their pronunciation in real time, leading to more rapid language acquisition. For educators, the platform serves as a valuable tool that can supplement classroom instruction and provide additional support for students who may need extra practice. This can lead to more efficient use of classroom time and allow educators to focus on higher-order teaching tasks.

### **3.2. Theoretical foundations**

The theoretical foundations of this thesis are anchored in several well-established theories and models of language learning, which provide a comprehensive framework for understanding the integration and effectiveness of Text-to-Speech (TTS) technology in enhancing pronunciation and overall language proficiency.

Firstly, Stephen Krashen's Input Hypothesis is central to understanding the role of comprehensible input in language acquisition. According to Krashen (1982), language learners progress in their knowledge of a language when they are exposed to input that is slightly beyond their current level of proficiency ( $i+1$ ). TTS technology provides a valuable source of such input by allowing learners to hear accurate and fluent speech, which they can then mimic and practice. The repetitive nature of TTS systems ensures that learners receive ample exposure to target language sounds, structures, and rhythms, facilitating their acquisition of pronunciation and fluency.

Secondly, Lev Vygotsky's socio-cultural theory emphasizes the importance of social interaction and scaffolding in cognitive development and learning. According to Vygotsky & Cole (1978), learners construct knowledge through mediated interactions with more knowledgeable others or tools. TTS technology acts as a mediating tool that supports language learners by providing models of correct pronunciation and intonation. The interactive exercises and immediate feedback offered by the TTS platform enable learners to engage in self-regulated

learning, while the scaffolding provided through guided practice helps them internalize linguistic patterns and improve their spoken language skills.

Thirdly, Universal Design for Learning (UDL) is a framework that aims to optimize teaching and learning for all people based on scientific insights into how humans learn. UDL advocates for providing multiple means of engagement, representation, and expression to accommodate diverse learners (Rose & Meyer, 2002). The TTS platform embodies the principles of UDL by offering various ways for learners to interact with content. The auditory feedback and visual highlighting of text cater to different learning preferences and needs, ensuring that learners with varying abilities and backgrounds can benefit from the technology. This inclusivity makes the TTS platform a versatile tool in language education.

Additionally, Communicative Language Teaching (CLT) emphasizes the importance of interaction as both the means and the ultimate goal of learning a language (Ng, 2020). CLT advocates for the use of authentic language in real-life contexts to promote communicative competence. TTS technology supports CLT by providing learners with opportunities to practice speaking and listening in a controlled yet realistic environment (Losi & Nasution, 2022). The platform's interactive exercises simulate conversational scenarios, enabling learners to develop their communicative skills in a meaningful and practical way.

Subsequently, the Cognitive Load Theory focuses on working memory capacity and the importance of managing cognitive load during learning (Janssen & Kirschner, 2020). TTS technology helps to manage the cognitive load by

breaking down complex language tasks into manageable parts and providing auditory and visual support. The use of TTS can reduce the extraneous cognitive load associated with decoding text, allowing learners to focus more on pronunciation and comprehension (Paas & van Merriënboer, 2020).

Finally, the Behaviorist theories, particularly those related to operant conditioning, also underpin the use of TTS technology in language learning (Anastasiu & Cyprien, 2021). The repetitive practice and immediate feedback mechanisms inherent in TTS systems align with the principles of reinforcement and shaping. Positive reinforcement through accurate pronunciation models and corrective feedback helps to strengthen desirable language behaviors, gradually leading to the improvement of pronunciation and fluency.

### **3.3. Characteristics of the proposal**

The proposed Text-to-Speech (TTS) interaction platform is designed to enhance the pronunciation accuracy and overall language proficiency of second language (L2) learners. It leverages advanced natural language processing and speech synthesis technologies to provide a user-friendly, interactive, and engaging learning experience. The platform incorporates several key features aimed at facilitating effective language learning. These include intuitive user interfaces, high-quality synthesized speech, synchronized text highlighting, and interactive exercises. Each of these features is meticulously designed to provide learners with immediate auditory and visual feedback, allowing them to practice and refine their pronunciation skills in a structured and supportive environment.

One of the core characteristics of the proposal is its adaptability and inclusivity, aligning with the principles of Universal Design for Learning (UDL). The TTS platform caters to a diverse range of learners, accommodating various learning styles and needs. It provides multiple means of representation and engagement, enabling learners to interact with the content in ways that suit their individual preferences. For instance, learners can listen to the pronunciation of specific words by clicking on them, repeat exercises to reinforce learning, and receive instant feedback to correct errors. This flexibility ensures that all learners, regardless of their backgrounds or abilities, can benefit from the platform.

Moreover, the proposal emphasizes the integration of the TTS platform into regular classroom activities and curricula. The platform is designed to be easily customizable by instructors, allowing them to custom exercises and tasks to meet the specific needs and goals of their students. This feature promotes seamless pedagogical integration, ensuring that the TTS platform complements and enhances traditional teaching methods rather than replacing them. Additionally, the platform includes tools for tracking and assessing student progress, providing valuable data that instructors can use to inform their teaching strategies and provide targeted support. By combining technological innovation with educational best practices, the proposal aims to create a comprehensive and effective tool for improving L2 pronunciation and fluency.

### **3.4. Structure, and dynamics of the components of the proposal**

The proposal involves a structured intervention aimed at enhancing pronunciation accuracy and overall language proficiency among A2-level students at the University of Babahoyo, specifically those enrolled in Level 3 at CENID. The intervention period spans 30 days, during which students engage daily with a Text-to-Speech (TTS) reader platform. This process is bookended by a pre-test and a post-test to measure the impact of the intervention.

#### **3.4.1. Pre-test before intervention**

**Objective:** The pre-test aims to establish a baseline for students' pronunciation accuracy and fluency.

**Date and Duration:** The pre-test was conducted on January, 15<sup>th</sup>-16<sup>th</sup> 2024.

**Participants:** Sixteen A2-level students were randomly selected from Level 3 of CENID at the University of Babahoyo.

**Procedure:**

- a. Students were given a standardized reading passage appropriate for their proficiency level.
- b. They read the passage aloud while being recorded.
- c. The recordings were evaluated using a rubric that measured aspects such as pronunciation accuracy, intonation, stress patterns, and overall fluency. Each performance was scored on a scale from 1 to 10, with detailed feedback provided to the students.

### 3.4.2. 30-Day Intervention Period

**Objective:** To improve students' pronunciation accuracy and fluency through consistent practice using a TTS reader platform.

**Dates:** The intervention period ran from January to February 2024.

**Structure:**

- a) Daily Practice: Students were required to practice reading using the TTS reader platform daily. Each session lasted approximately 30 minutes.
- b) Key Words: Students familiarized themselves with 3-4 essential words from the day's reading.
- c) Pre-Test (within the platform): Students read the text for the first time independently, timed by the platform to establish a personal benchmark for the session.
- d) Read Along: The text was read by a recorded voiceover, allowing students to hear accurate pronunciation, prosody, and speech rate.
- e) Practice: Students read the text again with the option to click on any word for pronunciation help. Speech rate scores were provided for self-assessment.
- f) Quiz: Multiple-choice questions assessed reading comprehension, with immediate feedback provided.
- g) Passing Task: Students read the text aloud to an instructor for final assessment, focusing on speech rate and prosodic markers. A holistic expression score was assigned.

**Feedback and Support:** Instructors provided individual feedback based on the daily reading sessions, helping students identify and work on specific pronunciation issues.

### 3.4.3. Post-Test activity after intervention

**Objective:** To evaluate the improvement in pronunciation accuracy and fluency after the intervention.

**Date and Duration:** The post-test was conducted on February 20<sup>th</sup>, 2024.

#### **Procedure:**

- a) Students were given a new standardized reading passage of similar difficulty to the pre-test passage.
- b) They read the passage aloud while being recorded.
- c) The recordings were evaluated using the same rubric as the pre-test, allowing for direct comparison of scores.

### 3.4.4. Analysis and Results

The Pre-test and Post-test scores were analyzed to determine the effectiveness of the TTS intervention. Key metrics included:

- **Pronunciation Accuracy:** Measured by the percentage of correctly pronounced words.
- **Fluency:** Assessed through intonation, stress patterns, and overall reading fluidity.

- **Self-Perception:** Students' reflections on their improvement, confidence, and overall satisfaction with the intervention were collected through a survey.

The proposed intervention, structured around consistent daily practice with the TTS reader platform, is designed to provide measurable improvements in pronunciation and fluency for A2-level students. This structured approach not only aims to enhance linguistic skills but also seeks to build students' confidence and self-efficacy in their language-learning path.

### **3.5. Description of the methodological requirements**

The methodological requirements for this study are centered around a mixed-methods approach, specifically utilizing an explanatory sequential design. This design involves first collecting and analyzing quantitative data, followed by qualitative data to provide further insight into the quantitative results. The study involved sixteen randomly selected students from Level 3 (A2 according to the CEFR) at the University of Babahoyo, ensuring a representative sample. The data collection instruments included pre-test and post-test assessments using standardized reading passages to evaluate pronunciation accuracy and fluency. The TTS reader platform was the primary tool for daily practice sessions, supplemented by a Likert scale survey to capture students' perceptions of the platform, focusing on aspects such as first impressions, anxiety, self-perceived improvement, language areas affected, discipline of study, and personal satisfaction.

The intervention procedure was structured to facilitate effective learning. Over a 30-day period, students engaged with the TTS platform for approximately 30 minutes each day. These sessions included keyword familiarization, pre-test reading, read-along exercises, practice readings, quizzes, and a passing task with instructor feedback. Regular instructor support was provided to help students improve their pronunciation and fluency.

The technological requirements were equally important to the success of the study. The TTS reader platform needed to incorporate advanced speech synthesis technology to provide accurate and natural-sounding voice output. Key interactive features included text highlighting synchronized with speech, clickable words for pronunciation help, and adaptive interactive exercises. A user-friendly interface was essential to maintain student engagement and facilitate effective learning.

Students needed access to devices such as computers, tablets, or smartphones capable of running the TTS platform smoothly, along with reliable internet connections. Good quality recording equipment and headsets were required to ensure clear audio for Pre-test and Post-test evaluations and minimize background noise during practice sessions. Then, data management was another important aspect, requiring secure storage of all collected data to ensure privacy and confidentiality. Additionally, ongoing technical support was essential to address any issues with the TTS platform or related technology.

### 3.6. Results and Discussion

The results were obtained through the approach employed in this research project. This was reflected in the grades obtained in the pre-and post-test and the student surveys analysis.

#### 3.6.1. Intervention results

Initially, a Pre-test activity was made, and it was already presented in the last chapter. Then, a Post-test activity was implemented and evaluated with the students' sample.

##### 3.6.1.1. Post-test results

The Post-test results provided a final overview of the pronunciation accuracy of students in Level 3 (CEFR A2) across courses F, G, H, I, and J. The scores were based on a rubric, with each student receiving a score out of 10 and it is presented in Table 3:

**Table 3**

*Post-test scores of 16 students of Level 3 which are the intervention group.*

No	Post-Test scores
S1	4
S2	6
S3	8,5
S4	7
S5	6
S6	6
S7	5,5

S8	6
S9	6
S10	8
S11	6
S12	6
S13	8
S14	5
S15	7
S16	7

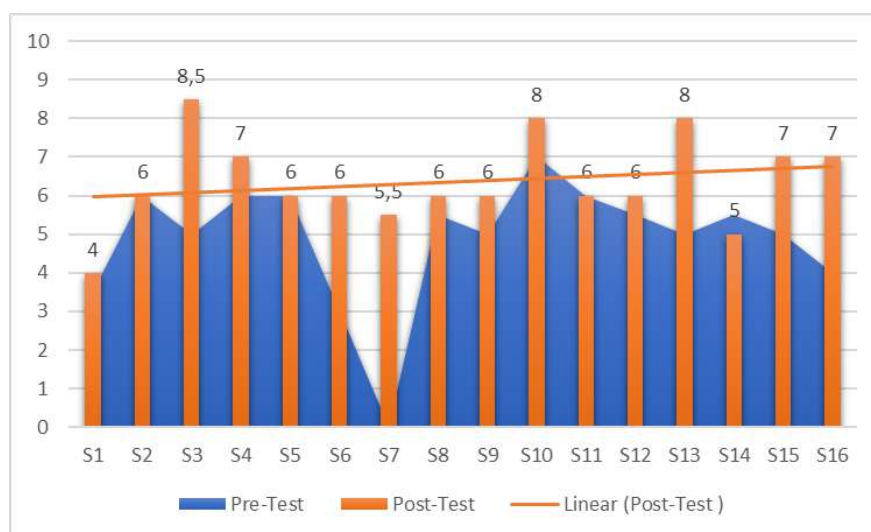
---

*Note.* This table was elaborated by the Authors.

As a potential increase, the scores show variability, ranging from 4 to 8,5 for the intervention period. This suggested much of an improvement as there is a diversity of proficiency levels within the group. Then, Figure 2 presents the variation among Pre- and Post-tests.

**Figure 2**

*Pre- and Post-tests variations*



*Note.* This table was elaborated by the Authors.

According to Figure 2, it is graphically evident an improvement in the Post-test among the intervened group of students. Thus, Table 4 presents the main statistics obtained from these results:

**Table 4**

*Post-test main statistic of Scores of Level 3 students of the intervention group*

		<b>Intervention</b>
N	Valid	16
	Lost	0
Mean		6.375
Standard error of mean		0,290
St. Deviation		1,1619
Variance		1,35

*Note.* This table was elaborated by the Authors.

The results of the intervention using the TTS reader platform for Level 3 students at the University of Babahoyo reveal significant insights into the effectiveness of the program. The study included a valid sample of 16 students, with no participants lost during the intervention period. The Post-test scores showed an improvement, with a mean score of 6.375, indicating that the average performance of the students was substantially higher after the intervention compared to their Pre-test scores.

The standard error of the mean was 0.290, suggesting a relatively high level of precision in the sample mean as an estimate of the population mean. This low

standard error highlights the consistency of the students' performance improvements and implies that similar improvements might be expected in the population of Level 3 students using the TTS platform.

The standard deviation of 1.1619 reflects the degree of variation in the students' Post-test scores around the mean. A standard deviation of this magnitude indicates moderate variability, meaning that while there was some variation in how much each student improved, most scores were clustered around the mean score of 6.375. This suggests that the intervention was generally effective across the majority of participants.

Finally, the variance of 1.35, which is the square of the standard deviation, further supports the observation of moderate dispersion in the scores. This level of variance indicates that while there were differences in individual outcomes, these differences were not extreme, reinforcing the overall effectiveness of the TTS reader platform in enhancing pronunciation skills among the participants.

As a result, this statistical analysis demonstrates that the 30-day intervention period, during which students practiced daily with the TTS reader, was successful in improving their reading pronunciation skills. The consistency variability in the scores suggests that the TTS platform could be a reliable tool for similar educational contexts.

#### **3.6.1.2. Paired T-Test**

It was used the paired t-test as an appropriate method to explore effectiveness because it accounts for the fact that the same participants are being

measured twice, which helps to control individual variability. Here's how you can conduct the paired t-test and interpret its results:

- Null Hypothesis (H0): There is no significant difference between the Pre-test and Post-test scores.
- Alternative Hypothesis (H1): There is a significant difference between the Pre-test and Post-test scores.

The formula used was:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Where:

- $\bar{d}$  is the mean of the differences.
- $S_d$  is the standard deviation of the differences.
- $\sqrt{n}$  is the number of pairs.

Then, if the calculated t-value is greater than the critical t-value, we reject the null hypothesis and conclude that there is a significant difference between the Pre-test and Post-test scores. On the other hand, if the calculated t-value is less than or equal to the critical t-value, we fail to reject the null hypothesis and conclude that there is no significant difference between the Pre-test and Post-test scores.

Let's conduct this paired t-test using the Pre-test and Post-test scores provided.

**Table 5***Paired T-test results*

<b>Calculations</b>	<b>Values</b>
T-Statistic:	6.75
P-Value:	0,000657
Degrees of Freedom:	15
Standard Deviation of Differences:	0.704
Standard Error of the Mean Difference (SEM):	0.176

*Note.* This table was elaborated by the Authors.

The t-statistic of 6.75 is substantially higher than the critical value of t for 15 degrees of freedom at any conventional significance level (e.g., 0.05, 0.01). then, the p-value is 0,000657 which is much less than 0.05, indicating that the difference in scores is highly statistically significant. Subsequently, the mean difference between the Post-test and Pre-test scores is 1.1875. This positive mean difference indicates an overall improvement in scores after the intervention. Finally, the standard deviation of the differences is 0.704, and the standard error of the mean difference is 0.176. These values suggest that the differences are consistent among students and the estimation of the mean difference is precise.

Based on the results of the paired t-test, we reject the null hypothesis and conclude that there is a statistically significant improvement in the scores from the

Pre-test to the Post-test. The intervention involving the use of the TTS platform appears to have had a positive impact on the students' pronunciation accuracy and overall language proficiency. This significant improvement reinforces the initial idea of this study which is to validate the efficacy of the TTS platform in enhancing language learning outcomes for A2-level students at the University of Babahoyo.

### 3.6.2. Descriptive statistics based on students' perceptions.

Based on the surveys launched to 145 randomly selected students, this part looked to understand students' perceptions and the impact generated using a TTS reader platform in class. Table 6 informed descriptive statistics of the main students' demographic variables.

**Table 6**

*Demographic statistics of survey*

<b>Variables</b>	<b>Mean</b>	<b>Min</b>	<b>Max</b>
<b>Age</b>	23,02	18	39
<b>Gender</b>	<b>Male</b>	<b>Female</b>	
	31,72%	68,28%	
<b>Location</b>	<b>Babahoyo</b>	<b>Other - Urban</b>	<b>Other - Rural</b>
	37,93%	37,24%	24,83%
<b>Semester</b>	<b>3th - 4<sup>th</sup></b>	<b>5th - 6th</b>	<b>7th - 8th</b>
	59,31%	28,97%	11,72%
<b>Field of study</b>	18		

*Note.* This table was elaborated by the Authors.

The demographic characteristics of the student sample explain different components, firstly, the mean age of the students surveyed is 23.02 years, with ages ranging from 18 to 39 years. This indicates a relatively young cohort, predominantly in their early twenties, which is typical for university-level students.

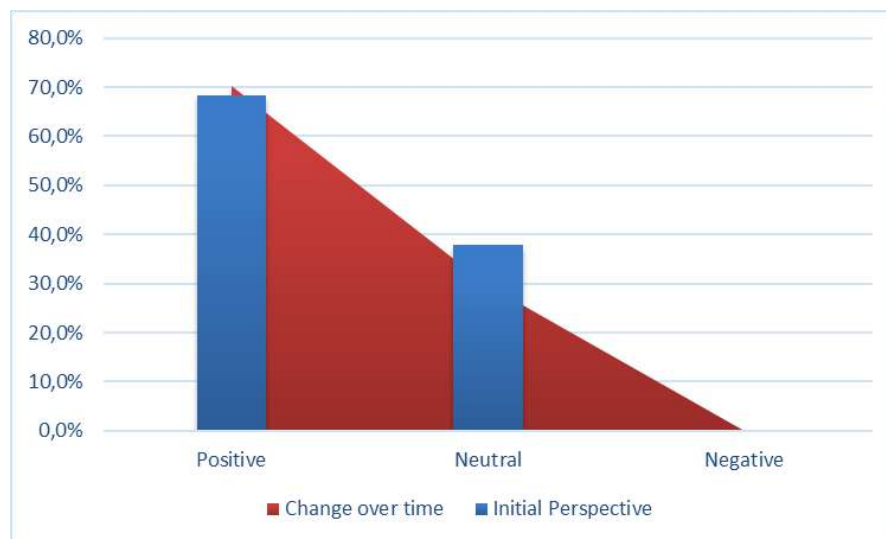
In terms of gender distribution, the sample consists of 31.72% male and 68.28% female students. This significant gender disparity suggests that there are more female participants in the survey, potentially reflecting the gender distribution in the university's English language courses. This aspect could be relevant when analyzing responses, as gender differences might influence perceptions and experiences with language learning technologies (Rodríguez-Fuentes & Calle-Díaz, 2023).

On the other hand, the students come from various locations, with 37.93% residing in Babahoyo, 37.24% from other urban areas, and 24.83% from rural areas. This geographic representation was expected due to the nature of a small and more rural approach university, being this diversity is important for assessing the accessibility of the TTS platform across different settings (Yang et al., 2022).

Additionally, the majority of the respondents are in their 3rd to 4th semesters, accounting for 59.31% of the sample. Students in their 5th to 6th semesters make up 28.97%, while those in their 7th to 8th semesters comprise 11.72%. Then, Figure 3 explains the perspective of students over time about the TTS reader platform.

**Figure 3**

*The perspective of students about the platform over time*



*Note.* This Figure was elaborated by the Authors.

The students' perspectives on the TTS platform show a generally positive trend over time. Initially, 68.3% of students had a positive view of the platform, with 37.9% remaining neutral and no negative responses recorded. Over time, the positive perspective slightly increased to 70.3%, while the neutral responses decreased to 29.7%, and still, no negative responses were reported. This shift indicates that as students continued to use the TTS platform, their satisfaction and positive perception of its effectiveness in improving their language skills grew.

In addition, it was asked why this perception could change and the students' feedback on the TTS platform highlights a positive impact, many students expressed high levels of satisfaction, noting that the platform allowed them to achieve things in English they previously thought unattainable. Initially skeptical,

some students found the TTS platform to be more effective than tools like Google Translate, particularly for understanding and practicing pronunciation.

While the platform's ability to provide clear pronunciation models and immediate feedback was frequently praised, students also reported improvements in their pronunciation, comprehension, and overall confidence in speaking English. The flexibility to adjust speech speed was also appreciated, as it helped them better understand and mimic the pronunciation. Then, Table 7 presents the usage rated by the students.

**Table 7**

*Usage of the TTS reader rated by students*

<b>Variable</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
The usage of the platform	4,14%	11,03%	32,41%	52,41%
The time used to complete the activity	2,76%	17,93%	33,79%	45,52%
Internet usage during the activity	4,14%	8,97%	39,31%	47,59%
The willingness to perform the tasks	2,07%	12,41%	33,79%	51,72%

*Note.* This table was elaborated by the Authors.

The survey results provide insights into the students' perceptions and experiences with the TTS platform across several dimensions. Firstly, regarding the usage of the platform, a significant majority of students (52.41%) rated their

engagement at the highest level, indicating frequent and consistent use. An additional substantial portion (32.41%) also reported high usage, suggesting that the platform was widely adopted and utilized. In contrast, fewer students rated their usage at moderate (11.03%) or minimal levels (4.14%), showing a generally positive reception.

In terms of the time used to complete activities, the majority of students (45.52%) reported spending a substantial amount of time, indicating thorough engagement with the material. A sizable group (33.79%) spent a moderate amount of time, reflecting a balanced approach to task completion. Meanwhile, smaller percentages of students spent less time, with 17.93% reporting moderate engagement and only 2.76% indicating minimal engagement.

The survey also highlighted internet usage during activities, with nearly half of the students (47.59%) utilizing the internet extensively. This suggests that the platform's integration with online resources was significant. Another 39.31% reported moderate internet usage, indicating that online components were a regular part of their interaction with the platform. Fewer students reported low (8.97%) or minimal (4.14%) internet usage, further supporting the importance of internet access in their learning process.

Finally, regarding the willingness to perform tasks on the platform, over half of the students (51.72%) expressed a strong willingness, highlighting high motivation and engagement. A significant portion (33.79%) also showed a good level of willingness, while fewer students were less inclined, with 12.41% at a

moderate level and only 2.07% indicating minimal willingness. These results suggest that the TTS platform was well-received and effectively utilized by the students, contributing positively to their learning experience. Then, Table 8 analyze the level of agreement with some statements related to their perception.

**Table 8**

*Students' perception of the intervention*

Variable	Strongly disagree	Disagree	Agree	Strongly agree
1. The user interface of the TTS Interaction Platform is intuitive and easy to navigate.	5,52%	8,28%	36,55%	49,66%
2. The quality of the synthesized speech on the platform is clear and natural.	3,45%	12,41%	38,62%	45,52%
3. The alignment of text highlighting with speech enhances my ability to follow and comprehend the text.	4,83%	11,72%	33,79%	49,66%
4. The interactive exercises on the platform are engaging and contribute to my learning experience.	4,14%	9,66%	33,79%	52,41%
5. The TTS Interaction Platform has been effectively integrated into my language learning lessons.	3,45%	11,03%	36,55%	48,97%

*Note.* This table was elaborated by the Authors.

Analyzing Table 8, firstly, regarding the user interface, a substantial majority of students (49.66%) "strongly agree" that the platform is intuitive and easy to navigate. Additionally, 36.55% "agree" with this statement, indicating a high overall satisfaction with the platform's usability. Only a small percentage of

students "disagree" (8.28%) or "strongly disagree" (5.52%), suggesting that the interface generally meets user expectations.

Secondly, the quality of the synthesized speech is perceived positively, with 45.52% of students "strongly agreeing" that it is clear and natural. Another 38.62% "agree," reflecting a significant endorsement of the speech quality. However, some students expressed reservations, with 12.41% "disagreeing" and 3.45% "strongly disagreeing," pointing to a potential area for improvement.

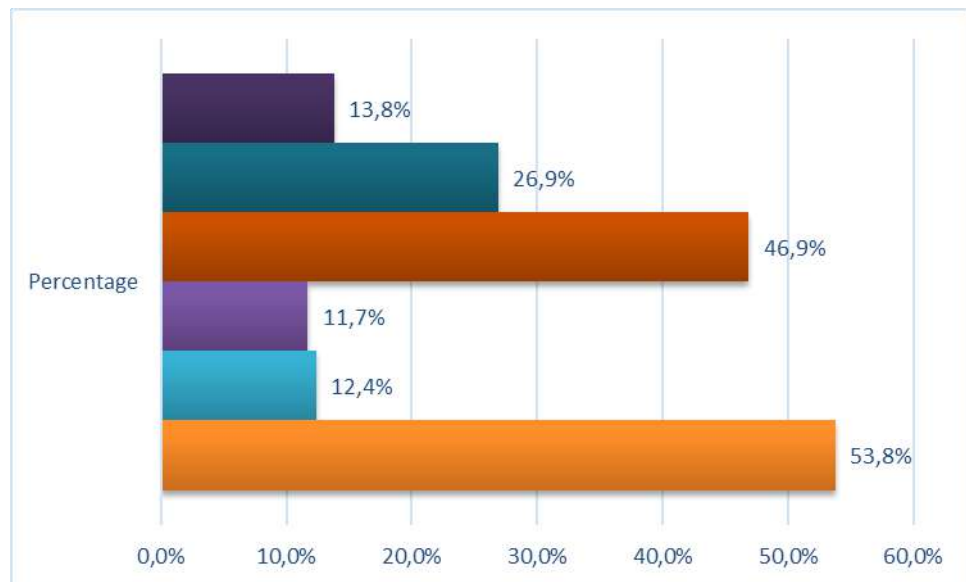
Thirdly, the alignment of text highlighting with speech is highly regarded, as 49.66% of students "strongly agree" that it enhances their ability to follow and comprehend the text. An additional 33.79% "agree," indicating widespread approval. A smaller portion of the students "disagree" (11.72%) or "strongly disagree" (4.83%), suggesting that most users find this feature beneficial. Fourthly, the interactive exercises on the platform are considered engaging and contributory to the learning experience, with 52.41% of students "strongly agreeing" and 33.79% "agreeing." Only 9.66% "disagree" and 4.14% "strongly disagree," showing that the interactive elements are well-received and effective in promoting learning.

Lastly, the integration of the TTS Interaction Platform into language learning lessons is viewed positively by most students. Nearly half (48.97%) "strongly agree" that it has been effectively integrated, and 36.55% "agree." A minority of students "disagree" (11.03%) or "strongly disagree" (3.45%), indicating that while the platform's integration is generally effective, there is some room for improvement. Thus, results suggest that the TTS Interaction Platform is well-

regarded by most students for its intuitive interface, high-quality synthesized speech, effective text-speech alignment, engaging exercises, and successful integration into language learning lessons. Figure 4 presents the aspects that students consider they have had improvements using the platform.

**Figure 4**

*Aspects of improvement of using the platform*



*Note.* This Figure was elaborated by the Authors.

Then, Figure 4 presents data on different aspects of language skills improvement as perceived by students using the TTS Interaction Platform. The most significant area of improvement reported by students is pronunciation, with 53.8% of them highlighting this aspect. This high percentage indicates that the platform is particularly effective in helping students enhance their pronunciation skills, likely due to the clear speech provided by the TTS technology.

The second highest area of improvement is words recognition, noted by 46.9% of the students. This suggests that the platform aids significantly in helping students recognize and understand individual words better, an essential component of language learning that underpins both reading and listening skills. Then, comprehension is the third most improved aspect, with 26.9% of students indicating a better understanding of the text. This improvement could be attributed to the alignment of text highlighting with speech, which helps students follow along more effectively and grasp the meaning of the content being read.

Fluency, including the smoothness and flow of speech, saw improvement for 13.8% of the students. This lower percentage compared to pronunciation and word recognition might suggest that while the platform aids in accurate pronunciation, additional practice might be needed for students to achieve natural fluency. Additionally, intonation, the variation of pitch in speech, was noted as an area of improvement by 12.4% of the students. This shows that the platform helps some students with the melodic aspects of language, though it may not be as pronounced an area of improvement as pronunciation or word recognition.

Lastly, phrase generation, or the ability to construct sentences and phrases, was identified by 11.7% of students as an improved skill. This indicates some benefit in constructing language, although it appears to be the least impacted area among the listed aspects. To sum up, the data in Figure 2 suggests that while the TTS Interaction Platform effectively improves specific areas like pronunciation and word recognition, it also provides some benefits across language skills in students.

### 3.6.3. Discussion

The Pre-test and Post-test scores highlight a clear improvement in students' pronunciation and overall language skills after the 30-day intervention period using the TTS platform. The mean Pre-test score was 5.18, while the mean Post-test score increased to 6.38. This significant improvement is further validated by the paired samples T-test, which yielded a t-value of -5.98 and a p-value  $< 0.0001$ , indicating that the changes in scores were not due to random chance but rather the result of the intervention.

Nachmani and Wolf (2019) analysis can be applied to explain these results, the hypothesis emphasizes the importance of providing learners with input slightly above their current proficiency level ( $i+1$ ). The TTS platform, with its clear and natural synthesized speech, provided students with a consistent model of accurate pronunciation and fluent speech, which they could mimic and practice. The repeated exposure to this input likely contributed to the improvement in their pronunciation and overall language proficiency.

The descriptive statistics provide further insights into the students' experiences and perceptions of the TTS platform. With a mean score of 6.38 in the post-test, the standard error of the mean was 0.29, and the standard deviation was 1.16, indicating a relatively consistent improvement across the student cohort. This consistency supports socio-cultural theory (Erbil, 2020), which emphasizes the role of social interaction and scaffolding in cognitive development. The TTS platform

acted as a scaffolding tool, providing students with immediate feedback and interactive exercises that supported their learning process.

In addition to pronunciation, the survey results showed that 46.9% of students reported improved word recognition, and 26.9% noted better comprehension. These findings align with the principles of Universal Design for Learning (UDL), which advocates for multiple means of engagement and representation to accommodate diverse learners (Griful-Freixenet et al., 2021). The TTS platform's combination of auditory feedback and visual text highlighting catered to different learning preferences, enhancing students' ability to follow and comprehend the text.

The survey data also revealed that 13.8% of students experienced improved fluency, while 12.4% and 11.7% reported enhancements in intonation and phrase generation, respectively. While these areas showed less dramatic improvement compared to pronunciation and word recognition, they highlight the need for more interactive and context-based practice to achieve substantial gains. This observation is consistent with the principles of Communicative Language Teaching (CLT), which emphasizes authentic language use in meaningful contexts to develop communicative competence (Chen and Han, 2020).

Moreover, behaviorist theories, particularly those related to operant conditioning (Anastasi and Cyprien, 2021), underpin the effectiveness of the TTS platform. The repetitive practice and immediate feedback mechanisms inherent in the TTS systems align with the principles of reinforcement and shaping, helping

students gradually improve their pronunciation and other language skills through positive reinforcement.

### **3.7. Limitations in the study**

Despite the positive findings, several limitations must be acknowledged in this study. Firstly, the sample size of 16 students, while providing valuable insights, is relatively small and may not be representative of the broader population of A2 level students. This limitation impacts the generalizability of the results, making it challenging to apply the findings to other contexts or larger groups. Secondly, the study was conducted over a short period of 30 days. Although this timeframe was sufficient to observe initial improvements in pronunciation and other language skills, it may not capture long-term effects or the sustainability of the gains achieved. Extended studies over several months or semesters would provide a more comprehensive understanding of the TTS platform's effectiveness.

Thirdly, the intervention relied solely on self-reported measures of student engagement and perceptions, which can be subject to bias. While these self-reports are valuable, incorporating objective measures such as recorded speech samples before and after the intervention could provide a more robust assessment of actual improvement in pronunciation and other skills. Additionally, the study did not account for potential external factors that might have influenced the results.

Variables such as students' prior exposure to English, the quality of their previous language instruction, and their individual learning strategies were not controlled for. These factors could have impacted the effectiveness of the TTS

platform and the overall results of the study. Lastly, the study focused predominantly on the TTS platform's impact on pronunciation and word recognition, with less emphasis on other critical language skills, then, future research should adopt a more holistic approach, evaluating the platform's effectiveness across a broader range of language competencies.

## **Conclusions**

The study aimed to evaluate the effectiveness of a Text-to-Speech (TTS) platform in enhancing the pronunciation and overall language proficiency of A2-level students at CENID. The findings revealed significant improvements in students' pronunciation and word recognition skills, supported by both quantitative and qualitative data. The Pre-test and Post-test comparisons demonstrated progress, with statistical analysis confirming the significance of the observed improvements.

Students' perceptions of the platform were positive, indicating high levels of engagement and satisfaction. The user-friendly interface, clear and natural synthesized speech, and interactive exercises were highlighted as key factors contributing to their positive learning experience. These features facilitated a better understanding of pronunciation and encouraged active participation, reinforcing the principles of Communicative Language Teaching (CLT) and Cognitive Load Theory.

However, the study also identified several limitations, including the small sample size, short intervention period, and reliance on self-reported data. These limitations suggest the need for further research with larger, more diverse samples and extended study durations to confirm the findings and explore long-term effects. Additionally, incorporating objective measures of language improvement and considering external influencing factors would provide a more comprehensive assessment of the TTS platform's effectiveness.

Despite these outcomes, the study also highlights several challenges that EFL teachers face in conducting classes entirely in the target language. One significant difficulty is ensuring that students comprehend instructions and lesson content when delivered exclusively in English, especially at lower proficiency levels. This can lead to frustration and disengagement if students struggle to follow along.

Additionally, teachers may find it challenging to balance the need for immersive language exposure with the practical necessity of ensuring understanding and maintaining classroom management. These challenges underscore the importance of supplementary tools like TTS, which can offer pronunciation models and aid comprehension through auditory and visual support.

In conclusion, the TTS platform has shown substantial potential in supporting language learning, particularly in improving pronunciation and word recognition among A2 level students. The positive feedback from students and the significant improvements observed in their language skills underscore the platform's value as a complementary educational tool. Future studies should aim to address the identified limitations and explore broader applications of TTS technology in diverse language learning contexts, ensuring that its benefits are maximized for all learners.

## **Recommendations**

Based on the findings of this study, several recommendations can be made to enhance the use and effectiveness of Text-to-Speech (TTS) technology in language learning environments. Firstly, educational institutions should consider integrating TTS platforms into their language curricula, this integration should include structured and regular use of TTS technology in various classroom activities. Additionally, teachers should receive training on how to effectively incorporate TTS tools into their lessons, maximizing their potential to support diverse learning needs.

To address the limitations identified in this study, future research should involve larger and more diverse sample sizes to generalize the findings across different learner populations. Extended intervention periods are also recommended to assess the long-term impact of TTS technology on language proficiency. Moreover, incorporating a mixed-methods approach that includes both qualitative feedback and objective performance measures, such as standardized language tests, can provide a more comprehensive evaluation of the technology's effectiveness.

It is also recommended to customize TTS platforms to better support individual learner differences, which involve features that allow for personalized learning paths, adjustable speech rates, and varied levels of complexity in practice activities. Additionally, it could provide more options for different accents and dialects can help learners become accustomed to the variations in spoken language, thereby improving their overall communicative competence.

### BIBLIOGRAPHICAL REFERENCES

- Acosta, H., & Cajas, D. (2018). Analysis of teaching resources used in EFL classes in selected Ecuadorian universities. *Indonesian Journal of Applied Linguistics*, 8(1), 100–109.
- Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 373–383.
- Afonso, M., Almeida, P., Beça, P., & Silva, T. (2022). *Usability Of Text-To-Speech Technology in Creating News Podcasts using Portuguese Of Portugal*.
- Alemi, M., & Khatoony, S. (2020). Virtual reality assisted pronunciation training (VRAPT) for young EFL learners. *Teaching English with Technology*, 20(4), 59–81.
- Al-Jarf, R. (2022). Text-To-Speech Software for Promoting EFL Freshman Students' Decoding Skills and Pronunciation Accuracy. *Journal of Computer Science and Technology Studies*, 4, 19–30.  
<https://doi.org/10.32996/jcsts.2022.4.2.4>
- Anastasiu, U., & Cyprien, T. (2021). Theories underpinning language acquisition/learning: behaviourism, mentalist and cognitivism. *International Journal of Contemporary Applied Researches*, 8(4).
- Arroba, J., & Acosta, H. (2021). Authentic digital storytelling as alternative teaching strategy to develop speaking skills in EFL classes. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 317–343.

- Benitez-Correa, C., Cabrera-Solano, P., Solano, L., & Espinoza-Celi, V. (2020). Improving Past Tense Pronunciation of Regular Verbs through the Use of Audacity: A Case Study of EFL Undergraduate Students in Ecuador. *Teaching English with Technology*, 20(1), 3–20.
- Bonces, J. A. R. (2019). IMPLEMENTING TEXT-TO-SPEECH TECHNOLOGY AS A MEANS OF ENHANCING L2 READING FLUENCY. In *The International Education & Learning Review* (Vol. 1, Issue 2).  
<http://journals.epistemopolis.org/index.php/learner/>
- Chen, H. C., & Han, Q. W. (2020). Designing and implementing a corpus-based online pronunciation learning platform for Cantonese learners of Mandarin. *Interactive Learning Environments*, 28(1), 18–31.  
<https://doi.org/10.1080/10494820.2018.1510422>
- Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350.
- Chen, Y., Assael, Y., Shillingford, B., Budden, D., Reed, S., Zen, H., Wang, Q., Cobo, L. C., Trask, A., & Laurie, B. (2018). Sample efficient adaptive text-to-speech. *ArXiv Preprint ArXiv:1809.10460*.
- Danielescu, A., Horowitz-Hendler, S. A., Pabst, A., Stewart, K. M., Gallo, E. M., & Aylett, M. P. (2023). Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

- Delić, V., Perić, Z., Sečujski, M., Jakovljević, N., Nikolić, J., Mišković, D., Simić, N., Suzić, S., & Delić, T. (2019). Speech technology progress based on new machine learning paradigm. *Computational Intelligence and Neuroscience, 2019*.
- Dongmei, L. (2021). Design of English text-to-speech conversion algorithm based on machine learning. *Journal of Intelligent & Fuzzy Systems, 40*(2), 2433–2444.
- Encalada, M. A. R., & Sarmiento, S. M. A. (2019). Perceptions about Self-recording Videos to Develop EFL Speaking Skills in Two Ecuadorian Universities. *Journal of Language Teaching & Research, 10*(1).
- Erbil, D. G. (2020). A review of flipped classroom and cooperative learning method within the context of Vygotsky theory. *Frontiers in Psychology, 11*, 1157.
- Fadillah, A. C. (2020). Pronunciation Difficulties of EFL Learners in Indonesia: A Literature Study. *Jurnal Budaya Brawijaya, 2*(1), 7–13.
- Frissell, N. A. (2019). *Text as speech: a new analytical framework for computer mediated communication*.
- Gilakjani, A. P. (2017). A Review of the Literature on the Integration of Technology into the Learning and Teaching of English Language Skills. *International Journal of English Linguistics, 7*(5), 95.  
<https://doi.org/10.5539/ijel.v7n5p95>
- Gottardi, W., Almeida, J. F. de, & Tumolo, C. H. S. (2022). Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. *Texto Livre, 15*, e36736.

- Griful-Freixenet, J., Struyven, K., & Vantieghem, W. (2021). Exploring pre-service teachers' beliefs and practices about two inclusive frameworks: Universal Design for Learning and differentiated instruction. *Teaching and Teacher Education, 107*, 103503. <https://doi.org/10.1016/J.TATE.2021.103503>
- Huensch, A. (2019). Pronunciation in foreign language classrooms: Instructors' training, classroom practices, and beliefs. *Language Teaching Research, 23*(6), 745–764.
- Jahara, S. F., & Abdelrady, A. H. (2021). Pronunciation Problems Encountered by EFL Learners: An Empirical Study. *Arab World English Journal, 12*(4), 194–212.
- Janssen, J., & Kirschner, P. A. (2020). Applying collaborative cognitive load theory to computer-supported collaborative learning: Towards a research agenda. *Educational Technology Research and Development, 68*(2), 783–805.
- Krashen, S. (1982). *Principles and practice in second language acquisition*.
- Kumar, Y., Koul, A., & Singh, C. (2023). A deep learning approaches in text-to-speech system: A systematic review and recent research perspective. *Multimedia Tools and Applications, 82*(10), 15171–15197.
- Lameras, P., & Arnab, S. (2021). Power to the teachers: an exploratory review on artificial intelligence in education. *Information, 13*(1), 14.
- Lindner, A., Romeike, R., Jasute, E., & Pozdniakov, S. (2019). Teachers' perspectives on artificial intelligence. *12th International Conference on Informatics in Schools, "Situation, Evaluation and Perspectives", ISSEP*.

- Losi, R. V., & Nasution, M. M. (2022). Students' attitudes toward communicative language teaching (CLT) in English speaking intensive program. *International Journal of English and Applied Linguistics (IJEAL)*, 2(1), 95–102.
- Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6(1), 1–29.
- Mirza, H. S. (2015). ESL and EFL Learners Improve Differently in Pronunciation: The Case of Lebanon. *Procedia - Social and Behavioral Sciences*, 199, 486–495. <https://doi.org/10.1016/J.SBSPRO.2015.07.536>
- Nachmani, E., & Wolf, L. (2019). Unsupervised polyglot text-to-speech. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7055–7059.
- Ng, C. H. (2020). Communicative Language Teaching (CLT) through synchronous online teaching in English language preservice teacher education. *International Journal of TESOL Studies*, 2(2), 62–73.
- Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, 29(4), 394–398.
- Pennington, M. C. (2021). Teaching pronunciation: The state of the art 2021. *RELC Journal*, 52(1), 3–21.
- Pourhosein Gilakjani, A., Namaziandost, E., & Ziafar, M. (2020). A survey study of factors influencing Iranian EFL learners' English pronunciation learning. *International Journal of Research in English Education*, 5(2), 103–123.

- Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023). Speech-to-Text and Text-to-Speech Recognition Using Deep Learning. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 657–666.
- Rodríguez-Fuentes, R. A., & Calle-Díaz, L. (2023). Fluency development beyond speech rate: A study on the effects of read aloud protocols on EFL learners. *Ampersand*, *11*, 100133. <https://doi.org/10.1016/j.amper.2023.100133>
- Sairanen, V. (2023). *Deep learning text-to-speech synthesis with Flowtron and WaveGlow*.
- 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.
- Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education*, *19*(1), 1–20.
- Sečujski, M., Nosek, T., Suzić, S., Pekar, D., Pakoci, E., Popović, B., & Delić, V. (2019). How the Advent of Machine Learning Transforms Speech Technologies. *SPEECH AND LANGUAGE 2019*, 105.
- Stroparo, T. R., & Floriani, N. (2023). SISTEMAS PARTICIPATIVOS DE CERTIFICAÇÃO: INOVAÇÃO SOCIAL, ECOINOVAÇÃO OU DESTERRITORIALIZAÇÃO? *Boletim de Conjuntura (BOCA)*, *13*(38), 325–341.

- Sun, Z., Anbarasan, M., & Praveen Kumar, D. (2021). Design of online intelligent English teaching platform based on artificial intelligence techniques. *Computational Intelligence*, 37(3), 1166–1180.
- Valentini-Botinhao, C., & Yamagishi, J. (2018). Speech enhancement of noisy and reverberant speech for text-to-speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(8), 1420–1433.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.
- Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities*, 51(1), 73–84.
- Wu, P., Yu, S., & Luo, Y. (2023). The development of teacher feedback literacy in situ: EFL writing teachers' endeavor to human-computer-AWE integral feedback innovation. *Assessing Writing*, 57, 100739.  
<https://doi.org/https://doi.org/10.1016/j.asw.2023.100739>
- Yang, C., Rassouli, A., Bhat, U. R., Endrano, A. A., Camiwet, D. Y., Shet, J. P., Hernández, R. M., & Sameem, M. A. M. (2022). Perceptions of English language teachers on the use of text-to-speech systems and the google site towards speaking competence. *Journal of Positive School Psychology*, 6(2), 1928–1937.

## APPENDIX B

Description: The following document details the Research Plan for the Thesis to be submitted for approval by the Program's Commission.

UNIVERSIDAD BOLIVARIANA DEL ECUADOR  
ANEXO II

**"PLAN DE INVESTIGACIÓN PARA LA TESIS, QUE DEBERÁ SER REVISADO Y APROBADO POR LA COMISIÓN DE TITULACIÓN DE CADA PROGRAMA AL CONCLUIR EL TALLER I"**

**TITLE OF THE THESIS:** "DEVELOPMENT OF A TEXT TO SPEECH INTERACTION PLATFORM TO IMPROVE PRONUNCIATION"

**AUTHOR 1:** Daniel Contreras Moscol

**RESEARCH CONTEXT:** EFL Students of the University of Babahoyo

**1. Brief analysis and description of the situation that justifies the presentation of this proposal.**

In recent years, the field of language education has undergone a transformative shift due to the integration of advanced technology into pedagogical practices. Among the multifaceted challenges faced by language learners, the mastery of accurate pronunciation and reading skills assumes a pivotal role in achieving fluency and effective communication. However, for non-native speakers, achieving a high level of accuracy in both reading and pronunciation remains difficult.

The emergence of text-to-speech (TTS) technology has presented a promising solution to address these challenges. TTS technology converts written text into spoken language, bridging the gap between written content and auditory input. This proposal aims to embark on the development of an interactive and dynamic TTS platform explicitly designed to propel the enhancement of reading accuracy and pronunciation proficiency among language learners.

The scholarly discourse and empirical findings in this domain have substantiated the significant impact of TTS technology on language acquisition. Al-Jarf's (2022) study underscores the potential of TTS software to elevate the decoding skills and pronunciation accuracy of EFL freshman students. Bonces (2019) has explored the implementation of TTS technology as a means to bolster fluency in second language (L2) reading. Meanwhile, Chen and Han (2020) have ventured into the development of a corpus-based online platform, leveraging TTS to facilitate pronunciation learning for learners of Mandarin.

Gottardi, Almeida, and Tumolo (2022) have offered reflective insights on the affordances of automatic speech recognition and TTS technologies in enhancing L2 pronunciation. These referenced studies collectively highlight the transformative potential of TTS technology in enhancing language learners' phonetic decoding, fluency, and pronunciation accuracy. Thus, this proposed platform seeks to synthesize and expand upon these insights to create an innovative learning tool that not only refines reading accuracy and pronunciation but also cultivates learners' self-assurance in their language learning journey.

**2. Statement of the problem**

Language acquisition, especially in a second language (L2) context, is accompanied by challenges that impede learners' attainment of accurate pronunciation and reading

---

proficiency. While technology has shown promise in aiding language learning, the effective integration of resources to enhance these specific skills remains an ongoing concern. Despite the potential of text-to-speech (TTS) technology in bridging the gap between written and spoken language, there is a scarcity of comprehensive platforms that target the improvement of reading accuracy and pronunciation skills.

The existing literature highlights the potential of TTS technology to improve aspects of language learning, but a comprehensive platform tailored explicitly to L2 learners' needs is lacking. Additionally, the intricacies of integrating TTS technology with pedagogical strategies and the assessment of its impact on learners' reading and pronunciation warrant further exploration. Therefore, a critical inquiry is needed to develop an innovative text-to-speech interaction platform that addresses these shortcomings and provides learners with an engaging tool for improving their reading accuracy and pronunciation skills. Then, the question that guides this research is: How can a text-to-speech interaction platform be designed and implemented to effectively enhance L2 learners' reading accuracy and pronunciation skills, and what are the pedagogical implications of its integration into language learning contexts?

### **3. Justification of the research**

This research aims to address a critical gap in language education by developing a text-to-speech interaction platform designed to enhance reading accuracy and pronunciation skills among second language (L2) learners. Despite the widespread recognition of the importance of these skills in language proficiency, learners often face obstacles in attaining the desired level of accuracy and fluency. The integration of technology, particularly text-to-speech (TTS) technology, presents a unique opportunity to tackle these challenges and revolutionize language learning.

The existing literature, as referenced in Al-Jarf (2022), Bonces (2019), Chen and Han (2020), and Gottardi et al. (2022), highlights the potential of TTS technology to improve decoding skills, fluency, and pronunciation accuracy. However, there is a lack of comprehensive platforms that offer an integrated approach to addressing these language learning needs. The proposed research aims to bridge this gap by creating a platform that combines the power of TTS technology with effective pedagogical strategies.

The incorporation of TTS technology in a user-friendly and interactive platform holds promise for catering to learners' diverse needs and learning preferences. Such a platform has the potential to provide learners with an immersive experience that fosters accurate reading and pronunciation through repeated exposure to spoken language. By developing a comprehensive platform, this research aims to contribute to the enhancement of both reading accuracy and pronunciation skills in a single integrated solution.

### **4. Description of the relationship between the proposal and the research lines of the university**

The selected thesis topic aligns with the research directions at the Universidad Bolivariana del Ecuador, specifically in the domain of educational methodologies and technology innovation. This proposition closely conforms to the university's dedication

---

to driving instructional techniques and improving the learning atmosphere through state-of-the-art approaches.

Furthermore, the pedagogical implications of this platform extend beyond its immediate impact on learners' language skills. The proposed research aligns with contemporary educational trends that prioritize personalized and technology-enhanced learning experiences. By investigating the integration of TTS technology with pedagogical strategies, this research addresses the broader landscape of technology-mediated education and its potential to reshape language instruction.

On the other hand, this research is led by the pressing need to address the challenges faced by L2 learners in achieving reading accuracy and pronunciation skills. By developing an innovative text-to-speech interaction platform, this study aims to contribute to the advancement of language education, offer a novel approach to addressing these challenges, and explore the pedagogical implications of technology integration in language learning.

**5. Description of the research object**

Reading and pronunciation improvement through TTS.

**6. Research aim**

To develop an innovative text-to-speech interaction platform that effectively enhances second language (L2) learners' reading accuracy and pronunciation skills, thereby contributing to their overall language proficiency and confidence.

**7. Specific research objectives**

- To design and implement a user-friendly text-to-speech interaction platform that seamlessly integrates advanced natural language processing and speech synthesis technologies.
- To explore the pedagogical implications and potential applications of the developed platform within language learning curricula.
- To evaluate the students' perception of the TTS's effectiveness in improving reading accuracy and pronunciation proficiency among L2 learners.

**8. Description of the conceptual and operational categories (independent variable & dependent variable)**

The independent variable of this study is the Text-to-Speech Interaction Platform which refers to a software tool developed to convert written text into spoken language. This innovative platform is designed to enhance the reading accuracy and pronunciation skills of second language (L2) learners. By bridging the gap between written and spoken language, the platform offers a unique and interactive learning environment. Learners engage with textual content through auditory means, facilitating the development of their reading accuracy and pronunciation abilities. The operationalization of this variable involves creating a user-friendly online platform that seamlessly integrates advanced natural language processing and speech synthesis technologies. This platform encompasses elements like user interface design, speech synthesis capabilities, synchronized text highlighting during speech, and interactive

exercises. Its primary purpose is to empower L2 learners to interact with written text in an auditory format, thereby contributing to improved language skills.

On the other hand, the dependent variable is Language Proficiency Improvement which signifies the overall enhancement of various language skills, encompassing both reading accuracy and pronunciation. This concept represents the advancement of learners' linguistic competencies, leading to improved comprehension, effective communication, and heightened confidence in the target language. The operationalization of this variable involves measuring the quantifiable enhancement in reading accuracy and pronunciation skills among L2 learners. This improvement is assessed through pre- and post-assessment scores obtained from exercises on the Text-to-Speech Interaction Platform. The term "reading accuracy improvement" encapsulates learners' ability to accurately decode and comprehend written text, while "pronunciation improvement" refers to their skill in articulating sounds, stress patterns, and intonation. The overarching goal is to evaluate the impact of the Text-to-Speech Interaction Platform on the language proficiency of learners.

#### **9. Description of the research approaches and methodologies to be used**

The research approach will primarily involve a combination of descriptive statistics and a survey-based analysis to comprehensively explore the platform's impact and learners' perceptions.

The research will employ a descriptive statistical approach to analyze the measurable outcomes resulting from the platform's utilization. Pre- and post-assessment tests will be administered to a selected group of L2 learners who engage with the Text-to-Speech Interaction Platform. These assessments will quantify improvements in reading accuracy and pronunciation skills. Descriptive statistics such as means, standard deviations, and percentage changes will be calculated to provide a clear overview of the extent of improvement achieved through the platform. This quantitative analysis will offer empirical insights into the concrete benefits of the intervention.

To gain a comprehensive understanding of students' perceptions and experiences, a survey will be administered to participants using the Text-to-Speech Interaction Platform. The survey will include questions addressing aspects such as user satisfaction, ease of use, perceived effectiveness, motivational factors, and challenges encountered. Likert scale and open-ended questions will be employed to capture both quantitative and qualitative feedback. Descriptive statistical methods will be utilized to summarize and interpret the quantitative survey responses. Moreover, thematic analysis will be conducted on qualitative responses to uncover recurring themes and sentiments expressed by the participants.

#### **10. Description of the research instruments**

Initially, it will evaluate the impact of the Text-to-Speech Interaction Platform on language proficiency, and pre- and post-assessment tests will be designed and administered. These tests will include reading accuracy exercises to assess learners' ability to decode and comprehend written text and pronunciation assessment exercises to evaluate their articulation of sounds, stress patterns, and intonation. The pre-assessment will establish a baseline, while the post-assessment will measure improvements after platform engagement. The scores from these assessments will

provide quantitative data for descriptive statistical analysis, offering insights into learners' proficiency enhancement.

A survey questionnaire will be developed to gauge students' perceptions and experiences with the Text-to-Speech Interaction Platform. The questionnaire will consist of both closed-ended and open-ended questions. The closed-ended questions will utilize Likert scale responses to quantify participants' satisfaction levels, perceived effectiveness, ease of use, and motivations for platform usage. Additionally, open-ended questions will allow participants to provide qualitative feedback, enabling them to express their opinions, challenges, and suggestions. The survey responses will provide valuable qualitative and quantitative data, which will be analyzed descriptively and thematically.

Finally, to understand the development and features of the Text-to-Speech Interaction Platform, a literature review will be conducted. This review will involve examining the platform's design principles, technical specifications, and intended learning outcomes. Analyzing the platform's documentation will provide contextual information and insights that aid in understanding its design rationale and potential impact on language proficiency improvement.

#### **11. Description of the beneficiaries of the proposal**

The proposed research holds the potential to benefit a diverse set of stakeholders, primarily encompassing 26 English teachers and approximately 6,000 students of the University of Babahoyo. Mainly, the research will directly impact the students at the University of Babahoyo, which have to study courses that go from A1 to B1 Level based on CEFR, enabling them to benefit from innovative teaching methodologies, interactive content, and tailored learning experiences facilitated by the strategic use of the TTS tool.

#### **12. Description of the research context**

This research will be conducted at the University of Babahoyo, a Higher Education Institution, located in Babahoyo, Los Ríos. This university operates from morning to evening, the whole week, where around 9000 students are educated. In the language center called CENID, there are 26 teachers that virtually present their classes daily. At this current semester, 6000 students are enrolled approximately, from A1 to B1 levels. They need to complete 5 English courses where they reach the English proficiency necessary for completing their studies and graduating.

Students are from 18 to 30 years and come from different backgrounds and places. Students are from different parts of Los Ríos province, and most of them live in rural areas. The group is highly diverse, culturally and linguistically, for that reason, CENID authorities try hard to get a supportive and collaborative learning environment, encouraging active engagement in meaningful tasks and authentic communication.

#### **13. Description of the practical contributions of the proposal, its importance, and its social need**

The proposed research holds several significant practical contributions to both the field of language education and technology integration. By developing and implementing the Text-to-Speech Interaction Platform, this study offers an innovative approach to

## APPENDIX D

Description: The following document is a letter from the candidate to the Director of the University where the Research would be conducted.

Babahoyo, 27 de agosto del 2023

Estimado Director,

**Julio Mora Aristega, MSc**

**Director del Centro de Idiomas de la Universidad Técnica de Babahoyo**

Los docentes del Centro de idiomas, Lic. Daniel Contreras Moscol con cédula de identidad 1204946147, y el Ing. Vicente Coello Vásquez con cédula de identidad 1204924391, actualmente alumnos de la Universidad Bolivariana del Ecuador, en donde actualmente cursamos la Maestría En Pedagogía del Inglés como Lengua Extranjera, nos dirigimos a usted con el fin de solicitar su autorización y apoyo para llevar a cabo una investigación en el Centro de Idiomas de la Universidad Técnica de Babahoyo. La investigación propuesta lleva por título "Development of a Text to Speech Interaction Platform to Improve Reading Accuracy and Pronunciation" y busca abordar una brecha crítica en la educación de idiomas mediante el desarrollo de una plataforma de interacción de texto a voz diseñada para mejorar la precisión en la lectura y las habilidades de pronunciación entre los estudiantes de segunda lengua (L2).

El enfoque de esta investigación radica en desarrollar una plataforma de interacción de texto a voz diseñada para mejorar la precisión en la lectura y las habilidades de pronunciación entre los estudiantes de segunda lengua (L2). Dado el contexto actual de la educación, donde la tecnología está desempeñando un papel cada vez más influyente, es crucial explorar cómo estas herramientas pueden ser aprovechadas para potenciar la pedagogía y brindar experiencias de aprendizaje más personalizadas y atractivas.

Las razones para llevar a cabo esta investigación son:

- Existe una brecha significativa en la educación de idiomas en cuanto a la precisión en la lectura y las habilidades de pronunciación entre los estudiantes de segunda lengua (L2). Esta investigación busca abordar esta brecha y mejorar la calidad de la enseñanza y el aprendizaje en este aspecto.
- En la era actual, la tecnología juega un papel crucial en la educación. La integración de una plataforma de interacción de texto a voz aprovecha las capacidades de la inteligencia artificial para ofrecer experiencias de aprendizaje más efectivas y personalizadas.
- Los educadores buscan herramientas efectivas que puedan mejorar la competencia en lectura y pronunciación de los estudiantes L2. Esta investigación se propone desarrollar una solución innovadora que se ajuste a las necesidades actuales de la educación de idiomas.
- Al empoderar a los docentes con herramientas tecnológicas avanzadas, se puede mejorar su capacidad para impartir una instrucción más efectiva y centrada en el estudiante, lo que a su vez mejora la experiencia de aprendizaje de los estudiantes.
- La mejora en la precisión en la lectura y la pronunciación tiene un impacto directo en la competencia comunicativa de los estudiantes L2. Una mejora en estas

## APPENDIX G

Description: This appendix contains the lesson plans used for the Proposal in the pre-test and post-test.

### Lesson Plan for Pre-test

**Institution:** University of Babahoyo

<b>Students:</b> Level 3 (A-2 students)		<b>Date:</b> January 15 <sup>th</sup> and 16 <sup>th</sup> , 204	
<b>Topic:</b> Reading – Focus on Pronunciation			
<b>Objective:</b> Familiarize students with the TTS-reader platform and practice pronunciation			
Stage	Interaction	Procedure or Activity	Timing
Introduction	TTS	Greet participants.	15 minutes
		Provide an overview of the platform program.	
		Explain the process of the 30-days interventions	
Pre-test Activity	TTS	Present the three texts selected for the pre-test in the first week	30 minutes
		Start the practice using the TTS reader.	
Q&A and Feedback	TTS	Invite questions and gather feedback.	15 minutes
Evaluation process	TTT	Review the process in the platform	2 hours
		Grade students' pronunciation practice according to the rubric.	

### Lesson Plan for Post-test

**Institution:** University of Babahoyo

<b>Students:</b> Level 3 (A-2 students)		<b>Date:</b> February 20th, 2024	
<b>Topic: Reading – Focus on Pronunciation</b>			
<b>Objective:</b> Measure students' improvement and advancements using the TTS reader platform during the intervention period.			
Stage	Interaction	Procedure or Activity	Timing
Introduction	TTS	Greet participants.	15 minutes
		Give general feedback on the intervention period.	
Pre-test Activity	TTS	Present the selected text for the post-test activity.	30 minutes
		Students record their pronunciation practice on the TTS reader.	
Q&A and Feedback	TTS	Invite questions and gather feedback.	15 minutes
Evaluation process	TTT	Review the process in the platform	2 hours
		Grade students' pronunciation practice according to the rubric.	

## APPENDIX H

Description: The chart below contains the schedule based on a Gantt Diagram with the Research Proposal activities.

No	Activity	Se p	Oc t	No v	De c	Ja n	Fe b	Ma r
1	Development of Literature Review							
2	Select variables to study according to Lit Rev							
3	Research secondary information to select the best instruments							
4	Write Introduction and Chapter 1: Literature Review							
5	Start creating the instruments used for the research							
6	Write Chapter 2: Methodology initial steps							
7	Validate instruments and create an initial diagnosis							
8	Writing methodology (analysis and data presentation)							
9	Starting pre-test activity							
10	Working on the 30-day intervention period							
11	Collecting post-test data							
12	Analyzing Data Using Descriptive Analysis							
13	Writing Proposal and Results							
17	Writing Discussion and Conclusions							
18	Feedback and modifications							



## APPENDIX I

Description: The following document contains the letter, operationalization matrix, and validation dossier. This was sent to an expert in the field to validate the tools that were going to be used in the Research Proposal.

**INSTRUMENTS CONTENT VALIDATION DOSSIER THROUGH  
EXPERT'S JUDGEMENTS**

## APPENDIX J

Description: The chart below contains the Operationalization Matrix of Variables

Variable	Conceptual Definition	Operational Definition	Dimensions	Indicators	Measurement Scale
Independent Variable: Text-to-Speech Interaction Platform	A software-based tool that converts written text into spoken language, designed to enhance pronunciation accuracy among L2 learners.	A user-friendly online platform incorporating advanced natural language processing and speech synthesis technologies.	Text-to-Speech Interaction Platform usage	- Interface intuitiveness - Quality of synthesized speech - Alignment of text highlighting with speech - Effectiveness of interactive exercises	Ordinal/Interval/Ratio (depending on specific indicators)
Dependent Variable: Language Proficiency Improvement	The overall enhancement of language skills, encompassing pronunciation accuracy.	The quantifiable improvement in pronunciation accuracy was measured by pre-and post-assessment scores on platform exercises.	Reading Accuracy	- Percentage of correctly pronounced words - Accuracy in sentence comprehension exercises	Ratio/Percentage
			Pronunciation Skills	- Percentage of accurately pronounced sounds - Evaluation of intonation patterns - Correctness of stress patterns	Ratio/Percentage

			<p>User Experience and Perceptions</p> <ul style="list-style-type: none"> <li>- Likert scale rating of satisfaction</li> <li>- Task completion time and ease of navigation</li> <li>- Likert scale rating of perceived effectiveness</li> </ul>	Ordinal
			<p>Pedagogical Integration</p> <ul style="list-style-type: none"> <li>- Frequency of platform usage in lessons</li> <li>- Customization of platform exercises to cater to individual needs</li> </ul>	Ordinal

## APPENDIX K

Description: This Appendix contains the Pre-test activity used with students and its rubric for evaluation.

### **Text 1: A Future of Technological Harmony**

In this version of the future, technology is completely integrated into our lives, improving our capabilities and enriching our experiences. Artificial intelligence has evolved into a trusted partner, assisting us in various aspects of our daily routines, from managing our homes to optimizing our work processes. Education has undergone a transformation, with personalized learning platforms adapting to individual needs and supporting a lifelong love of learning. Healthcare has advanced in many ways, diagnostic tools and personalized treatments enabling us to live longer, healthier lives. The world is more interconnected than ever before, with virtual communication bridging physical distances and supporting global collaboration.

### **Text 2: A Future of Environmental Harmony**

In this version of the future, humanity has reconciled its relationship with nature, recognizing the interconnectedness of all living things. Renewable energy sources power our societies, leaving a minimal carbon footprint. Sustainable practices govern agriculture, ensuring food security for a growing population. Sustainable urban planning has created green cities that coexist harmoniously with the natural environment. Conservation efforts have restored biodiversity, preserving the delicate balance of ecosystems. Awareness of environmental issues has permeated all aspects of society, driving individuals and organizations to make conscious choices that protect our planet.

### **Text 3: A Future of Human Harmony**

In this version of the future, humanity has overcome its divisions, embracing diversity and supporting a global community. Education has emphasized empathy, understanding, and

conflict resolution, cultivating a generation of peacemakers. Cultural exchange has become commonplace, breaking down barriers and supporting mutual respect. Global cooperation has addressed shared challenges, from climate change to poverty, ensuring a more equitable and sustainable future for all. Differences of opinion are embraced as opportunities for growth, leading to more informed and inclusive decision-making processes. Technology has facilitated cross-cultural communication and collaboration, bridging the gaps that once separated people.

### **Extra – Focus On Accuracy and understanding of linguistics**

Good pronunciation is an essential part of effective communication. When we speak clearly and accurately, we are more likely to be understood by others. This can lead to better relationships, improved job prospects, and a greater sense of confidence. Additionally, good pronunciation can help us to learn new languages more effectively. Practice pronunciation with the following words:

Though	Thought	February	weather	Thoughtful
Routine	Although	Library	Secretary	Thorough
Neither	Thoughtful	Necessary		Weather
Schedule	Thorough	Probably	Should	
Through	Thoroughfare	Recipe	Through	

**Rubric:**

Criteria	2	1.5	1	0.5
Pronunciation	Pronunciation is clear and accurate.	Pronunciation is mostly clear and accurate with minor errors.	Pronunciation is sometimes unclear or inaccurate.	Pronunciation is unclear and inaccurate.
Reading Accuracy	Reading is fluent and accurate.	Reading is mostly fluent and accurate with minor errors.	Reading is somewhat fluent and accurate with noticeable errors.	Reading is choppy and inaccurate.
Expression	Expression enhances understanding of the text.	Expression mostly enhances understanding of the text with minor inconsistencies.	Expression is sometimes flat or inconsistent.	Expression is flat and lacking in emotion.
Decoding and Word Recognition	Demonstrates excellent decoding and word recognition skills.	Demonstrates proficient decoding and word recognition skills with occasional errors.	Demonstrates partial decoding and word recognition skills with noticeable errors.	Demonstrates limited decoding and word recognition skills.
Comprehension	Demonstrates full comprehension of the text.	Demonstrates partial comprehension with some gaps in understanding.	Demonstrates limited comprehension with significant gaps in understanding.	Demonstrates minimal comprehension of the text.
Overall Presentation and Fluency	Presents the text with confidence and fluency.	Presents the text with mostly confident and fluent delivery with minor hesitations.	Presents the text with occasional hesitations and interruptions.	Presents the text with hesitations and lack of fluency.

## APPENDIX L

Description: This Appendix contains the Post-test activity used with students and its rubric for evaluation.

### **Topic: The Future of Technology**

In the future, technology will change our daily routines in ways we can only imagine. Though we might not see flying cars everywhere, there will be many advancements that will make our lives easier. Neither long commutes nor traffic jams will be an issue as self-driving cars will take us through thoroughfares quickly and safely.

Our schedules will be more flexible with the help of advanced artificial intelligence. A thoughtful AI secretary could help us organize our tasks and remind us of necessary appointments. It might even predict the weather accurately, helping us plan our days better. In February, when the weather is unpredictable, our smart homes will automatically adjust the temperature and lighting to keep us comfortable. Although it may seem like a distant dream, these innovations are already in development.

Libraries will probably have more digital collections, allowing us to access books from anywhere. This will be especially useful for students who need to research and study. Recipes for cooking will also be easily accessible, and our kitchen appliances will guide us through each step thoroughly.

We should remember that with all these advancements, it is important to use technology thoughtfully. While it can improve our lives, we must also be aware of its impacts. By being thorough in our understanding and use of technology, we can ensure a better future for everyone.

**Rubric:**

Criteria	2	1.5	1	0.5
Pronunciation	Pronunciation is clear and accurate.	Pronunciation is mostly clear and accurate with minor errors.	Pronunciation is sometimes unclear or inaccurate.	Pronunciation is unclear and inaccurate.
Reading Accuracy	Reading is fluent and accurate.	Reading is mostly fluent and accurate with minor errors.	Reading is somewhat fluent and accurate with noticeable errors.	Reading is choppy and inaccurate.
Expression	Expression enhances understanding of the text.	Expression mostly enhances understanding of the text with minor inconsistencies.	Expression is sometimes flat or inconsistent.	Expression is flat and lacking in emotion.
Decoding and Word Recognition	Demonstrates excellent decoding and word recognition skills.	Demonstrates proficient decoding and word recognition skills with occasional errors.	Demonstrates partial decoding and word recognition skills with noticeable errors.	Demonstrates limited decoding and word recognition skills.
Comprehension	Demonstrates full comprehension of the text.	Demonstrates partial comprehension with some gaps in understanding.	Demonstrates limited comprehension with significant gaps in understanding.	Demonstrates minimal comprehension of the text.
Overall Presentation and Fluency	Presents the text with confidence and fluency.	Presents the text with mostly confident and fluent delivery with minor hesitations.	Presents the text with occasional hesitations and interruptions.	Presents the text with hesitations and lack of fluency.

## APPENDIX M

Description: This Appendix presents a semi-structured questionnaire to evaluate students' perceptions and satisfaction.

### **Informed Consent**

This questionnaire is intended for English students of the Technical University of Babahoyo and it aims to analyze the perceptions and usage of Text-to-speech reading platforms and its practice. All information is confidential and will be used for academic purposes. The estimated time for completing the survey is 15 minutes. Do you agree to proceed with the survey?"

Yes                       No

### **Section I: General Information**

1. Age: \_\_\_\_\_
2. Gender:
 

Male                       Female                       Other
3. City where you live: \_\_\_\_\_
4. Career: \_\_\_\_\_
5. Semester: \_\_\_\_\_

### **Section 2: Using TTS reader**

6. How were your initial impressions of the TTS reader platform and the read aloud protocols?
 

Positive                       Negative
7. Did your first impressions change over the course of the intervention?
 

Yes                       No
8. If so, in what ways? \_\_\_\_\_

9. Please rate the usage of the TTS reader in your English classes. Consider that 4 is the highest level, while 1 is the lowest.

<b>AI Tool</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
1. The platform (TTS Reader)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The time used to complete the activity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The internet usage during the activity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The willingness to perform the tasks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### **Section 3: User Experience and Perceptions**

10. Did you experience any anxiety or nervousness before or during the read aloud protocols?

Yes                       No

11. Reflecting on the 4-week period, how would you rate your perceived improvement in English fluency?

Much better                       Somehow better       Same                     

Worst

12. What specific aspects of fluency do you believe have improved the most?

Pronunciation                       Intonation                       Phrasing

Word recognition                       Comprehension                       Sentence Fluency

13. Did you find the protocols compatible with your study routine?

Yes                                       No

14. Were there any specific challenges or strengths you have? Describe

\_\_\_\_\_

15. Please rate your level of agreement with the following statements regarding the integration of AI-powered tools in virtual English classes. Consider 4: strongly agree, 3: agree, 2: disagree, 1: strongly disagree.

<b>Statement</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
1. The user interface of the TTS Interaction Platform is intuitive and easy to navigate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The quality of the synthesized speech in the platform is clear and natural.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The alignment of text highlighting with speech enhances my ability to follow and comprehend the text.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The interactive exercises on the platform are engaging and contribute to my learning experience.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. The TTS Interaction Platform has been effectively integrated into my language learning lessons.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

16. On a scale from 1 to 5, how satisfied are you with your overall experience with the read aloud protocols?

<b>Statement</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
My satisfaction level is:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

17. What factors contributed most to your satisfaction or dissatisfaction?

---