TECHNOLOGY-ASSISTED TRAINING IN L2 SPANISH PRONUNCIATION: EVALUATION FROM PHONETIC AND PSYCHOMETRIC PERSPECTIVES

by

STACEY MICHELLE JACOBSON

ERIN E. O'ROURKE, COMMITTEE CHAIR
ALICIA B. CIPRIA
BRYAN J. KORONKIEWICZ
DOUGLAS J. LIGHTFOOT
DARRIN J. GRIFFIN

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Modern Languages
and Classics in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2021

ABSTRACT

Pronunciation is often not given the attention it deserves in the foreign language classroom, especially as research continues to show that poor pronunciation can impede communication (Agostinelli, 2012; Arteaga, 2000). Recent attempts have been made to determine the utility of technology for the purposes of acquiring another language; specifically, speech recognition technology, along with the feedback it can provide, has been found to assist with pronunciation acquisition (Golonka et al., 2014).

This study investigated the effects of technology-enhanced Spanish pronunciation training and its potential to improve beginning L2 learners' accuracy in pronouncing grapheme-phoneme mismatches as well as a decrease in voice onset time (VOT) duration when producing the voiceless stops /p t k/. Over the course of a semester, the pronunciation of learners using an app with speech recognition functionality was compared to that of learners interacting with pronunciation activities in the online companion to the course textbook, and to learners who completed grammar activities with no pronunciation component. Overall, both the grapheme-phoneme mismatch accuracy and VOT duration showed improvement with use of technologies and over time.

Beyond the improvement in pronunciation, this study investigated the participants' opinions regarding the technology they interacted with. The software Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, et al., 2015) was utilized to conduct psychometric analyses of participants' thoughts and emotions regarding their assigned technology from the language used in the questionnaire responses. Despite participants' generally positive views of

their assigned technology, the findings also supported previous claims that enjoyment does not necessarily translate into improvement (Youngs et al., 2011), and suggest that too much focus on the technology itself may hinder pronunciation improvement.

DEDICATION

To my Grandpa Frank, whose pronunciation of *Via de la Valle*
inspired a Master's thesis and now a dissertation

To Mary, who pointed out that the word *echo*
is inside the word *techo*

To my parents, Alan and Gigi, for never giving up
on trying to figure out exactly what it is that I research,
and for supporting me no matter how many times I go back to school

And to my sister, Amy, for teaching me math over the phone
so I could take programming classes,
without which I might still be analyzing data

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

"My fingers," said Elizabeth, "do not move over this instrument in the masterly manner which I see so many women's do. They have not the same force or rapidity, and do not produce the same expression. But then I have always supposed it to be my own fault – because I would not take the trouble of practising. It is not that I do not believe *my* fingers as capable as any other woman's of superior execution."

~ Elizabeth Bennet, *Pride and Prejudice*, Chapter XXXI

"If I had ever learnt, I should have been a great proficient."

~ Lady Catherine de Bourgh, *Pride and Prejudice*, Chapter XXXI

LIST OF ABBREVIATIONS AND SYMBOLS

LINGUISTIC ABBREVIATIONS

1	First person	PTCP	Participle
2	Second person	SBJV	Subjunctive mood
3	Third person	SG	Singular
adj.	Adjective	*	Unacceptable
COND	Conditional tense	[ ]	Allophone
F	Feminine	< >	Grapheme
IMP	Imperative	{ }	Group, subset
IMPER	Imperfect	/ /	Phoneme
M	Masculine	‘ ’	Translation
n.	Noun	ˈ	Primary stress
PL	Plural	ˌ	Secondary stress
PRET	Preterit tense	ʰ	Aspiration
PRS	Present tense	̩	Syllabic consonant

PSYCHOMETRIC ABBREVIATIONS

affect	Category containing words pertaining to affective processes
anger	Category containing anger-related words
anx	Category containing anxiety-related words
focus future	Includes modals for future tense, words that reference the future, e.g., *hope*
focus past	Includes past tense conjugations, words that reference the past, e.g., *remember, already*
focus present	Includes present tense conjugations, words that reference the present, e.g., *today*
i / I	Category containing first person singular referents
ipron	Category containing impersonal pronouns
negemo	Category containing negative emotion words
posemo	Category containing positive emotion words
ppron	Category containing personal pronouns
pron	Category containing pronouns
sad	Category containing sadness-related words
tentat	Category containing tentative language
they	Category containing third person plural referents
we	Category containing first person plural referents
you	Category containing second person singular referents

STATISTICAL ABBREVIATIONS

B	Estimate
CI	Confidence Interval
df	Degrees of freedom
Exp(B)	Odds ratio(s) for the predictor(s) in regression analysis
F	F statistic
H	Kruskal-Wallis H test statistic
Int.	Intercept
M	Mean
ms	*Milliseconds*
N	Total
nsd; *(**)	No significant difference; Levels of statistical significance
p	Probability, denotes level of significance
R	Statistical software
R^2	McFadden's Pseudo R^2; calculates effect size
SD	Standard deviation
SE	Standard error
Sig.	Significance
t	t-test statistic
X^2	Chi-Square test statistic

PHONETIC SYMBOLS – CONSONANTS

Stops

p	Voiceless bilabial stop
t	Voiceless dental stop (Spanish) Voiceless alveolar stop (English)
c	Voiceless palatal stop
k	Voiceless velar stop
b	Voiced bilabial stop
d	Voiced dental stop (Spanish) Voiced alveolar stop (English)
ɟ	Voiced palatal stop
g	Voiced velar stop

Fricatives

θ	Voiceless interdental fricative
s	Voiceless alveolar fricative
z	Voiced alveolar fricative
ʃ	Voiceless alveopalatal fricative
ʒ	Voiced pre-palatal fricative
ʝ	Voiced palatal fricative
x	Voiceless velar fricative
h	Voiceless glottal fricative

Affricates

t͡s	Voiceless alveolar affricate
t͡ʃ	Voiceless pre-palatal affricate
d͡ʒ	Voiced pre-palatal affricate

Rhotics, Laterals, and Approximants

β	Voiceless bilabial approximant
ð	Voiced interdental approximant
ɾ	Voiced alveolar tap (Spanish) Voiced alveolar flap (English)
r	Voiced alveolar trill
ɹ	Voiced alveolar approximant
l	Voiced alveolar lateral approximant
ɫ	Voiced velarized alveolar lateral approximant
ʎ	Voiced palatal lateral
j	Voiced palatal approximant
w	Voiced labio-velar approximant
ɣ	Voiced velar approximant

Nasals

m	Voiced bilabial nasal
ɱ	Voiced labiodental nasal
n	Voiced alveolar nasal
ɲ	Voiced palatal nasal
ŋ	Voiced velar nasal

PHONETIC SYMBOLS – VOWELS

i	High front unround tense vowel
ɪ	High front unround lax vowel
e	Mid front unround tense vowel
eɪ	Diphthong beginning with mid front unround tense vowel, ending with high front unround lax vowel
ɛ	Mid front unround lax vowel
ə	Mid central unround lax vowel (unstressed, English)
ɚ	Rhotacized mid central unround lax vowel
ʌ	Mid central unround lax vowel (stressed, English)
æ	Low front unround tense vowel
a	Low central unround tense vowel (Spanish)
ɑ	Low back unround tense vowel (English)
ɑɪ	Diphthong beginning with low back unround tense vowel, ending with high front unround lax vowel
o	Mid back round tense vowel
oʊ	Diphthong beginning with mid back round tense vowel, ending with high back round lax vowel
ʊ	High back round lax vowel
u	High back round tense vowel

ACRONYMS

ANOVA	Analysis of Variance
ASR	Automatic Speech Recognition
CAH	Contrastive Analysis Hypothesis
CALL	Computer Assisted Language Learning
CAPT	Computer Assisted Pronunciation Training
CEFR	Common European Framework of Reference
CG	Control Group
CLT	Communicative Language Teaching approach
CPH	Critical Period Hypothesis
CSV	Comma Separated Values
EFL	English as a Foreign Language
EG1	Experimental Group 1 (Spanish SOLO app)
EG2	Experimental Group 2 (online textbook-companion activities)
ELL	English Language Learner
ESL	English as a Second Language
L1 / L2	First language / Second language
LIWC	Linguistic Inquiry and Word Count (psychometric text analysis software)
LMS	Learning Management System
MALL	Mobile Assisted Language Learning
MMORPG	Massively Multiplayer Online Role-Playing Game

MOOC	Massive Online Open Courses
NLP	Natural Language Processing
OPI	Oral Proficiency Interview
SLA	Second Language Acquisition
SLM	Speech Learning Model
TBLT	Task-Based Language Teaching
TOEFL	Test of English as a Foreign Language
TTS	Text to Speech
VHL	Vista Higher Learning (publisher); name used for textbook-companion website
VOT	Voice Onset Time
WI, WM	Word-Initial, Word-Medial
ZPD	Zone of Proximal Development

DEFINITIONS

Allophone: the spoken representation of a phoneme; it is possible for one phoneme to have more than one allophone (e.g., [ð] and [d] are allophones of /d/ in Spanish), and these allophones are mutually exclusive in terms of where they occur

Contrastive: two phonemes that occur in the same environment, e.g., in the minimal pair *he* [ˈhi] and *she* [ˈʃi] which differ only by one sound in one position

Digraph: the combination of two graphemes (e.g., <c> and <h> become the digraph <ch>) which functions as one grapheme when corresponding to a phoneme

Grapheme: the orthographic (written) representation of a speech sound

Language interference: the negative impact of one language on the perception, production, acquisition, etc. of another

Opaque / Opacity: describes languages with higher numbers of consonants in which more than one phoneme maps to a grapheme; antonyms: transparent/transparency

Phoneme: a unit of sound, designated by / /, that has meaning and distinguishes one word from another (e.g., /p/ and /b/ in English as they occur in the same environment, as in pat [ˈp^{h}æt] and bat [ˈbæt])

Spectrogram: a visual representation of sound frequencies, used for acoustic analysis

Transparent / Transparency: describes languages with higher numbers of consonants that have one-to-one correspondence with their graphemes; antonyms: opaque/opacity

ACKNOWLEDGMENTS

There are so many people without whom this dissertation might never have been written. In addition to my parents and sister, to whom this dissertation is dedicated, I want to thank my committee chair, Dr. Erin O'Rourke, for literally everything. You were my rock throughout and also as we figured out how I would graduate in the middle of a pandemic. I would also like to take this opportunity to thank my committee: Thank you to Dr. Bryan Koronkiewicz for making it possible for me to have participants, and for pushing me to examine my own pedagogical methodology. Thank you to Dr. Alicia Cipria for challenging me to consider other viewpoints, one of which became a main theme of my dissertation. To Dr. Doug Lightfoot: I'm glad I got the opportunity to work with you again, and thank you for your suggestion of the native-like pronunciation debate. And to Dr. Darrin Griffin: your enthusiasm is contagious! I enjoyed working with you and learning about LIWC, and I still hope to work with grounded theory one day.

I would especially like to thank Dr. Paul Reed for his willingness to advise me on which statistical tests best fit my project and data. Also, a big thank you to the Sanford Media Center and its employees, in particular Scott Barnes and Shane Miesse, for allowing me to run my study in the sound booth. Finally, I want to thank the instructors who graciously worked with me during the Fall 2018 semester: Stephanie B., Ivano, Mary, Jenna, Gerardo, Sarah, Stephanie S., and Andrew, as well as Samantha, Jana, Caleb, Elizabeth, and Catanya who supported me and kept me sane.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The pronunciation of consonants has the capacity to impact L2 communication. Pronunciation is an often-overlooked element of language acquisition, particularly in controlled second language acquisition settings that are governed by pedagogical methodologies and due to class time constraints, among others. The present study compares the (now defunct) Spanish SOLO app, which had automatic speech recognition (henceforth ASR) functionality, to pronunciation practice activities already available in order to determine if practicing with the app (and ASR) can help improve pronunciation accuracy pertaining to grapheme-phoneme mismatches (i.e., consonants that have more than one correspondence between their written and spoken forms).

Depending on the language in question, the written form of words (i.e., orthographic representation) can assist or hinder the pronunciation of unfamiliar words. While native speakers do encounter unfamiliar words in their own languages and need to figure out how to pronounce them, this situation is much more familiar to the L2 learner. Unlike L1 acquisition, in L2 acquisition there is a greater likelihood that learners will encounter the written and spoken forms of the language simultaneously (Bassetti, 2008), creating the need to associate graphemes (i.e., letters) with phonemes (i.e., mental representations of speech sounds). However, the number of phonemes that are associated with a given grapheme varies by language; some languages have more of a one-to-one correspondence, like Spanish, while others can have multiple phonemes associated with the same grapheme, like English.

It is likely not difficult to come up with an example of a word that took a lot of thought to figure out how to pronounce, whether in the L1 or a language learned later on. However, the decisions made pertaining to how to pronounce individual letters probably went unnoticed. Consider, then, Figures 1 and 2, which map out a process that is typically subconscious. In this particular example, the learner's L1 is English and the L2 is Spanish.

Figure 1[1]

Decision Map for hecho *'done' [ˈe.t͡ʃo]*

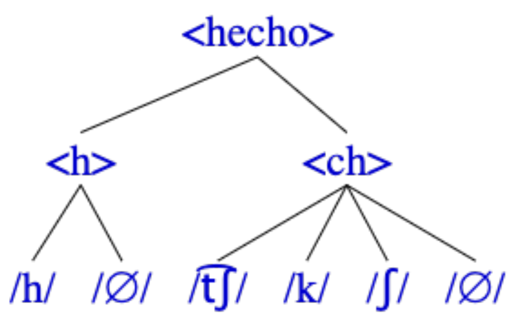
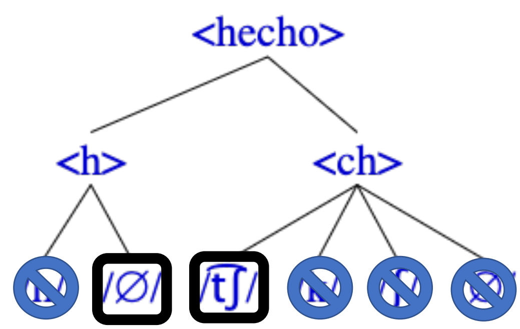


Figure 2

Target Pronunciation for Figure 1



Considering just the consonants in the word *hecho*, grapheme-phoneme associations in English provide the L2 Spanish learner with two options for the pronunciation of <h>, and as many as four options for <ch>. In this instance, at least, the desired output (indicated in Figure 2 by the black boxes) is already familiar to native English speakers, as words like *hour* have a written <h> that is not pronounced, and <ch> in Spanish is pronounced like the <ch> in *choose*.

However, other graphemes, and specifically ones whose phonemes do not exist in one of the two languages in question, create even more potential for confusion, likely leading to mispronunciation and potentially, problems with intelligibility that negatively impact

[1] Decision maps were created using jsSyntaxTree (Eisenbach & Eisenbach, 2020).

communication. Figures 3 and 4 illustrate the same subconscious process, but with more complicated graphemes for an L1 English speaker learning Spanish.

Figure 3

Decision Map for ajedrez *'chess' [a.xe.ˈðres]*

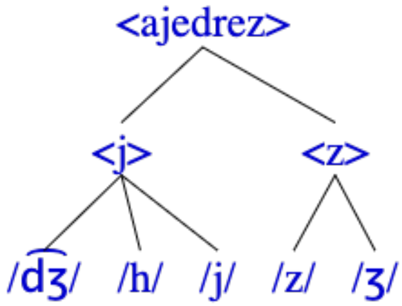
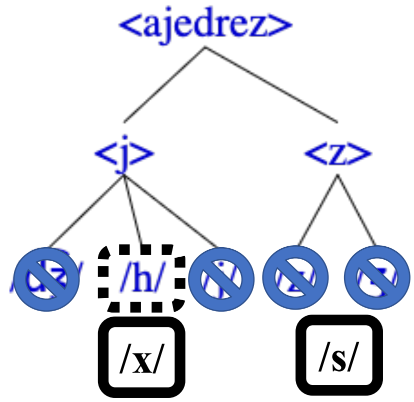


Figure 4

Target Pronunciation for Figure 3



Figures 3 and 4 illustrate the possible pronunciations for <j> and <z> based on their pronunciation in English (the L1 in this example). <d> and <r> can be set aside, as they have quite similar pronunciations in English and Spanish, but <j> and <z> illustrate potential pitfalls. As was the case in the previous example, both graphemes are associated with more than one phoneme in English: <j> occurs as /d͡ʒ/ in *just*, as /h/ in *jalapeño*, and as /j/ in *hallelujah*; <z> occurs as /z/ in *zoo* and as /ʒ/ in *azure*. However, Figure 4 presents problems with the use of English grapheme-phoneme associations as a tool to decode Spanish pronunciations. First, the canonical Spanish pronunciation of <j> (/x/) is a sound that does not occur in English. Second, while /s/ occurs in words like *summer* in English, it is rarely associated with the letter <z>. Now, the learner has to form a new association for <z> as well as figure out how to best approximate the sound /x/ in response to <j>, while avoiding the temptation to use the higher-frequency pronunciation of <j> (/d͡ʒ/ as opposed to /h/ or /j/), exemplified in words like *jump* and *just*.

As this example and its decision maps illustrate, it can be difficult for language learners to learn to associate graphemes with their phonemes in another language, and has the potential to create unintelligible words or unintentionally form a completely different word. Considering *jump* and *just*, a native Spanish speaker who is beginning to learn English might pronounce these words as something approximating *hump* (which is an English word) and **hust* (which is not)[2]. Returning to the L2 learner of Spanish, pronouncing *llamo* 'I call' (where <ll> is pronounced similarly to the English pronunciation associated with <y>) with an [l] sound (as in *lick*) creates *lamo* 'I lick' which is a Spanish word, but one that is far from the intended word.

Errors such as those illustrated previously exemplify the gap filled by the present study: L2 communicability can be affected by the pronunciation of consonants, especially when the L1 and L2 differ in terms of grapheme-phoneme associations (i.e., orthographic transparency[3] or opacity[4]). Despite the potential impact on communicability, however, for many reasons, including the sheer amount of material to cover, and the emphasis on meaning over form in the current pedagogical trend – the communicative approach (e.g., Lee & VanPatten, 2003) – pronunciation typically is not given the attention it deserves (Agostinelli, 2012; Darcy et al., 2012; MacDonald, 2002; Miller, 2012; Morin, 2007). However, this lack of attention is counterintuitive if the purpose of the communicative approach is indeed communication. If students cannot pronounce words well enough to convey which word they are using, communication may be impeded if not lost altogether (Agostinelli, 2012; Anthony, 1996; Arteaga, 2000; Gleason, 2012; Gordon, 2012; Llanes et al., 2017; MacDonald, 2002; McCrocklin, 2012; Morin, 2007; Schairer, 1992; Zielinski, 2012).

[2] Or as *yump and *yust
[3] (Carreiras et al., 2014); Hualde, 2014); Meschyan & Hernandez, 2005)
[4] (Ellis & Hooper, 2002); Timmer & Schiller, 2012)

As technology has become more accessible, more attempts have been made to utilize it to assist (or augment) the process of acquiring another language (within or outside of a classroom). With the invention of and dramatic rise in the use of smartphones (and other portable technology), mobile applications (apps) have become more popular and are available in more concise forms and without the need for a gaming system or a computer. Research has shown that existing technology has been frequently repurposed for language learning uses, and that students rank mobile apps as highly useful for language learning (Steel, 2012). Despite the draw, technology should not be incorporated into the classroom and/or the L2 acquisition process simply for the sake of using the latest technological advancements, but rather it should be supported by language learning theory as well as methodologically and linguistically sound principles (Martín Monje et al., 2014; Youngs et al., 2011). For example, the review of hundreds of studies focusing on the use of technology in L2 classrooms (and in the L2 acquisition process generally) conducted by Golonka et al. (2014) revealed few instances where the use of technology had a meaningful impact. However, Golonka et al. (2014) did find that technology did have an effect on pronunciation, mentioning ASR as a feature that improved pronunciation acquisition, specifically citing feedback as a highly useful and beneficial feature.

The feature that is most relevant to the present study is ASR. Apps such as Duolingo, Babbel, and Rosetta Stone have attempted to incorporate this technology to support the process of acquiring L2 pronunciation. Research has shown that ASR is a feature learners would like to have added to existing technology (Culbertson et al., 2016), as the potential for providing feedback is typically favored by learners (McCrocklin, 2016) and has been shown to positively impact pronunciation acquisition (García et al., 2018; Golonka et al., 2014; McCrocklin, 2016; Morgan & García, 2016; Olson, 2014; Zhao et al., 2013). The present study aims to add to the research on the utility of technology for pronunciation acquisition, as the app allowed the use of ASR in the majority of activities, providing either positive (i.e., a pleasant-sounding rising tone with a green check mark) or negative (i.e., a sharp, abrupt lower tone with a red 'x') feedback. In contrast, currently available pronunciation practice activities are often of the listen-and-repeat style, requiring learners to compare their own recorded productions to recordings of native speakers. Those utilized in the present study were included as activities within the online companion website to the textbook used in first-semester Spanish courses.

Beyond the investigation and analysis of pronunciation itself, this study also investigates attitudes toward and opinions of pronunciation-related instructional technologies. Regardless of how effective a technology may theoretically be, if the application of it does not resonate with the target audience (students, typically, in the case of language learning), it will most likely have some type of negative, or at the very least, ambivalent, connotation associated with it. In order to discover learners' emotions, feelings, and perspectives regarding technology used for pronunciation practice/training, the present study utilizes psychometric analysis to uncover the psychological impact behind the sentiments expressed (Pennebaker, 2013).

1.1 Research Questions

The present study aims to investigate the efficacy of speech recognition technology in improving pronunciation of problematic consonants in beginning learners of Spanish. Specifically, this study aims to address the following questions:

1. To what extent does the use of speech recognition software when practicing pronunciation with a language learning app impact beginning students' improvement of problematic areas of pronunciation in the L2?
    a. In which of the problematic pronunciation areas under investigation can significant improvement be found?
    b. Does training using speech recognition (and its feedback complement) produce better results than using already available technology?

2. What are students' attitudes toward and/or perceptions of currently available technology for foreign language pronunciation practice?
3. Is there a connection between impressions of/attitudes toward using the app and improvement in pronunciation?

The following chapters aim to address these questions. First, Chapter 2 presents a review of relevant literature, including (but not limited to) topics such as second language acquisition, adult acquisition of L2 phonology, computer-assisted pronunciation training, articulatory and acoustic phonetics, and research pertaining to the function and use of the psychometric software, Linguistic Inquiry and Word Count (LIWC). Next, Chapter 3 describes the methodology followed in conducting the present study, including descriptions of stimuli, instruments, participants and procedures as well as information regarding data extraction and analysis. Then, Chapter 4 explores and discusses the articulatory and acoustic phonetics findings, and Chapter 5 presents a quantitative interpretation of the participants' questionnaire responses using psychometric analysis, accompanied by quantitative analyses of word frequency trends. Additionally, a qualitative analysis of the topics frequently mentioned in the questionnaire responses is also included, as well as the discussion of both analyses. Finally, Chapter 6 addresses the research questions through summaries of the study's findings. This is followed by an exploration of contributions to various fields, the limitations of the present study, considerations for future study, and recommendations for future research.

LITERATURE REVIEW

This section explores the articulatory and acoustic features of English and Spanish. Although the sounds and graphemes pertinent to the current study correspond to Spanish, it is important to understand the features of the language learners' first language (L1). English is presented first as, in the present study, it is the participants' L1. Thus, if they were to rely on intuition for pronunciation, it would come from English. Discussing English first provides a frame of reference from which the consonantal variation that can affect the acquisition of the second language (L2) Spanish phonetics can be more clearly understood.

This chapter begins with an exploration of second language acquisition in terms of its relationship with pronunciation. Then, adult L2 acquisition of phonology is presented, including a discussion of L1 interference and contrastive analysis. Next, Technology-Assisted Language Learning is explored, with a specific focus on Computer-Assisted Pronunciation Training (CAPT). The subsequent two sections explore English and Spanish in terms of articulatory and acoustic phonetics, respectively. Finally, the research pertaining to the psychometric software Linguistic Inquiry and Word Count (LIWC) is presented, followed by applications to the present study.

2.1 Second Language Acquisition and its Relationship with Pronunciation

To understand the current state of foreign language instruction, it is important to understand the history of the field, as most second language acquisition (SLA) theories have at least some instructional element to them. Moreover, this also could provide insight into the status of pronunciation in modern foreign language instruction. However, the acquisition of pronunciation in a second language (L2) seems to be rather mysterious, as it can be difficult to obtain a clear picture of the manner in which pronunciation has been addressed throughout the evolution of SLA theory. Nor does the situation become much clearer upon consideration of the various models and approaches to language instruction, as communication is often the focus. While many of these theories discuss speech, pronunciation may not be the intended target, as speech is often mentioned in reference to fluency (Lee & VanPatten, 2003; Skehan, 1998).

As its name suggests, the Audiolingual Approach heavily emphasized the spoken form of the language, and is historically one of the few approaches to do so. According to Lightbown and Spada (2013), typical activities that corresponded to this approach included memorizing dialogues and sentence patterns, as well as mimicking the pronunciation of the teacher, who often focused on specific sounds that occur in minimal pairs and are often perceptually difficult for language learners, such as the vowels in *feet* [fit] and *fit* [fɪt]. Additionally, drill-style activities were frequent and commonplace, emphasizing the target structure during the various iterations of the sentence or dialogue while using a variety of vocabulary words (Cook, 2001, pp. 206-207). This approach was also greatly concerned with habit formation, as its proponents were concerned that learners would form and fossilize bad habits (Lightbown & Spada, 2013). Thus, while students were speaking (and thus pronouncing) the language, "no free use of language was allowed because it was believed that it would cause errors, which would interfere with the formation of correct habits in the foreign language" (de Bot et al., 2005, p. 78). Overall, while this method did emphasize oral production, the lack of utterance freedom combined with the fear of poor habit formation hindered its effectiveness.

Subsequent approaches appeared to deemphasize pronunciation, or at least make no overt mention of it, though some promoted ideas that theoretically could be applied to pronunciation.

For example, behaviorism introduced the ideas of transfer and contrastive analysis, as it was generally thought that L2 errors were caused by interference from the L1 (i.e., transfer). However, knowledge of the L1 would allow for the prediction of these errors, as they are caused by differences between the L1 and L2, indicating which parts of the language needed to be actively learned (i.e., contrastive analysis) (Gass & Selinker, 2008; Lightbown & Spada, 2013). Though not specifically discussed in conjunction with pronunciation, both of these ideas could easily be applied to it, as the L1 and L2 phonological inventories may overlap but not be identical, creating a situation in which contrasts could be highlighted and therefore learned. Gass and Selinker (2008) note that only in the North American interpretation is contrastive analysis applied to language learning and teaching; the goal of the European interpretation was to further the understanding of the nature of language itself (pp. 95–96).

Also containing ideas that could be extrapolated out to include pronunciation is the Natural Approach, which Krashen (1988) describes as focused solely on input in the target language (from the teacher) and potential output in the target language from the students; more importantly, only in cases of impeded communication are students' errors addressed and corrected (p. 106). It is unclear, however, what kind of errors are corrected. Moreover, although the related Natural Approach presents learners with input in the L2, the focus is still on communication (rather than accuracy) unless communication is affected by errors (Krashen, 1988); again, it would be possible to extend the principles of these approaches to address pronunciation, but without that extension, Krashen's theory does not appear to deem pronunciation particularly (or exceptionally) important.

Despite the focus of the cognitive perspective on issues pertaining to the acquisition of grammar (Lightbown & Spada, 2013; Mitchell & Myles, 2004), there are two principles presented that could be applied to pronunciation: fossilization and perceptual saliency, as the fear of errors fossilizing has been noted as a concern in a previous approach (Audiolingualism). Additionally, perceptual saliency refers to the manner in which humans perceive and organize information (Mitchell & Myles, 2004); there is nothing in this description that precludes it from applying to pronunciation. In terms of evaluation, however, there is at least a reference to fluency; Skehan (1998) provides a model for assessment of oral performance which alludes to the "smooth flow of language" as an indicator of a high level of proficiency, and mentions tempo as pertaining to a more intermediate level of proficiency (p. 178). Though these factors are tangentially related, they do address the manner in which the speaker produces the actual spoken language, just not at the segmental (phonetic) level. Additionally, Task-Based Language Teaching (TBLT), subsumed under the umbrella of the cognitive perspective, appears to allude to pronunciation, or at least the idea of linguistic needs, of which pronunciation could be one. Long (2015) states that "TBLT's way of handling students' linguistic needs is through provision of appropriate input, task-based practice, focus on form (not forms), and negative feedback, all during work on pedagogic tasks, not by task-supported language teaching" (p. 333). As appears to be a trend, though no explicit mention of pronunciation is made, one could argue that ideas such as appropriate input and negative feedback (Long, 2015, p. 333) would definitely be applicable to pronunciation instruction and/or acquisition. Thus, it appears that the cognitive perspective at least allows for individual interpretation to apply their concepts and principles to pronunciation.

Among the models, the connectionist model produced the most direct and overt reference to pronunciation that has occurred thus far. Gass and Selinker (2008) describe the connectionist model as one in which "[l]earning does not rely on an innate module, but rather takes place based on the extraction of regularities from the input [and] [a]s these regularities or patterns are used over and over again, they are strengthened" (p. 219). It is rather straightforward to utilize this model in conjunction with pronunciation as the focus is on language patterns. In fact, connectionism is closely related to models commonly found pertaining to L2 phonological acquisition, such as the Perceptual Assimilation Model (Best & Tyler, 2007) and the Speech Learning Model (Flege, 1995), as these theories function under the assumption that the sound systems of the L1 and L2 share the same phonological space and therefore interact (Leather, 2000, p. 19). Given that this model embraces extraction and repetition of patterns, it is unsurprising that pronunciation would be a focal point.

In contrast to the cognitive perspective, the sociocultural perspective posits that social interaction facilitates language acquisition and development; one of the most well-known theories therein is the Zone of Proximal Development (ZPD) which states that support from an interlocutor can assist the learner in performing at a higher level (Lightbown & Spada, 2013, p. 118). As this perspective places great emphasis on conversation and social interaction in order to construct knowledge, the only pronunciation elements included appear to focus more on suprasegmentals, such as style, intonation, and prosody, which have been thought to impact communication (Thorne, 2000; Lightbown & Spada, 2013). While still not placing too much focus on pronunciation, at least some elements are explicitly listed as factors that should be included when considering the context for communication (Thorne, 2000), an idea adopted by the subsequent approach, at least when it was first proposed.

The Communicative Language Teaching (CLT) approach was aptly named in that the overarching goal of this approach is the use of language in order to communicate. As communication and fluency are the main focus, one of the most salient features is the focus on meaning rather than on form (e.g., Lee & VanPatten, 2003); thus, if meaning can be conveyed successfully, strict adherence to correct formation of the language is not necessary. Although the emphasis is on communication, the modern application of the Communicative Approach appears to deprioritize pronunciation, despite the importance of comprehension noted in the early phases of the approach (Brumfit, 1984). According to Lightbown and Spada (2013):

> [w]hen communicative language teaching was first introduced in the late 1970s, little attention was given to the teaching of pronunciation. If it was taught, the emphasis was on suprasegmentals (rhythm, stress, and intonation) – aspects of pronunciation that were considered more likely to affect communication. (p. 68)

While this may be an accurate description overall, Brumfit (1984) points out that specifically pertaining to the goal of comprehension, activities in which the focus is more on form and accuracy (and less on meaning) can be beneficial because students may need to be informed of phonological and orthographic special cases in the L2 (p. 82). However, it appears that the communicative approach prioritizes other elements.

Lee and VanPatten (2003) outline the tenets of this method, explaining that typical activities and interactions in the foreign language classroom should be meaningful (i.e., conveying meaning or a message) and focus on language use. Additionally, they note that the role of the instructor differs from previous methods; in communicative language teaching, the instructor functions as more of a resource or facilitator. In keeping with this role, even the manner in which feedback is given differs: recasting and requesting clarification are common ways to attempt to call learners' attention to specific elements for improvement, rather than more explicit, meta-linguistic explanations of errors. Moreover, one of the few times Lee and VanPatten (2003) mention pronunciation is in conjunction with language testing, and even then the evaluation criteria are limited to only considering transfer and fluency (p. 110). In terms of what it views as important, this approach appears to value the communication and negotiation of meaning over simply producing the correct forms of the language (a description that could be applied to pronunciation), in which there is little connection to potential language use.

More recently, alternatives to the CLT approach have gained popularity, especially in areas where the population may already speak more than one language. Donato et al. (2012) explore intercomprehension and plurilingualism. Both of these approaches "develop language competencies by building on what students already know to foster more efficient and effective language acquisition" (Donato et al., 2012, p. 50). The authors note that these theories are prevalent in language instruction in Europe, where students often learn languages concurrently. The main tenet of intercomprehension is that the starting point of acquisition is grounded in what the learners already know about language. The use of prior knowledge (i.e., meta-linguistic awareness, and potentially contrastive analysis) creates an environment in which learners can progress faster due to support offered through the linguistic knowledge they already possess. Related to the concept of intercomprehension, Donato et al. (2012) also discuss the idea of plurilingualism. This theory centers around the idea of using a second (or third, fourth, etc.) language as a bridge, utilizing the languages the learners know as a stepping stone toward the acquisition of the new language. It is interesting to note that in most, but not all, cases of the implementation of these methods, the learners were young adults, suggesting that adult learners can benefit from meta-linguistic awareness between the language(s) they already speak and the one they hope to acquire.

The next section approaches second language acquisition from the standpoint of the acquisition of the phonological system of the second language by adults. Specifically, elements are explored that contribute to, are affected by, or can affect the acquisition of the phonological system of the second language. Subsequently, the section concludes with a discussion L1 interference and contrastive analysis.

2.2 Adult L2 Acquisition of Phonology

Much research has been done pertaining to the acquisition of L2 phonology, especially in cases in which the sounds in question only exist in one of the two languages. According to Zampini (2008), the majority of the research done in the field of the acquisition of L2 phonology assumes that a learner's L1 influences the L2, and thus focuses on the interaction between the languages mainly in this direction. Studies in which English is either the L1 or L2 often focus on the acquisition of vowels (Baker et al., 2008; Cobb & Simonet, 2015; Counselman, 2015; Flege et al., 2003; Gong et al., 2015; Müller Pograjc & Markič, 2017; Simon et al., 2010; Simon & D'Hulster, 2012) due to the larger quantity of vowels in the English phonological inventory than in the inventories of many other languages. Those that investigate Spanish but do not focus on vowels tend to investigate the voiced stops /b d g/ (Mackay et al., 2001; Mirisis, 2020; Nagle, 2017; Zampini, 1994), rhotics (Colantoni & Steele, 2006; Face, 2006; Olsen, 2012; Rafat, 2015), suprasegmentals (Gordon & Darcy, 2016[5]; Henriksen, 2013; Lozano-Argüelles et al., 2020; Saalfeld, 2012), and the voiceless stops /p t k/ (e.g., Flege, 1991; González López & Counselman, 2013; Kissling, 2013; Nagle, 2018; Stölten et al., 2014, 2015), among others.

The discussions pertaining to native-like pronunciation (in contrast to intelligibility or comprehensibility) and the potential effects of pronunciation instruction are likely to continue, especially as there seems to be little consensus in either one. The argument pertaining to native-like pronunciation is two pronged: (1) whether it is physiologically possible, and (2) whether it is

[5] In this study, suprasegmentals were investigated simultaneously with vowels.

a desirable outcome. Then, the debate regarding explicit instruction centers on whether it positively impacts pronunciation accuracy in L2 learners. Together, these ideas influence how the L2 learner is perceived (or judged) by others, particularly by native speakers of the L2. That people are able to make these kinds of judgments demonstrates another element that can positively or negatively affect communication, due to negative thoughts coming from both strangers and L2 learners themselves.

2.2.1 Physiological and/or Psychological Possibility of Native-like Acquisition of L2 Pronunciation

Stemming from early language acquisition theories, and particularly the Critical Period Hypothesis (CPH), it has long been debated whether second language learners could acquire sounds not present in their native inventories and, if they were able to acquire them, to what extent they could approximate the pronunciation of native speakers. The Speech Learning Model (SLM) in Flege (1995) stated that the phonetic systems of the L1 and L2 coexist, and as such, they interact with each other; this leads to the formation of new phonetic categories if the perception of the L2 sound is different enough from similar sounds in the L1. Though variables such as age play a role in the success or failure of category formation, studies have shown that it is possible to acquire sounds not present in the L1 phonetic system. Flege et al. (2003) found that early L1 Italian L2 English bilinguals were able to form a separate phonetic category for /eɪ/ (though they often over-articulated the diphthong), while late bilinguals were unable to do so; early bilinguals were categorized as those who moved to an English-speaking country between the ages of 2 and 13, while late bilinguals were designated as those who moved between the ages of 15 and 26 (p. 472). The over-articulation described by Flege et al. (2003) suggests that the bilinguals were able to recognize the sounds and their differences at the phonemic level, but also that they understood the physical properties of pronouncing the sounds.

Adult SLA, and particularly the acquisition of L2 phonology, differs from child and/or bilingual acquisition in ways that may either help or hinder the acquisition process. Admittedly, the CPH sparked an ongoing debate regarding the role that age might play in the acquisition of Spanish phonology (Flege et al., 2003), but this does not account for the variation found in the research. Many researchers have claimed that there is a relationship between the sound system of the L1 and that of the L2. Specifically, the Speech Learning Model (SLM), proposed by Flege (1995), "assumes that the development of speech skills remains intact across the life span of individuals and that L1 and L2 sound categories reside in a common phonetic space [. . .] [and therefore] interactions will arise between L1 and L2 sound categories" (as cited in Simonet, 2014a, pp. 731-732).

In keeping with SLM, the investigation conducted by Olsen (2012) supports the idea of the phonological systems of the L1 and the L2 interacting, as he found that L2 learners of Spanish were more likely to accurately produce taps if the tap occurred in the same context in both languages. Olsen (2012) claimed that not only were the L2 learners accessing information about both languages simultaneously, but they were also able to compare the use of the sound across languages. Based on potential conclusions drawn from Olsen (2012), it is possible that metalinguistic awareness can impact adult learners' perception and production of L2 sounds. For example, Counselman (2015) found that the more attention paid to contrasts between the L1 and L2 vowels (in this case through explicit instruction), the more likely the L2 Spanish learners were to perceive the differences and form new categories rather than assimilate to existing ones, and that this improvement in perception can result in more accurate production. Additionally, pertaining to lexical stress, Henriksen (2013) presents a study that found differences in perception of stress in minimal pairs that differed by that element, attributing the findings to the age of L2 acquisition. The study found that the late learners had high numbers of errors distinguishing between words that differed by stress, but performed much better when distinguishing between minimal pairs that differed by one phoneme.

Despite previous research having described (relatively) successful interactions between learners' L1 and L2 inventories, it has been much debated in the research whether, and/or to what extent, L2 learners can perceive non-native sounds they might encounter when acquiring a second language, as well as whether or not they can successfully produce them in a manner akin to that of a native speaker. Perception, therefore, has been of great interest to research in this field and others, particularly when considering the beginning stages of acquisition. Unlike acquisition at higher levels where there is already some foundation to build from, in the beginning the experience of the L2 learner can be similar to the process of loanword adaptation: the L2 learner hears a word containing unfamiliar sounds and attempts to represent that word utilizing the phonological system of the L1. However, L2 learners are not simply adapting words in isolation; thus, a variety of methods have been proposed in an attempt to represent and explain the manner in which L2 learners treat non-native sounds. In contrast, previous research, especially from proponents of the perceptual approach in the field of loanword phonology, has argued that non-native sounds are simply adapted to the closest sound in the native inventory because the phonetic or phonological information of the L2 is inaccessible (Paradis, 2006; Peperkamp & Dupoux, 2003). This highlights the difference between borrowing and acquisition: in a borrowing situation the speaker may not encounter the need to adapt a particular sound again for some time, while in an acquisition setting, such as the foreign language classroom, the learners' perception of L2 sounds may improve due to more consistent input.

Although studies have supported the idea of native-like pronunciation attainment, research has also shown that there may be limitations to the acquisition of non-native sounds. Primarily, the ability to fully reach native or near-native proficiency has been questioned. For example, Flege and Hillenbrand (1984) found that more experienced learners are able to produce sounds in their L2 in a more native-like manner; in spite of this, Flege and Hillenbrand (1984) also stated that “Flege’s [SLM] model leads to the prediction that, with sufficient experience, L2 learners will produce stop consonants differently in L2 than L1 (if the L1 and L2 stops differ phonetically), but will never perfectly match native speakers of L2” (p. 709). In this particular instance, the L2 learners in question spoke English as their L1 and were able to produce the French voiceless stop /t/ with a VOT shorter than the accepted values for English /t/. The differences in VOT duration found by Flege and Hillenbrand (1984) fit within research on adult L2 acquisition of VOT which has shown that VOT values often fall between the accepted values for the L1 and L2 for adult L2 learners (Zampini, 2014; Nagle, 2019). Findings such as these, where native-like pronunciation was not achieved, may reflect the idea of the CPH, which claimed that after a certain period of time, it is difficult if not impossible to acquire new sounds and/or produce them with native-like proficiency (Gass & Selinker, 2008; Lightbown & Spada, 2013). Research has even progressed to the point of investigating whether explicit instruction in pronunciation affects the acquisition of non-native sounds (Alves & Luchini, 2020; Alves & Magro, 2011; Camus, 2020; Counselman, 2015; González-Bueno, 1997a; González López & Counselman, 2015; Kissling, 2013, 2015; Yang, 2017), but thus far, there appears to be little impact on pedagogy, potentially due to the lack of a clear consensus on the effects of pronunciation instruction.

2.2.2 Native-like Acquisition of L2 Pronunciation as an Instructional Goal

Particularly in the classroom, communication has become the focus and goal of L2 acquisition (Lee & VanPatten, 2003), and pronunciation has often been deemphasized in order to achieve communicative competency in the limited timeframes typical of classroom foreign language instruction (Hurtado & Estrada, 2010; Martinsen et al., 2014). The focus on communication without considering pronunciation, however, has a flawed logical premise: that communication can occur successfully at any level of intelligibility. It is true that communication can occur between people with few if any linguistic commonalities, as is the case in the creation of pidgins (Brumfit, 1984; Edwards, 1995; Hughes, 1983), but if the goal is the acquisition of the language, the aim ought to be greater proficiency than that of a pidgin. Admittedly, comparing the early stages of learning a foreign language to a pidgin is an extreme viewpoint, but it highlights a gap in the research pertaining to effective communication in the L2, as pronunciation appears to still largely be deprioritized but communication is paramount, despite good or poor pronunciation potentially impacting the success (or failure) of communication (Agostinelli, 2012; Anthony, 1996; Arteaga, 2000; Gleason, 2012; Gordon, 2012; Hurtado & Estrada, 2010; Llanes et al., 2017; MacDonald, 2002; Martinsen et al., 2014; McBride, 2015; McCrocklin, 2012; Morin, 2007; Schairer, 1992; Zielinski, 2012).

Studies focusing on attitudes toward and perceptions of speech accentedness may support the argument in favor of attaining native-like pronunciation (often through explicit instruction) for reasons related to affect. McBride (2015) found that, when presented with speech samples from adult Spanish language learners whose speech contained a clear L2 accent, native Spanish speakers easily identified both segmentals and suprasegmentals that contributed to the speech sounding foreign (the trill /r/ was mentioned frequently, as was intonation), and these inaccuracies were linked to negative perceptions of the speaker. In terms of how these findings relate to L2 instruction, McBride (2015) concluded that "[t]he most fundamental implication of the results of [the] study is that L2 pronunciation matters, not only in terms of comprehensibility, but because it has a serious impact on native and non-native speakers' initial impression of the speaker" (p. 27). Therefore, to combat negative feelings (such as anxiety) that L2 learners may have toward their own pronunciation, addressing and amending the features that give others a negative impression of L2 speech can benefit learners greatly.

Relatedly, Martinsen et al. (2014) investigated the perception of foreign accent in Spanish L2 learners from various levels of instruction, as well as with varying amounts of time spent abroad. They found that while more exposure to the foreign language did create significant differences in pronunciation, the pronunciation of L2 learners who had spent a large amount of time abroad was still perceived as significantly different than that of native speakers. Martinsen et al. (2014) attribute this to a developmental plateau, stating that "as students' progress toward native-like pronunciation reached the level of comprehensibility at which their communicative needs could be easily met, improvement in their pronunciation slowed" (p. 74). Due to this plateau, they argue that L2 language growth should still be encouraged even though coherent communication has been achieved, and that this should be accomplished by emphasizing and teaching pronunciation as well as evaluating how native-like the learners' communication is, rather than whether or not communication is successful (Martinsen et al., 2014).

In addition to impressions of L2 speech, previous studies investigated explicit instruction in pronunciation and its effects on the acquisition of non-native sounds (Alves & Luchini, 2020; Alves & Magro, 2011; Camus, 2020; Counselman, 2015; González-Bueno, 1997a; González López & Counselman, 2015; Kissling, 2013, 2015; Yang, 2017). González López and

Counselman (2015), who investigated students enrolled in second-semester Spanish classes, found that training in articulatory phonetics positively impacted VOT values in the L2, observing that even at beginning levels, L2 learners are able to produce separate VOT values for voiceless stops in their L1 and L2. Similarly, Counselman (2015) concluded that the combination of explicit instruction in articulatory phonetics and metalinguistic awareness of perceptual differences between English and Spanish pronunciation produced more improvement in the learners' pronunciation accuracy than instructor-provided feedback. Additionally, Casas Solà (2014) found a marked difference in the VOT values of L2 English speakers between their first and second years of university study, which was attributed to the course on phonetics and phonology the students took between those years.

Some of the most current research to support the idea of native-like pronunciation being a goal in L2 acquisition stems from Computer-Assisted Pronunciation Training (CAPT). McCrocklin (2016) supports the idea of technology, and particularly ASR, providing learners with autonomy when learning and practicing pronunciation, especially as she mentions that previous research on ASR has found that it can improve pronunciation accuracy. She acknowledges that "pronunciation is often downgraded as a teaching goal and pushed aside in favor of other skills" (McCrocklin, 2016, p. 25), and thus promotes the use of ASR outside of the classroom as a way to work toward more accurate (and therefore, more native-like) pronunciation without sacrificing class time. Additionally, McCrocklin (2016) also addresses an element related to the affective filter (Krashen, 1988), namely anxiety, which working with ASR may minimize as students can practice on their own but still receive instruction and/or feedback. Thus, their affective filter is down, and they can receive input and acquire the language.

Relatedly, Olson (2014) investigated the effects of the use of speech analysis software on beginning learners' pronunciation accuracy. His participants utilized PRAAT to compare their speech to that of native speakers through the use of waveforms and spectrograms. Olson (2014) stated that "[t]he design [of the study] allows students to analyze their own recordings, compare their productions with the productions of a native Spanish speaker, and attempt to produce the target sounds in a more native-like manner" (p. 53). Though it may not be feasible to spend an exorbitant amount of time on the instruction of pronunciation in the classroom, Olson (2014) supports the claim that that pronunciation-specific technology can be implemented even in beginning levels of foreign language classes. While the purpose of his study pertained more to the methodology behind the implementation and evaluation of the use of PRAAT for pronunciation instruction, Olson (2014) does acknowledge anecdotally that many participants were able to improve their own pronunciation by identifying the differences between their pronunciation of the Spanish /d/ intervocalically (where it should be pronounced [ð] rather than [d]) using spectrograms. With more empirical evidence, the observations noted in Olson (2014) could indicate that even beginning L2 learners can benefit from this kind of instruction and become aware of pronunciation differences from the early stages of L2 acquisition.

In contrast to the aforementioned viewpoints, there is also evidence against native-like pronunciation being a goal of second language acquisition. One argument stems from the idea of intelligibility, which refers to the ability of a speaker to successfully communicate a message, while less importance is placed on pronunciation accuracy. Munro (2008) argues that "the fact that millions of second language users around the world communicate using foreign-accented speech indicates that accent-free pronunciation is *not* a necessary goal for either learners or teachers of second languages" (p. 194). Additionally, in a survey of pronunciation instruction, Derwing (2008) found that while there is a specific set of segmentals that is frequently the focus of pronunciation instruction in English as a Second Language (ESL) classrooms, much more emphasis has been placed on suprasegmentals such as rhythm, stress, intonation, and prosody when instruction was aimed to improve learners' intelligibility. Moreover, two studies cited in Derwing (2008) mentioned that although segmentals are more frequently noticed by the learners themselves as problematic, often the sounds in question are less frequently used in the language and thus less likely to impede communication when they do occur (pp. 334–335).

Regarding pronunciation instruction in foreign language classes, the argument in favor of intelligibility (and thus against the need for native-like accuracy) presented in Munro (2008) fits well with communicative teaching methodology, as both favor communication over precision. Arguments against native-like pronunciation being a goal of second language instruction are also supported by the observation that research related to pronunciation instruction does not provide a clear picture of success or failure. For instance, Díaz-Campos (2014) mentions that studies may have found that L2 learners' pronunciation improved after instruction, but the improvement was often incidental. Relatedly, Simon and D'Hulster (2012) claim that "neither pronunciation instruction nor continued exposure [to the L2] alone guarantee success [in L2 pronunciation improvement]" (p. 271), indicating that multiple factors are likely at work in the acquisition of L2 pronunciation.

In comparison to Díaz-Campos (2014) and Simon and D'Hulster (2012), some studies have found more conclusive evidence that receiving a treatment of phonetics instruction did not show significant differences. Kissling (2013) found that instruction in articulatory phonetics had very similar effects on the pronunciation accuracy of university students enrolled in first, second, and third-year Spanish courses when compared to activities consisting of listening with a dictation element added. Hurtado and Estrada (2010) found that L2 learners enrolled in Spanish pronunciation classes were equally effective in improving the pronunciation of Spanish vibrants when the classes were taken in the US or abroad; studying abroad, however, did improve the frequency of correct utterances. However, the apparent lack of consensus, at least in terms of the effectiveness of the instruction of articulatory phonetics, may stem from methodological differences, as Kissling (2013) stated that the instruction provided to both groups was very similar, and the introduction of variables not directly relevant to instruction in Hurtado and Estrada (2010) may have had an impact, as they do acknowledge that "[t]he explicit instruction helped students become aware of the articulation of the Spanish vibrants, their phonological contrast, allophonic distribution, and differences with the English retroflex" (p. 84). These extra variables, combined with their findings that the pronunciation of Spanish vibrants improved, may indicate the interference of some of their sociolinguistic variables.

In addition to student-centered rationales, the standpoint of the instructor could influence decisions regarding explicit pronunciation instruction. As noted in Derwing (2008), studies that have investigated ESL instructors have found that they may often believe themselves to be ill-equipped to teach pronunciation because for many, phonetics or phonology courses were not part of their graduate degrees. Thus, when these instructors were required to teach pronunciation, only segmentals were taught because that was the focus of the pronunciation software available to the instructors. Relatedly, Derwing (2008) also cites the transmission (or the fear of transmission) of inaccurate information regarding pronunciation as another reason why ESL teachers are hesitant to teach pronunciation. Studies have found that without sufficient training in phonetics, ESL instructors have erroneously attributed differences between English tense and lax vowels as pertaining to length rather than quality, for example (Derwing, 2008, pp.

362–363). Still pertaining to ESL, MacDonald (2002) mentioned other rationales in addition to those found in Derwing (2008). From in-depth surveys, MacDonald (2002) received commentary in which instructors expressed that they disliked teaching pronunciation and/or that they felt they "were not good at teaching it" (p. 4).

Related to the idea of being *ill-equipped*, mentioned by Derwing (2008), there are other areas that could impact the drive or desire to teach pronunciation: specifically, course textbooks (Arteaga, 2000) and standards for teacher training (Morin, 2007). Arteaga (2000) surveyed ten textbooks published for use in beginning Spanish courses, but found that less than half contained sections covering phonetics, and many of the sections were deficient, incorrect, and/or confusing (pp. 346–347). While the textbooks investigated by Arteaga (2000) seemed to contain too little information, the analysis conducted by Morin (2007) of the NCATE Program Standards for the Preparation of Foreign Language Teachers revealed an overwhelming amount of expected skills, specifying:

> foreign language teachers who understand the rules of the sound system of the target language, can describe how target language sounds are articulated, are able to describe target language phonological features, and can diagnose their own target language pronunciation problems and those of their students. (p. 342)

The large amount of variation in support for pronunciation instruction between Arteaga (2000) and Morin (2007) likely had an impact on instructors' attitudes. In either case, the paucity or overabundance of knowledge and information could make teaching pronunciation appear daunting.

2.2.3 Assessment for L2 Acquisition of Pronunciation

Evidence against native-like pronunciation being a goal of second language acquisition, particularly in terms of instruction, can also be found in standards and evaluation rubrics. Current pedagogical methods and language learning standards reference speech production, as oral communication is one of the main goals, but place comparably little emphasis on pronunciation, even in conjunction with proficiency evaluations (Lee & VanPatten, 2003). According to Lord and Fionda (2014):

> the American Council on the Teaching of Foreign Languages (ACTFL) 2012 Guidelines for Speaking Proficiency repeatedly talk about fluency and accuracy, but fail to define those constructs in any particular terms. The definitions and descriptions given lead us to believe that these are largely with respect to lexical, morphological, and syntactical structures used. Pronunciation and phonology, or accent, are mentioned infrequently and only in the negative: when they will cause problems. (p. 516)

Relatedly, Lee and VanPatten (2003) present a test that can be used to evaluate a language learner's oral proficiency, the Oral Proficiency Interview (OPI), created by the American Council on the Teaching of Foreign Languages (ACTFL, 2015). In their discussion of the evaluation of learners' responses during these tests, Lee and VanPatten (2003) note that the criteria for evaluation for the OPI are derived directly from the ACTFL Proficiency Guidelines, which appear to reference pronunciation in terms of the potential effects of L1 transfer (pp. 109–110). Also mentioned by Lord and Fionda (2014), the most recent ACTFL Proficiency Guidelines reference evidence of L1 influence or transfer as well, but only at the Novice Low, Novice High, Intermediate Low, and Intermediate Mid levels, and in terms of them detracting from communication (ACTFL, 2012).

Moreover, the ACTFL World Readiness Standards (ACTFL, 2015) do not explicitly mention pronunciation. These standards have five main goals, referred to as the *5 Cs*:

> The five "C" goal areas (Communication, Cultures, Connections, Comparison, and Communities) stress the application of learning a language beyond the instructional setting. The goal is to prepare learners to apply the skills and understandings measured by the Standards, to bring a global competence to their future careers and experiences. (The National Standards Collaborative Board, 2015, p. 2)

Within the goals denoted in the World Readiness Standards, it can be difficult to determine where pronunciation might fit. However, the fourth *C*, Comparisons, does allow for the development of some metalinguistic awareness, as it states that "[l]earners use the language to investigate, explain, and reflect on the nature of language through comparisons of the language studied and their own" (ACTFL, 2015, par. 9), though this does not directly reference pronunciation as something to make comparisons about. Additionally, in Cutshall (2012), which explored the fourth *C* in depth, the only practical example included that could be related to pronunciation was a comparison through the use of cognates (which often are similar both phonetically and orthographically) (par. 10), though pronunciation itself was not mentioned. Hence, it appears that it may be left to the discretion of the instructor to determine whether pronunciation will be taught, and if so, how and how much. Current L2 teaching methodologies tend to favor communication over accuracy, and those principles are reflected in the focus of standards and rubrics.

2.2.4 L1 Interference and Contrastive Analysis

It is not unusual for the claim to be made that the phonological system of a speaker's first language has the potential to interfere with the process of acquiring a subsequent system (Chang, 2012; Flege et al., 2003; Major, 2008; Piccinini & Arvaniti, 2015). In order to explain this phenomenon, Major (2008) contrasts surface and abstract transfer. While surface transfer pertains to the interaction of the sounds produced, phonological abstract transfer focuses on the perception of the L2 phonological inventory in comparison to elements of the L1 (Kartushina & Frauenfelder, 2014). Relevant to the present study, Kartushina and Frauenfelder (2014) state that "[s]tudies showing the effects of L1 orthography on L2 perception, for example, are generally taken to favor abstract transfer [. . .] since orthography is an abstract representation" (p. 1). Relatedly, Zampini (2008) comments that abstract elements may be more predisposed to transfer (p. 236).

Though the present study does not investigate perception, at least not in an aural sense, the presence of the written form of the language may cause transfer of the abstract, symbolic representations of language between the L1 and the L2, especially as English and Spanish overlap in their use of the same symbolic system. According to Goikoetxea (2006),

> [t]he Spanish orthography has 29 graphemes: the 26 letters of the English alphabet plus *ñ* and two digraphs, *ch* and *ll*. But in contrast to English, which has over 40 phonemes, Spanish only has 25 phonemes. Of the 29 Spanish graphemes, 24 graphemes [. . .] represent a unique sound. (p. 334)

To illustrate an example of L1 orthographic interference, Major (2008) mentions the issue of the interpretation of the Spanish letter (i.e., grapheme) <r>. Even though [ɾ] exists as an allophone in English (Eddington & Elzinga, 2008), the typical adaptation of the Spanish grapheme <r> is [ɹ]; it is probable that this is due to the association of <r> with [ɹ] in English, while the orthographic representation corresponding to [ɾ] in English is <tt>, as in words like *letter* [ˈlɛɾɹ̩]. Additionally, pertaining to phonetic interference and vowels, Flege et al. (2003) concluded that

L2 English learners tended to produce a monophthong instead of the diphthong /eɪ/ due to interference from their L1, Italian, which contains /e/ but not the diphthong that is present in English.

Although pronunciation would benefit from receiving more attention in the foreign language classroom, unfortunately, due to time constraints and pedagogical trends, teachers may elect to move on to another topic or activity after such a comment instead of briefly focusing the students' attention on the pronunciation issue. While pausing to address every pronunciation error would leave little time for anything else to be accomplished, pointing out, explaining, and training larger L2 pronunciation issues can lead to improvements in students' pronunciation (Counselman, 2015; Huensch & Tremblay, 2015; Inceoglu, 2016; Kissling, 2015). In cases where the L1 and the L2 do not align, previous research has shown that increased metalinguistic awareness positively impacted pronunciation accuracy (Counselman, 2015; Kissling, 2015; Simon & D'Hulster, 2012). Thus, this section explores misalignments between the grapheme-phoneme correspondences in the L1 as compared to the L2 through the lens of Contrastive Analysis, with an explanation of its relevance to the present study.

The present study aligns itself with the tenets of Contrastive Analysis because it investigates the acquisition of an L2 (Spanish) that has an almost-identical writing system as the L1 (English), but the correspondence between grapheme and phoneme differs between the two languages. Leather (1999) describes the Contrastive Analysis Hypothesis (CAH) as a hypothesis that proposes that structural differences between a person's L1 and L2 will be problematic (p. 26). As this is a rather general definition, Cichoki et al. (1999) expand on this definition, not only by adding in the idea that not only does the CAH pick up on differences between the L1 and the L2, but also by introducing the idea that the CAH contrasts the two languages and is able to predict which elements will be more or less problematic to learn depending on how similar or different the given element is in one language versus the other (p. 97). Additionally, Eckman (1987) mentions a weak version of the CAH; this version simply charges linguists to utilize their knowledge of linguistics to explain difficulties observed during the process of L2 acquisition (p. 56). These definitions and explanations help to pinpoint exactly what the CAH does: it promotes the comparison (or contrast) of the L1 and the L2 in order to uncover similarities and differences, and from those, it attempts to predict the errors that may occur in L2 speech based on the divergent features of the two languages. One salient feature of the CAH as described in Eckman (1987) is the use of knowledge of linguistics, which fits in well with what the CAH proposes to do, as well as with potential interpretations of the hypothesis.

It is interesting to note that the definitions of the CAH could be considered rather vague due to little explanation of what the predictions entail, or how to go about making them. They also often involve syntactic elements. For example, Odlin (2014) claims that:

> [r]ecent studies [. . .] show results that make plausible some predictions for novel SLA contexts (e.g., 'Speakers of Finnish as a group will have a greater difficulty with the articles of Portuguese than will speakers of Swedish as a group'). Although predictions involving structures such as articles seem viable [. . .], there are limits on predictability. Group tendencies often seem predictable but individual behavior much less so. (p. 27)

Due to skepticism such as this, others have attempted to amend the CAH to fix its perceived issues. For example, Odlin (2014) mentions *a priori* contrastive analysis, which is based in actual linguistic comparisons between languages, which seemed to focus the application of the hypothesis somewhat. However, Odlin does acknowledge that prediction still cannot account for everything. In particular, he mentions bidirectionality, which refers to the prediction that what is problematic in one language will also be problematic in the other. All of these are issues that must be considered when using the CAH as framework for analysis, which is why the present study aligns itself with tenets of the approach instead of adopting it completely. Specifically, the present study does not investigate bidirectionality (Odlin, 2014), and all predictions relate to the likelihood of a target response based on group membership and/or word location, rather than predicting the problematic elements themselves (Cichoki et al., 1999).

With current advancements in technology, SLA generally has attempted to utilize these advancements to enhance language learning, particularly in the classroom. The acquisition of pronunciation has also been impacted by new technology, as it has provided opportunities for learners to compare their voices to that of native speakers and receive feedback on their own pronunciation, for example. The subsequent section describes the implementation of Technology-Assisted Language Learning, divided into the subcategories of Computer-Assisted Language Learning (CALL), Mobile-Assisted Language Learning (MALL), and Computer-Assisted Pronunciation Training (CAPT), in L2 acquisition settings.

2.3 Technology-Assisted Language Learning

Beyond perceived and/or ideal pedagogical utility, incorporating technology in the language learning process must have a clear purpose and goal. Otherwise, it is simply using technology for the sake of using technology, and the benefits to the learners are unclear, obscured, or absent altogether. Golonka et al. (2014) reviewed more than 350 studies regarding the use of technology in foreign language learning and in the foreign language classroom. While too numerous to describe in their entirety, the total value of the use of technology was found to be small, although there were some key areas in which Golonka et al. (2014) were able to claim that there had been an impact on the learning of the foreign language, namely focusing on pronunciation (through the use of speech recognition). This result, they argued, was due to the feedback provided to the students in the form of a comparison between their pronunciation and what the technology was programmed to recognize.

Feedback being provided by a non-human entity, as was mentioned in Golonka et al. (2014), is one of the many ways that modern technology can impact the foreign language classroom and foreign language instruction. As technology becomes more ubiquitous, and as the students present in the classroom may not remember a time when a specific technology (e.g., smartphones) did not exist, modernizing language learning tools is crucial in order to connect with these students. Thus, an implementation of technology in the classroom has begun, first with computers and later with smartphones, tablets, and other mobile technology.

As technology advanced, Computer-Assisted Language Learning (CALL) developed due to the continuing implementation of computers (and later, other technology) in the language classroom. Arnold and Ducate (2011) define CALL as language improvement through the use of technology, which in the beginning, was limited to only computers. They also enumerate the potential uses of CALL: in the early stages, it was utilized as a tutor or a tool; subsequently, communicative principles were added to it, shifting the focus to acquisition rather than explicit learning and instruction. More recently, computer-mediated communication, as well as wikis, blogs, and social networks, have been adapted and tailored to the foreign language classroom. Youngs et al. (2011) add online distance learning, synchronous and asynchronous tools, and virtual learning environments to this extensive list. However, despite all of these uses, Arnold and Ducate (2011) acknowledge that CALL is still not perceived as precisely vital to L2 instruction.

As can be seen from the variety of potential uses of the technology, Youngs et al. (2011) describe CALL as methodologically flexible. However, they also caution against the use of CALL technology without proper grounding in learning objectives, pedagogy, and second language acquisition theory and research, but also acknowledge that proper implementation can positively impact areas such as learner autonomy, motivation, and collaborative learning. The following sections outline recent developments in computer-assisted language learning, including the emergence of Mobile-Assisted Language Learning (MALL); this section ends with a critical evaluation of MALL applications, reiterating the standpoint expressed by Youngs et al. that language learning technology needs to be linguistically and methodologically sound.

A popular topic within CALL is the use of digital games to facilitate language learning and practice. One of the more popular gaming styles is MMORPG (Massively Multiplayer Online Role-Playing Game), in which players connect online from all over the world, often working together to accomplish tasks in the game. Role-playing games generally put players into different categories or characters (avatars) with specific abilities or skills. These avatars are typically highly customizable, so much so that Pasfield-Neofitou et al. (2015) claim that "students often do not distinguish between themselves and their avatars; the boundaries between the real and virtual environments are highly (cognitively) permeable" (p. 709). This creates a unique environment for the development and facilitation of language learning and practice.

In addition to virtual worlds, digital platforms also have the potential to support and augment foreign language instruction. For example, previous studies that have investigated the effectiveness of incorporating *Second Life*, a massive online platform accessed by means of computer software, in the classroom have used it more as a conversation tool, where students were able to meet together and chat, or, in only a few cases, as a place to create a presentation together (Hsiao et al., 2015; Zhang, 2013). However, feedback from participants in studies such as Zhang (2013) reported that limitations of using *Second Life* as a conversation tool, mainly that conversation partners were unreliable and that more formal language was unavailable. Similarly, almost all the feedback about the game utilized in Culbertson et al. (2016) included a comment or a request for a voice feature to be incorporated into the game. These virtual worlds have the potential to fit well with foreign language learning because communication is essential to effective collaboration. There is also a great deal of potential to connect with speakers of other languages, though the technology is by no means perfect.

In many cases, the games investigated in CALL research required the use of a computer or a gaming console, limiting the areas in which the technology could be used. The smartphone, however, debuted in 1993[6], though the term itself dates back to the early 1980s and was written as a separate adjective and noun[7]; the first Blackberry was available in 1999[8]. With 3G, which was introduced in 2001, smartphone features became more advanced, including the capability to send and receive photos, videos, and music, though arguably the most notable addition was Internet access[9]. Then, the first iPhone launched in late 2007, selling 6.9 million units in one fiscal quarter[10], followed by Android in 2008[11]. As smartphones were developed and subsequently increased in popularity, many games that once would have needed computers or full consoles were transformed into mobile applications, permitting anyone to play them anywhere as long as their device had cell service (or WiFi) and enough battery. Thus, the MALL subfield emerged.

---

[6] (Hosch, 2020, par. 2)
[7] (Oxford University Press, n.d.-r)
[8] (Heathman, 2017, par. 10)
[9] (Hosch, 2020, par. 2)
[10] (Johnson, 2008, par. 4)
[11] (Heathman, 2017, par. 14)

Though often connected with smartphone use, MALL refers to any handheld technology that can be used to assist with language learning, differing from CALL in that, in order to be considered mobile, the technology has to be personal and portable (Demouy & Kukulska-Hulme, 2010). Thus, this definition extends the scope of "mobile technology" to include iPods and MP3 (in newer iterations, also MP4) players (Demouy & Kukulska-Hulme, 2010; Stockwell & Hubbard, 2013), smartphones (Chuang, 2015; Duman et al., 2015; Stockwell & Hubbard, 2013), eReaders and tablets (Duman et al., 2015; Stockwell & Hubbard, 2013), as well as electronic dictionaries, digital voice recording devices, and handheld game consoles (Duman et al., 2015). Some studies have even included laptops in this category (Stockwell & Hubbard, 2013), as they are mobile, personal, handheld, and portable, though there is debate over whether laptops will be considered as part of MALL going forward (Viberg & Grönlund, 2013). It is important to distinguish between technology considered part of CALL versus MALL because the addition of mobility changed the dynamics the technology; the fact that MALL technology can be accessed and used anywhere provides a flexibility that CALL technology does not have, but it also introduces troublesome elements such as the potential for distraction.

Out of the aforementioned technologies, smartphones are among the most frequently researched, and as such, so are the applications that they can run. Steel (2012) reported on the manner in which mobile apps are used by language learners. The original 590 responders to a survey given to university students were narrowed down to the 134 who listed mobile apps within their top three most useful technologies for language learning outside of the classroom. These students represented ten foreign languages in total: Chinese, French, German, Indonesian, Italian, Japanese, Korean, Russian, Spanish, and Portuguese. Analysis of student comments led to the investigation of the following themes: "(a) the ways students used mobile apps to maximise time, location and opportunity to fit language learning into their daily lives, and b) the specific learning benefits students perceived through their use of mobile apps for language learning" (Steel, 2012, p. 2). Common themes among student responses pertained to efficiency pertaining to time management and use, as well as ease of use. The comments on language-specific learning benefits centered around apps being used for vocabulary or grammar purposes (translators, dictionaries, etc.), which in turn assisted with acquisition of vocabulary. Steel reports an interesting finding, or really a lack of a finding, in that few if any students mentioned the fact that the technology they were currently using for language learning was originally not meant for this purpose, but rather for connecting them with others through phone calls, text messages, or various mobile applications.

While the use of technology in the foreign language classroom can be beneficial, it also must be done thoughtfully. The assumption that something being published and made available for public consumption makes it legitimate and accurate is erroneous. Especially as the development and publication of mobile applications has become more accessible, a critical eye is necessary in determining the value of apps in general, but especially language learning apps. Martín Monje et al. (2014) reported the process of creating a rubric designed to evaluate language learning mobile applications (i.e., apps). It centered particularly on apps that could be used in an English as a Foreign Language (EFL) setting. A total of 67 apps were included in the assessment for the creation of the rubric. The first phase of analysis focused on the functionality of the apps, and found that many had technical problems from the start, ranging from not functioning at all to not functioning properly. The majority were available for use on most Apple products, and about 25 percent were available on Android devices, and there were even fewer that were available for platforms other than the aforementioned two. Pertaining to pricing, the apps they reviewed were categorized into expensive, a few euros, and free. They also created different categories for the apps, from games to dictionaries, grammar/vocabulary/pronunciation practice, adaptation of online courses, and using language in context (podcasts, videos, films, cartoons). Finally, they noted the special features that some of the apps had, such as drag-and-drop, social network connectivity, and avatars. The rubric then was created focusing on two areas: pedagogical criteria and technical criteria; the Common European Framework of Reference (CEFR) was used as a benchmark. The investigators concluded that "[t]his evaluation made clear the fact that apps initially attractive to the user of MALL are not necessarily backed up by a sound linguistic content that is adequate for steady language learning" (Martín Monje et al., 2014, p. 10) as many of the apps did not score very high once the rubric was combined with the elements from the CEFR.

Despite the fact that the present study does not investigate the use of technology in the classroom, and in fact opines that the mobile technology in question should not be used during class time but instead as supplementary practice, the relevance of instruction to MALL cannot be completely set aside. One could argue that all three of the technologies used in the present study (online grammar-related activities, online pronunciation activities, a language learning app) still contain elements of instruction even though none of them require or have a human instructor who is physically present. Of particular importance is the type of instruction found in the app (see Section 3.4.1). First and foremost, the app was bilingual, with the avatar instructor, Sofia, presenting concepts in both English and Spanish. Even the activities contained this bilingual element; for example, in the app's conversation building activity (present in most lessons), the learner heard and saw Spanish, but also saw English simultaneously. This was the app's biggest nod to *focusing on meaning*, the cornerstone of the communicative approach. It also should be noted that the app, which was designed for those who wish to learn a language but cannot do so in a more traditional environment, demonstrated a flexibility in instructional methodology that is often seen in more traditional language learning settings (in comparison to online instruction, for example).

2.3.1 Computer-Assisted Pronunciation Training (CAPT)

While both CALL and MALL have changed the manner in which it is possible to learn a language, a specific element of language acquisition is still often overlooked: pronunciation. As has been mentioned previously, many of the uses of technology in the classroom and/or for language learning in general utilize technology as a medium for communication (Culbertson et al., 2016; Liu et al., 2016) rather than as a tool for practicing and improving a given language element, particularly speech. Though some studies acknowledged that it was possible to communicate orally using the technology in question (Hsiao et al., 2015; Linden Research, Inc., 2018; Zhang, 2013), few addressed the potential for pronunciation improvement (Golonka et al., 2014). In general, it is often the case that pronunciation is passed over even in traditional learning environments in favor of other topics (Levis, 2007; Offerman & Olson, 2016), especially in non-immersive environments where there are significantly fewer contact hours between teacher and students. The deprioritizing of pronunciation is surprising due to the high importance placed on communication in current foreign language teaching pedagogy (e.g., Hurtado & Estrada, 2010; Lee & VanPatten, 2003; Martinsen et al., 2014; Wong & VanPatten, 2003). As technology continues to develop yet still remains accessible to the masses, its potential utility in facilitating the acquisition and development of pronunciation through computer-assisted training will only increase over time.

Technology itself continues to advance and expand the capabilities of artificial intelligence and machine learning. For example, in 2011 Apple introduced Siri, its virtual assistant (Hempel, 2016), which can process speech and, from that input, perform a variety of tasks including making a (usually) intelligible response, following commands, and converting the audio input to text. The debut of Siri was followed closely by Amazon's Alexa in 2014 (Hempel, 2016), which has similar capabilities. By 2019, Siri spoke and understood 23 languages, as well as regional varieties of some of those languages (iOS Feature Availability, n.d.). Although this technology has been in existence for nearly a decade, and has clearly been shown to be able to accommodate multiple languages, its use in SLA, and more specifically the acquisition of L2 pronunciation, is still quite limited. Thus, the present study aims to demonstrate that CAPT (i.e., interactive practice specifically focused on targeted pronunciation improvement) can positively impact the acquisition of L2 phonology, and that current advances in technology can facilitate pronunciation accuracy. The subsequent section first outlines the ways in which technology and pronunciation training can be used to improve L2 pronunciation, describing previous research into pronunciation feedback generally before specifically focusing on ASR and its potential for providing feedback to real-time L2 speech.

2.3.1.1 Pronunciation Improvement and Feedback. Pronunciation has been shown to either hinder or facilitate communication (Agostinelli, 2012; Arteaga, 2000; i.a.); thus, techniques to improve pronunciation accuracy would be of great benefit to L2 learners, and technology provides innovative methods for pronunciation training and practice. However, before production can begin, it is crucial to understand what L2 learners can perceive in terms of sounds not found in their native inventories; after taking this information into account, feedback can then be provided to L2 learners which enables them to reflect on their pronunciation and use the feedback as a tool to hone their pronunciation through training.

The type of feedback typically seen in conjunction with pronunciation training is corrective. Unlike recasts or repetition typically seen in conjunction with language instruction, in which the responsibility is placed on the student to understand what errors have been made and how the instructor corrected them (e.g., Lightbown & Spada, 2013), the type of corrective feedback provided through the use of technology, and made use of in CAPT, removes the responsibility from the language learner and provides feedback in a more concrete manner, such as audiovisual training demonstrating significant improvement in the production of French nasal vowels (Inceoglu, 2016). Additionally, Zhao et al. (2013) investigated the use of exaggerated speech, presented in audiovisual form, as a form of phonetic corrective feedback to English language learners. In their study, auditory features were exaggerated, such as increasing pitch level and contours, adding length to the target phoneme, inserting pauses before and after the target phoneme, and including visual features such as increasing the range of movements of the articulatory organs but slowing the articulatory movements around the target sound. In doing so, they found a significant difference between the learners' perceptual abilities since the group provided the exaggerated stimuli outperformed those given non-exaggerated stimuli.

Of interest to the present study is that there was one class of sounds that worsened under the treatment implemented in Zhao et al. (2013): they found that stop minimal pairs (such as *pat* [ˈp^{h}æt] and *bat* [ˈbæt], *toe* [ˈt^{h}oʊ] and *doe* [ˈdoʊ], and *cane* [ˈk^{h}eɪn] and *gain* [ˈgeɪn]) worsened with this particular treatment. They attributed this result to the fact that the articulation of stops is difficult to exaggerate, as almost all of the articulatory gestures happen internally, meaning that the movement involved in producing the sound is not visible from the outside. For example, it is possible to observe lip rounding (an external gesture) when producing certain vowels, but it is not possible to observe tongue position or height (both internal gestures) since they occur within the vocal tract. It is the same for stops; it is possible to see lip closure (though that only pertains to /p/ and /b/), but overall it is not possible to observe tongue position or occlusion to distinguish between /t d/ and /k g/.

In contrast to the more visually noticeable sounds investigated in Zhao et al. (2013), Offerman and Olson (2016) investigated the implementation of a system of visual feedback for the instruction of VOT in the Spanish voiceless stops, /p t k/, in the L2 classroom. Between four elicitation tasks, the participants received explicit instruction about each of the three stops, one at a time, and were provided feedback in the form of waveform and spectrogram images of their own speech, which they then compared to native speaker images (with guidance). Afterward, they compared their recording to that of the native speaker, and practiced and re-recorded themselves. They found that, overall, after receiving and interacting with visual feedback, there was significant improvement in the participants' pronunciation, with the VOT values moving toward those accepted as native VOT values.

The findings of Offerman and Olson (2016) suggest that VOT can be impacted by pronunciation training and feedback in adult L2 learners. The amount of feedback provided in their experiment was extensive and detailed; conversely, the feedback provided by app utilized in the present study identifies utterances as either "correct" or "incorrect". Since this is an intermediate step between the acoustic feedback provided in Offerman and Olson (2016) and not providing feedback at all, if the present study finds that this binary type of feedback produces similar results, it could indicate that a small amount of feedback can positively impact L2 pronunciation, even in segmental features.

2.3.1.2 Automatic Speech Recognition (ASR). As more research explores the use of technology in pronunciation acquisition, automatic speech recognition (ASR) has become more mainstream due to the popularity of personal assistants such as Siri and Alexa. However, it is by no means flawless, nor is it being used to its full potential in terms of language learning. This section explores the potential impact that ASR could have on second language acquisition, and more specifically, pronunciation training. Kolesnikova (2017) investigated error identification and correction that could negatively impact the efficacy of intelligent tutor systems in two languages with similar consonant inventories: English and Spanish. Specifically, she investigated the acquisition of American English consonants by L1 speakers of Mexican Spanish. She states that "it is important for a human teacher or an intelligent tutor model to successfully perform the task of providing relevant feedback by identifying errors in the learners' speech, explaining the causes of such errors, and offering adequate corrective exercises" (Kolesnikova, 2017, p. 197); despite the importance of these elements, error detection and correction with relevant feedback, however, is an area where technology, such as ASR, still falters. Kolesnikova (2017) attempts to address this by proposing error patterns which could then be incorporated into computer-assisted pronunciation technology in order to better identify and address learner errors. First, from a theoretical standpoint, she classified each consonant in each of the languages by its features, identified any allophones (including their contexts), and noted the similarities and differences between the American English and Mexican Spanish inventories (every consonant in both languages was included in the study). Based on her comparisons of the two inventories, she then proposed three error patterns at the phonemic level: substitution, insertion, and deletion. From these, she proposed adding transcriptions of probable erroneous pronunciations to the phonetic databases that already exist in intelligent tutor and ASR systems.

Though theoretical in nature, the findings of Kolesnikova (2017) are highly relevant to the present study, as it investigates the same sounds, except Spanish is the L2, while American English is the L1, and it explores the efficacy of ASR. The suggestions made by Kolesnikova (2017) pertaining to the augmentation of the phonetic databases with likely mispronunciations have merit, as learners could greatly benefit from a system that recognizes errors and provides feedback on them, but systems such as intelligent tutors and ASR can only work with what they are "taught."

As studies such as Kolesnikova (2017) have explained, ASR provides feedback to learners without the requirement of a human being present to do so, allowing students to practice meaningfully on their own. Continuing with this theme, McCrocklin (2016) investigated the use of ASR for pronunciation training from the perspective of learner autonomy since ASR can potentially provide the L2 learner with feedback on his or her pronunciation independent of an instructor or other human entity. Within an ESL listening course, participants took part in a pronunciation workshop for three weeks, and were placed into one of three groups: the control group, which received only face-to-face instruction; one experimental group, which received mostly face-to-face instruction, but with a small amount of supplementation with strategy training; and another experimental group, which received a hybrid treatment of strategy training and interaction with the ASR software. Overall, she found that using the ASR software increased participants' feelings of autonomy; additionally, the presence of feedback within the ASR feature was one of the main reasons the participants favored that specific technology. Despite these positive findings, McCrocklin (2016) does acknowledge the fact that frustration likely had an impact on her findings, as ASR is still developing, and the program she required the participants to use was less developed than other forms of ASR such as Siri or Google Voice. While the use of ASR in McCrocklin (2016) still occurred within the context of the classroom environment, the sense of autonomy reflected in her findings presents the potential for similar results outside of the language classroom.

As the research in this section illustrates, CAPT is an innovative approach to pronunciation acquisition, as it combines phonetic principles with ever-changing and improving technology. It is also important to note that the idea behind CAPT is training, which typically focuses on a small number of problem areas and combines the multimodal presentation of feedback with targeted articulatory and/or acoustic information. As such, CAPT represents an intertwining of phonetics/phonology and technology that can benefit L2 learners in their quest to achieve native-like proficiency, or at the very least provide them with extra opportunities to train and receive feedback. This is especially true if pronunciation is not highlighted much in their language classes or they are attempting to learn the language on their own.

Since technology such as ASR relies on the information encoded into it by humans, and previous research has suggested improvements to ASR systems that incorporate phonetic and phonological information (e.g., Kolesnikova, 2017), it is important to understand the phonological systems of any language that one might wish to add to ASR databases. Additionally, considering the use of ASR specifically for foreign language pronunciation practice and/or training, it would be beneficial to understand the phonological systems of both the L1 and the L2 in question. The subsequent two sections provide phonetic, phonological, and orthographic information for consonants in both English and Spanish, the first approaching consonants from an articulatory standpoint, and the second from an acoustic standpoint.

2.4 Articulatory Phonetics

When investigating the acquisition of pronunciation, especially in combination with orthographic representation, it is important to first understand the features and functions of the sounds under investigation from the standpoint of each language, as the already-acquired language (i.e., native language, or L1) may affect, or *interfere with*, the acquisition of the L2. *L1 interference*, specifically pertaining to pronunciation, can appear as "phonological interferences (confusion arising from similar sounds regardless of spelling similarity) [or as] orthographic interferences (confusion arising from similar spellings regardless of phonological similarity)" (Yamashita, 2013, p. 54). This section describes some of the canonical features of the sound systems of English and Spanish; it is not exhaustive, but rather focuses specifically on the consonants (and their orthographic representations) that were investigated in the present study in order to provide background for comparisons and claims pertaining to L1 interference (Major, 2008; Strange & Shafer, 2008; Yamashita, 2013) and/or contrastive analysis (Cichoki et al., 1999; Leather, 1999, Odlin, 2014).

2.4.1 Phonemes and Graphemes in English and Spanish Consonant Inventories

It is generally accepted that the English phonological inventory contains 23 consonants, defined by place and manner of articulation, as well as by voicing status (Wickelgren, 1966). English consonants appear to be more straightforward in terms of allophonic variation, since much of the regional variation is seen in vowels (Clopper et al., 2005), and prosody (Clopper & Smiljanic, 2011), among others. However, the orthographic representation of English consonants often highlights inconsistencies in 1) the relationship between graphemes and phonemes, 2) the existence of duplicate referents, and 3) an unpredictability that could contribute to L1 interference in the L2.

The consonants in the Spanish phonetic inventory overlap with those in English. At the phonemic level, Spanish contains 17 consonants, also defined by voicing status as well as by place and manner of articulation (Hualde, 2014). Much research has focused on regional variation in consonants (e.g., Hualde, 2005; Lipski, 1994), though variation in vowels has been found in multilingual contexts (Alvord & Rogers, 2014; Konopka & Pierrehumbert, 2008; O'Rourke, 2010; Nadeu, 2014) and also linked to sociolinguistic variation (Lipski, 2011; see also de la Fuente Iglesias & Pérez Castillejo, 2020; Garrido, 2007). Additionally, the relationship between graphemes and phonemes in Spanish is quite transparent (Carreiras et al., 2014); in most cases, there is a one-to-one correspondence between grapheme and phoneme, with some notable exceptions (Hualde, 2014). Most of the graphemes and phonemes included in the current study exemplify cases in which the relationship is less transparent. As the focus of the present study is on the graphemes <c>, <g>, <h>, <ll>, <ñ>, <j>, <ch>, and <z> (plus the voiceless stops /p t k/), the following sections focus on these graphemes and their corresponding phonemes, presented by manner of articulation. Graphemes appear within angle brackets (< >), phonemes within slashes (/ /), and allophones within brackets ([ ]).

2.4.1.1 Stops. The English phonological inventory contains two classes of stops: the voiceless stops /p t k/ and the voiced stops /b d g/; primarily of interest to the present study is the set of voiceless stops and /g/. The English voiceless stops /p t k/ are generally considered to be aspirated, though studies have shown that the unaspirated form can be triggered by conditions such as consonant clusters, syllable stress (or rather lack thereof), and the sound immediately preceding the voiceless stop, in both inter- and intra-word conditions (Ito & Strange, 2009; Kahn, 1976; Keating et al., 1981; Kenstowicz & Kisseberth, 1979; MacKay et al., 2001; Nagle, 2018).

The Spanish sound inventory also contains two classes of stops: the voiceless bilabial, dental, and velar stops (/p/, /t/, and /k/, respectively) and the voiced bilabial, dental, and velar stops (/b/, /d/, and /g/, respectively). In Spanish, the class of voiceless stops is considered to be unaspirated, indicating that the amount of time (in milliseconds[12]) between the release of the stop and the beginning of the vowel is very short (e.g., Ladefoged, 1975; Lisker & Abramson, 1964). This is in direct contrast to the aspirated voiceless stops in English.

In both English and Spanish, the graphemes <c> and <g> have more than one phonemic association. In Spanish, the pronunciation depends on the vowel that follows the grapheme (i.e., <c{a, o, u}> are pronounced [k], while <c{i, e}> are pronounced [s]; <g{a, o, u}> are pronounced [g], while <g{i, e}> are pronounced [x]). However, in English, the grapheme-to-phoneme mapping is more complicated, as <c> can be pronounced as [k] (*cat, cot, cut*), and [s] (*certain, cinch*), which mirrors Spanish, but other pronunciations are also possible, such as [ʃ] (*special, suspicious, ocean*) and [t͡ʃ] (*cello*). As for <g>, in English, the grapheme <g> can occur with the [g] pronunciation followed by every vowel in the English vowel inventory, as in *gear* [ˈgiɹ], *give* [ˈgɪv], *gate* [ˈgeɪt], *get* [ˈgɛt], *gaffe* [ˈgæf], *gall* [ˈgaɫ], *gum* [ˈgʌm], *ghoul* [ˈguɫ], *good* [ˈgʊd], *goat* [ˈgoʊt], *got* [ˈgɔt], *gout* [ˈgɑʊt], *guide* [ˈgɑɪd], and *indigoid* [ˈɪn.də.ˌgoɪd]. However, words like *gymnasium* [d͡ʒɪm.ˈneɪ.zi.əm], *giraffe* [d͡ʒə.ˈɹæf], and *gel* [ˈd͡ʒɛɫ] still contain the grapheme <g> but are pronounced [d͡ʒ]. The orthographic inconsistencies in English (and especially in comparison to Spanish) highlight the potential for L1 interference on L2 pronunciation. Particularly for beginners, these inconsistencies in the L1 should be considered when attempting to explain pronunciation errors in the L2 because L1 interference is plausible, especially in speakers who are just starting the process of learning a second language.

[12] See Section 2.5 for a more thorough discussion of VOT in English and Spanish.

2.4.1.2 Fricatives. Three fricatives from the English sound inventory pertain to the grapheme-phoneme mismatches under investigation in the present study: /h/, /s/, and /z/. All three are considered phonemes, as they are contrastive with other sounds (i.e., the meaning changes based on the phoneme in question, as in *hip* /ˈhɪp/, *sip* /ˈsɪp/, and *zip* /ˈzɪp/), and have a one-to-one correspondence between phoneme and allophone. It is also contrastive with the absence of sound, exemplified by the minimal pair *hit* /ˈhɪt/ and *it* /ˈɪt/. The voiced and voiceless alveolar fricatives, /s/ and /z/ respectively, both have phonemic status in English; their contrast can be seen in minimal pairs such as *sip* /ˈsɪp/ and *zip* /ˈzɪp/. They also have a one-to-one correspondence with their respective allophones, [s] and [z] (ignoring the potential for voicing assimilation), as well as with their graphemes <s> and <z>. It is important to recognize the status of /s/ and /z/ as separate phonemes in English because this separation, similar to the distinction between /h/ and the absence of sound, potentially plays a role in the acquisition of pronunciation when the L2 treats these sounds differently (which Spanish does).

The Spanish fricatives investigated in the present study are the voiceless alveolar fricative, /s/, and voiceless velar fricative, /x/, and the voiced palatal fricative, /ʝ/. The phoneme /s/ corresponds to the graphemes <z>, as in *azul* [a.ˈsul] 'blue', and <c> when it is followed by <i> or <e>, as in *cita* [ˈsi.ta] 'appointment' and *hice* [ˈi.se][13] 'I did' (1SG-PRET)[14]. The phoneme /x/ corresponds to <j> as in *ajo* 'garlic' [ˈa.xo], as well as to <g> when it is followed by <i> or <e>, as in *lógico* [ˈlo.xi.ko] 'logic, logical' and *gemelo* [xe.ˈme.lo] 'twin'. The phoneme /ʝ/ corresponds to the grapheme <ll>, as in *llamo* 'I call' (1SG-PRS) [ˈʝa.mo] and *calle* 'street'

[13] or to the interdental fricative [θ], depending on regional variation; the present study assumes a Latin American dialect inventory.
[14] All glosses in this document follow the Leipzig Glossing Rules (The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses, 2015).

[ˈka.ɟe][15]. Although the grapheme <z> appears in Spanish, the phoneme /z/ does not; there are instances of voicing assimilation, such as in consonant clusters, in which the allophone [z] is phonetically possible (Hualde et al., 2010), but overall, the grapheme <z> typically corresponds to the voiceless alveolar fricative [s], rather than the voiced alveolar fricative [z]. As has been discussed previously, the graphemes <ci> and <ce> correspond to [si][16] and [se][17], respectively.

As for <ll>, depending on the region, the surface realization of /ɟ/ can be: the voiced palatal fricative, [ʝ]; the voiced pre-palatal fricative, [ʒ]; or the voiced palatal lateral, [ʎ][18] (Hualde, 2014). It is important to note that the voiced alveolar lateral approximant, /l/, exists in the phonetic inventory as well; it corresponds to the grapheme <l> (but never <ll>). Similar to the explanation for <z>, from the standpoint of an L1 English speaker, the phoneme corresponding to <ll> in Spanish is not the expected one (from the perspective of the L1), but rather [l], creating a situation in which L1 interference is likely.

2.4.1.3 Affricates. English has two affricate phonemes, /t͡ʃ/ and /d͡ʒ/, voiceless and voiced pre-palatal affricates, respectively. The former corresponds to the grapheme <ch>, and the latter corresponds to <j> (and sometimes also to <g>), but both of these graphemes correspond to more than one phoneme. The digraph <ch> appears in the spelling of *chip* [ˈt͡ʃɪp], *echo* [ˈɛ.koʊ], *machine* [mə.ˈʃin], and *yacht* [ˈjat], representing /t͡ʃ/, /k/, /ʃ/, and Ø, respectively. The affricate /d͡ʒ/ appears in the pronunciation of words written with both <j> and <g>, as in *just* [ˈd͡ʒʌst] and *germ* [ˈd͡ʒɚm]. This correspondence is not the case in other languages, Spanish

[15] The present study recognizes that other phonemes, such as /ʎ/, also correspond to <ll> depending on the region (Hualde, 2014).
[16] or [θi], depending on regional variation
[17] or [θe], depending on regional variation
[18] The voiced palatal lateral has phonemic status as /ʎ/ in some varieties of Spanish (Hualde et al., 2010).

included, and should be considered when investigating orthographic issues in more orthographically transparent languages from the standpoint of acquisition of pronunciation.

In contrast to English, the mapping of the voiceless pre-palatal affricate /t͡ʃ/ in Spanish, whose graphemic representation is <ch>, appears more straightforward, though allophonic variation can be found depending on the region. For example, /t͡ʃ/ can be realized as: the voiceless pre-palatal fricative, [ʃ]; the voiceless palatal stop, [c], or its voiced counterpart, [ɟ]; or the voiceless alveolar affricate, [t͡s], in addition to [t͡ʃ] (e.g., Hualde et al., 2010; Hualde, 2014). Despite these potential regional allophones, generally the mapping of <ch> is more transparent in Spanish than in English. However, the difference in transparency between the two languages creates a situation in which interference may be found, particularly for L1 English speakers.

2.4.1.4 Laterals and Approximants. In English, the digraph <ll>, which is more likely to be considered two <l> graphemes in a row, corresponds to the phoneme /l/, the voiced alveolar lateral approximant. In keeping with the lack of orthographic transparency in English, <l> and <ll> both correspond to same phoneme (/l/), as in the words *allot* [ə.ˈlɑt] and *lot* [ˈlɑt]. However, as has been mentioned previously[19], this is not the case for Spanish, in which the grapheme <l> corresponds to /l/, but the predominant pronunciation in Latin American dialects and in many Peninsular dialects for the grapheme <ll> is the voiced palatal fricative, /ʝ/, which does not appear in the English phonological inventory (Odden, 2005, p. 148). The lack of distinction between one <l> and two in one language but not the other has great potential to interfere in the acquisition of pronunciation, particularly when the two graphemes correspond to different phonemes.

[19] See Section 2.4.1.2

It is important to note that the phoneme /j/, the voiced palatal approximant (Delahunty & Garvey, 2010), exists in English, though it is associated with the grapheme <y>. Regardless, in terms of the adaptation of a non-native sound into English, it would be considered the closest phoneme (Paradis & LaCharité, 1997) to the voiced palatal fricative [ʝ]. The adaptation of /ʝ/ as [j] also would help to avoid the potential for confusion or miscommunication due to minimal pairs between /ʝ/ and /l/ in Spanish (e.g., *llamo* 'I call' (1SG-PRS) [ˈʝa.mo] and *lamo* 'I lick' (1SG-PRS) [ˈla.mo]). However, for an L1 English speaker, a strong influence of orthography is probable, as both <l> and <ll> map to /l/ (but never /ʝ/) in English, which would make <ll> susceptible to L1 interference.

2.4.1.5 Nasals. The only nasal formally investigated in the present study is the Spanish palatal nasal /ɲ/, associated with the grapheme <ñ>; the palatal nasal does not form part of the English phonological inventory (Odden, 2005, p. 148). However, similar to the situation regarding the adaptation of /ʝ/ as [j], English does contain [j] which, when pronounced in a sequence following [n], would represent the palatal quality of the nasal /ɲ/, exemplified by words such as *canyon* [ˈkʰæn.jən]. Adapting /ɲ/ as [n.j], (i.e., splitting the palatal nasal into a palatal and a nasal), illustrates a process referred to as *nativization*[20] (Fitt, 1998, p. 144), which could arguably be considered the least costly transformation that also preserves the most phonological information (Paradis & LaCharité, 1997). Additionally, adapting /ɲ/ as [n.j] would avoid issues with minimal pairs and dramatic shifts in meaning (e.g., *caña* 'cane' [ˈka.ɲa] and *cana* 'grey hair' [ˈka.na]), which would likely cause a lapse in communicability.

Despite these issues, from an orthographic standpoint, there is the potential for recognition of the grapheme <ñ> in English due to the prevalence of Spanish last names

[20] This appears using the British English spelling as *nativisation* in the original text.

containing said grapheme, and also due to its presence in loanwords, such as *jalapeño*[21], which has been accepted into the language as a whole entity despite <ñ> not being included in the graphemic inventory of English. There are even cases in which it has been erroneously overextended or hypercorrected, which in general refers to the application (i.e., extension) of a given linguistic rule onto something or into somewhere it does not belong (Schwegler & Kempff, 2007, p. 363). Pertaining to borrowed words, however, these kinds of overextensions are often referred to as *hyperforeignisms*, a term which describes "cases in which loanwords receive features not present in the original word, [but] added because they are associated with the source language by the recipient speakers" (Simonović, 2015, p. 29). As examples of hyperforeignisms, Moquin and Salmons (2020) describe "the non-pronunciation of /s/ in *coup de grace* on the awareness that French has orthographic-final consonants that are not pronounced, and the introduction of <ñ> into the Spanish borrowing *habanero*" (p. 5), creating the erroneous spelling, **habañero*.

The graphemic representation in Spanish of the palatal nasal /ɲ/ (<ñ>) illustrates a difference between English and Spanish orthography that, to an L1 speaker of English, could appear to be caused by a diacritic, in this case the tilde (~). However, nothing about the diacritic itself should indicate that the sound to which the grapheme <ñ> is mapped ought to be palatal. In fact, in some languages, the tilde (~) over a character can be used to represent vowel nasality, as in Portuguese (Müller Pograjc & Markič, 2017), glottalization in Vietnamese (Mixdorff et al., 2003), nunation in Arabic (Habash et al., 2012), and smooth rising accent in Lithuanian (Norkevičius et al., 2005). In Spanish, as well as in Basque (Hualde, 2003), Tagalog (Aspillera & Hernandez, 2014), Galician (Regueira, 1996), Guaraní (Estigarribia, 2017), and Quechua

[21] (Oxford University Press, n.d.-f)

(Noble & Lacasa, 2010), among others, <ñ> as a whole is considered its own letter (Daniels & Bright, 1996); this is also the case for the grapheme <õ> in Estonian (Zabrodskaja, 2007) and Guaraní (Estigarribia, 2017), as well as for the officially recognized graphemes representing nasal vowels in Guaraní in general (Estigarribia, 2017).

2.4.2 Grapheme-Phoneme Mismatches

Grapheme-phoneme mismatch, or polyphony[22], refers to instances in which there is more than one phoneme corresponding to a grapheme, causing the grapheme to be ambiguous in terms of how it should be conceptualized and subsequently pronounced (Timmer & Schiller, 2012). English and Spanish differ in the manner in which their graphemes correspond to phonemes, Spanish being more transparent (or shallow) than English. Due to both inter- and intra-language polyphony, language learners often find pronunciation to be problematic when presented with the written form of a word containing a grapheme that maps to more than one phoneme. In this particular case, as the L1 (English) is the less transparent of the two languages, an overabundance of choice can hinder proper pronunciation in a more transparent language, such as Spanish; additionally, this is exacerbated in cases in which the canonical pronunciation in the L2 is not part of the phonetic inventory of the L1.

2.4.2.1 Orthographic Issues. There are a number of areas in which learners' instincts in their L1 have a greater likelihood of leading them astray in the pronunciation of sounds in the L2. In this case, the issue is compounded by the orthographic representation, since "[alphabetic languages greatly differ in terms of the degree of their orthographic transparency or consistency [. . .] [and] Spanish [. . .] is considered to be one of the most orthographically transparent languages" (Carreiras et al., 2014, p. 806). English, on the other hand, cannot be considered

---

[22] (Oxford University Press, n.d.-j, definition 3)

orthographically transparent because in many cases its graphemes can be mapped to more than one phoneme. Combined with knowledge of the English and Spanish phonological systems, along with the idiosyncrasies of English orthography, it is possible to predict where there are likely to be pronunciation issues when these languages come into contact, keeping in mind that English and Spanish share identical orthographic representations, save for a small number of diacritics. This section first provides an explanation of English orthographic opacity in contrast to the transparency of Spanish, followed by an overview of each of the graphemes under consideration in the current study from the perspective of L1 interference (i.e., predictable problems in Spanish when considering English as the L1).

As has been mentioned previously, English is generally considered to be orthographically opaque (Ellis & Hooper, 2002; Timmer & Schiller, 2012). One of the most famous examples of English orthographic opacity was mentioned by George Bernard Shaw, who stated that *ghoti* could be pronounced as *fish* [ˈfɪʃ] due to the vagaries of English spelling (e.g., <gh> as in *enough* [ə.ˈnʌ**f**], <o> as in *women* [ˈw**ɪ**.mən], <ti> as in *nation* [ˈneɪ.**ʃ**ən]) (Gee, 1999); it has since been pointed out that a second (and likely tongue-in-cheek) interpretation exists in which *ghoti* is pronounced completely silently, as <gh>, <o>, <t>, and <i> all also occur in the spelling of words in which they are not pronounced (e.g., *light* [ˈlɑɪt], *people* [ˈpi.pl̩], *valet* [væ.ˈleɪ], *business* [ˈbɪz.nəs]) (Pereltsvaig, 2011). Conversely, Spanish is considered transparent, as one-to-one correspondences between the written and spoken forms are much more frequent (Carreiras et al., 2014; Hualde, 2014; Meschyan & Hernandez, 2005), but it is not completely transparent. Describing the function of <u> in conjunction with <g> and <q>, Goikoetxea (2006) states that:

> *u* does not represent any phoneme in two cases: when *g* and *q* are followed by *e* or *I* (i.e., *gue, gui,* and *que, qui*). In the case of *g*, mute *u* is used to keep the /g/ sound, in which case it is just an orthographic convention that then opens up the door to the use of dieresis in order to be pronounced[23]. In sum, *u* mute cannot be counted as a vowel (or context-dependant [*sic*] vowel [. . .]) for purposes of syllabification or accentuation. (p. 334)

Examples of the phenomenon described by Goikoetxea (2006) are words such as *guerra* ‘war’ [ˈge.ra], *guitarra* ‘guitar’ [gi.ˈta.ra] , *qué* ‘what’ [ˈke], and *química* ‘chemistry’[24] [ˈki.mi.ka] in which there is some opacity due to the presence of <u>. Aside from instances such as these, Spanish has relatively few instances in which there is greater than a one-to-one correspondence between grapheme and phoneme; rather, opacity is limited to a small group of consonant graphemes (Ijalba & Obler, 2015, p. 50).

Morgan (2006), however, claims that the situation is not quite as straightforward as it may appear. He brings into the discussion the idea of language contact and change, stating that even native Spanish speakers often have difficulty determining “the pronunciation of loan words, proper names, trademarks, and acronyms unfamiliar to them (and, unfortunately, unpredictable from the orthography given)” (Morgan, 2006, p. 122). Pronunciation issues such as those mentioned by Morgan (2006) are understandable, since native English speakers have issues with some of these as well (e.g., the pronunciation of the file type *gif* or the name *Xavier*, the oft heard mispronunciation of *habanero* as **habañero,* among others).

For learners, the grapheme <c> can be quite problematic, as not only is there variation in the pronunciation of words containing this grapheme in both English and Spanish, but also

[23] Goikoetxea (2006) is referring to words such as *lingüista* ‘linguist’ [liŋ.ˈgu̯is.ta] and *bilingüe* ‘bilingual’ [bi.ˈliŋ.gu̯e].

[24] also ‘chemist (M.SG or F.SG)’

because this particular grapheme does not have a phoneme that corresponds to it and only it. In English, it shares its phonemic representation with /k/, /s/, /ʃ/ and /t͡ʃ/, which also each have their own graphemic representations (<k>, <s>, <sh>, and <ch>, respectively). In Spanish, it is represented phonemically by /k/ and /s/, depending on the subsequent vowel. Thus, the differences in the manner in which graphemes relate to phonemes and vice versa in both languages may impede L1 English speakers in their pronunciation of Spanish words.

In addition to issues caused by variation type, the pronunciation of words containing the grapheme <c> by L2 learners can also be influenced by cognates, especially due to the historical influences of Latin and French on English. Ladefoged (1975) notes that historically, English words that end in <-tion> at one time were pronounced [sion], the coarticulation of [s] and [i] resulting in the [ʃ] found in the Modern English pronunciation of <-tion> (p. 236). Due to many periods of contact with Latin and/or much borrowing from French, depending on the word in question (Baugh & Cable, 2012), English and Spanish have many cognates; occasionally, these cognates have identical spellings in both languages, but often a change of some sort occurred, leaving the words in question similar enough to each other to be potentially problematic from the standpoint of L1 interference. This is the case for the words referenced by Ladefoged (1975), although his analysis did not mention the vagaries of English spelling beyond that the <t> in <-tion> was never actually pronounced (p. 236). The explanation found in Ladefoged (1975) left orthographic representations such as <-sion> and <-cion>, as in *ascension* and *suspicion*, respectively, unaccounted for but equally problematic from an L1 interference perspective, as all three of these representations are pronounced as [ʃ] in Modern English.

Similar to the issues pertaining to <c> due to English spelling, the graphemes <g> and <j> overlap in terms of potential L1 interference. From a phonetic standpoint, in English, <g>

can map to /ɡ/ as in *gall* [ˈɡɑl], *get* [ˈɡɛt], *gill* [ˈɡɪl], *gone* [ˈɡɑn], and *gull* [ˈɡʌl] (among others), but it also maps to /d͡ʒ/ as in *gym* [ˈd͡ʒɪm], *gem* [ˈd͡ʒɛm], and *gist* [ˈd͡ʒɪst]. The grapheme <j> also maps to /d͡ʒ/ as in *jam* [ˈd͡ʒæm], *jest* [ˈd͡ʒɛst], *jingle* [ˈd͡ʒɪŋɡl̩], *job* [ˈd͡ʒɑb], and *jump* [ˈd͡ʒʌmp]. Moreover, all of the phonetic representations ([ɡ] and [d͡ʒ] for <g>, [d͡ʒ] for <j>) are contrastive with /h/ in English, as in *hill* [ˈhɪl], *hem* [ˈhɛm], and *ham* [ˈhæm], among others. Therefore, the likelihood that an L1 English speaker would pronounce these as [h] (the logical adaptation[25] for /x/ which does not occur in English) is low; there is a greater likelihood, especially in the case of <g>, for said speaker to face difficulties deciding between /ɡ/ and /d͡ʒ/.

In contrast to <j> and <g>, L1 interference pertaining to the grapheme <z> is rather more straightforward: it is expected that L1 English speakers should function under the assumption that <z> maps to /z/ and is pronounced [z] (and conversely, <s> maps to /s/ and is pronounced [s]). As has been mentioned previously, place, manner, or voicing assimilation due to consonant clusters has been ignored pertaining to allophonic variation as consonant clusters are not the focus of the current study, though it is interesting to note that it is rare to find <z> surrounded by other consonants in English; *chutzpah*[26] [ˈhʊt.spə] and *patzer*[27] [ˈpʰɑt.sɚ], (borrowed from Yiddish and German, respectively) demonstrate that English can accommodate the voicing assimilation of /z/ as [s] in these words, but the words are not native to the English language. As such, one would expect that L1 English speakers would have certain expectations regarding the pronunciation corresponding to the grapheme <z> which could interfere with the acquisition of a

[25] See explanation of Paradis and LaCharité (2017) in Section 2.4.1.5
[26] (Oxford University Press, n.d.-c)
[27] (Oxford University Press, n.d.-i)

language in which <z> generally corresponds to pronunciation(s) that are not [z], and in which the production of [z], when it occurs, is often caused by processes such as voicing assimilation[28].

2.4.2.2 Reading Rules and Word Recognition. From an orthographic standpoint, especially as visual written stimuli were used in the present study, L1 reading rules or patterns may also have the potential to interfere with pronunciation. Porter (2010)[29] found that the influence of L1 English reading intuition was significantly negatively correlated with level of instruction; in other words, speakers with less exposure to the L2 (Spanish) had a greater chance of their pronunciation being influenced by L1 reading intuition, and vice versa. She found that the words that were most affected were those that ended in <e>, and that words ending in <ge> were usually pronounced [d͡ʒ]. Since Spanish contains words that end in <ge> and <je>, such as *elige* 'chooses' (3SG-PRS) [e.ˈli.xe] and *dije* 'I said' (1SG-PRET) [ˈdi.xe], the present study could potentially uncover a similar phenomenon. Due to English words ending in <gi> and <ji>, such as *magi* [ˈmeɪ.d͡ʒɑɪ] and *emoji* [ə.ˈmoʊ.d͡ʒi], both corresponding to the same pronunciation ([d͡ʒ]), their appearance at the end of words in Spanish, in words such as *tejí* 'I knitted' (1SG-PRET) [te.ˈxi] and *elegí* 'I chose' (1SG-PRS) [e.le.ˈxi], could produce similar results to those seen in Porter (2010) in response to word-final <ge>.

Extrapolating from the findings in Porter (2010), there appears to be the potential for the L2 speaker's L1 instincts pertaining to skills other than pronunciation to interfere and cause erroneous pronunciations, specifically reading. As has been mentioned previously, it is often the case in English that multiple phonemes have been mapped onto the same grapheme; <ch> is an

---

[28] It is possible for <z> to be pronounced as [z] in Spanish due to voicing assimilation that can occur when /s/ precedes /m/ as in *hazme un favor* 'do me a favor,' which can be pronounced [ˈaz.me]; similarly, /s/ can be pronounced as [z] for the same reason, as in *mismo* 'same,' which can be pronounced [ˈmiz.mo] (Hualde et al., 2010, p. 74).
[29] This study was conducted by the present author.

example of this, as words such as *parachute* [ˈpʰeɪ.ɹə.ʃut], *technology* [tʰɛk.ˈnɑ.lə.d͡ʒi], and *reach* [ˈɹit͡ʃ] all contain the same graphemes but different phonemes (/ʃ/, /k/, and /t͡ʃ/, respectively). This alternation is neither rule-governed nor prompted by phonology; therefore, it is unlikely that L1 English speakers would subconsciously attempt to apply restrictions from the L1 to the novel word from the L2. More investigation is necessary, but in a previous study, the investigator (of the present study) found that participants read the Spanish word *techo* 'roof' [ˈte.t͡ʃo] the same way they would pronounce the English word *echo* [ˈɛ.koʊ], but with [t] added to the beginning, resulting in the erroneous pronunciation of *[ˈtʰɛ.koʊ][30]. A possible interpretation of this outcome is that the participants recognized a word from the L1 (either *echo* or *tech*) within the novel word from the L2 (*techo*) and applied the pronunciation they were familiar with to the novel word, suggesting that there is a possibility for a unique type of L1 interference between English and Spanish: one based on the recognition of one word within another. While finding the word *echo* within *techo* may not be the intended use of literacy skills, the technique of finding smaller words within larger ones relates to intra-word analysis which Yamashita (2013) explains as "analysis of a word into its structural units such as phonemes or morphemes" (p. 53). Additionally, Whaley and Kibby (1981) claim that "dependence on within-word characteristics is more important for beginning reading success than is reliance on between-word analysis" (p. 315). Thus, the attempt at figuring out how to pronounce *techo* from the visual representation, but ending up with **teko* may indicate that beginning L2 learners attempt to use literacy skills they already have, but their L1 English orthographic opacity interfered.

[30] An alternate interpretation is that the participants recognized the English word *tech* inside the Spanish word *techo*.

2.5 Acoustic Phonetics

In addition to the aforementioned elements of articulatory phonetics, the present study also investigates the acquisition of L2 pronunciation from an acoustic standpoint. An acoustic analysis of speech sound provides not only a visual representation of the sound(s) in question, but also the opportunity to obtain quantifiable measurements which can then be compared within and across languages, speakers, etc. Hualde (2005) describes one of the visual representations, the spectrogram, as "a representation of the speech signal that gives us more detail by allowing us to observe the distribution of energy at different frequencies" (p. 61). The present study utilizes PRAAT (Boersma & Weenink, 2018) for spectrogram analysis of the class of voiceless stops (/p t k/), providing measurable, concrete evidence of speech production at two intervals during the acquisition process. This section defines voice onset time (VOT) as a theoretical concept, considering and explaining multiple factors that can affect realization. Subsequently, VOT is discussed pertaining specifically to English and Spanish.

2.5.1 Voice Onset Time

A phenomenon originally confined to occlusive consonants (Abramson & Whalen, 2017), VOT is defined as "the moment at which the voicing starts relative to the release of a closure" (Ladefoged, 1975, p. 124). Intertwined with VOT is the idea of aspiration. Aspiration refers to "a delay in the onset of the vibration of the [vocal] cords after the release of a preceding voiceless consonant" (Kenstowicz & Kisseberth, 1979, p. 10). The presence of aspiration results in longer VOT values, and in some languages, contrasts between aspirated and unaspirated stop consonants can be phonemic (Ladefoged, 1975).

The VOT of stop consonants can also be affected by the sounds that surround them. Typically, VOT is either measured word-initially and in a token in isolation (Dmitrieva et al., 2015; Keating et al., 1981; Kissling, 2015) or intervocalically (Antoniou et al., 2011). It is not uncommon for voiceless stops that are preceded by another consonant to lack aspiration, thus having lower VOT values in this context than they would otherwise (Kenstowicz & Kisseberth, 1979; Ladefoged, 1975). Moreover, the effect of consonant clusters on VOT values does not appear to be confined within word boundaries. From their discrimination tasks, which required participants to distinguish between [b] and [p], Repp and Lin (1991) found that while there was little difference in VOT values between tokens with word-initial voiceless stops read in isolation or within an utterance and following a vowel, there was a highly significant difference in VOT duration when the coda of the preceding syllable contained a consonant ([s] in their case, as [s] is the only consonant that English allows to precede [p] in a consonant cluster, as in *spot* [ˈspɑt] and *inspire* [ɪn.ˈspɑɪ.ɚ]).

In a different type of discrimination task, Ito and Strange (2009) manipulated word boundaries in order to investigate differences in VOT pertaining to word segmentation. Of particular relevance to the present study was their investigation of aspiration. They included stimulus pairs that differed by the syllabification of one phoneme, /s/; in all three subtypes in this category, the phoneme /s/ shifts between word boundaries, with either a Vowel /s/ Consonant{/p t k/} pattern (as in *loose tops* versus *Lou stops*), a Consonant /s/ Consonant{/p t k/} pattern (as in *keeps talking* versus *keep stalking*), or a Consonant /s/ Consonant{/p t k/} Consonant pattern (as in *cook's prints* versus *cook sprints*) (p. 2351). They found a highly statistically significant difference in VOT duration between the tokens containing /p t k/ in word-initial position and those containing /p t k/ in a consonant cluster; specifically, the VOT of the voiceless stop in the consonant cluster was significantly shorter. The findings from Ito and Strange (2009) support previous research in which VOT is much shorter when voiceless stops are part of a consonant cluster (Kenstowicz & Kisseberth, 1979; Ladefoged, 1975), but also suggests that word boundaries, especially those in which the final sound can be re-syllabified without creating phonotactically illegal word-initial consonant clusters, should be considered both in token and/or frame sentence creation and in analysis.

Stress also is a factor that can affect VOT. Using English as an example, Kenstowicz and Kisseberth (1979) state that "[t]he traditional analysis of [English] aspiration in a theory not appealing to the syllable says that voiceless stops are aspirated when followed by a stressed vowel and not preceded by *s*" (p. 256). However, that does not account for all the data, since Khan (1976) explains that even if they are in an unstressed syllable, English voiceless stops are still aspirated in word-initial position (as cited by Kenstowicz & Kisseberth, 1979). To complicate matters further, Kenstowicz and Kisseberth (1979) again cite Kahn (1976), who explains the difference in the aspiration of the /p/ in *support* in comparison to the one in *happy*. He claims that the /p/ in *support* is allowed to be aspirated because it is syllable-initial (and not in a complex onset, i.e., consonant cluster), and the syllable containing it is stressed. In contrast, the /p/ in *happy* is realized as unaspirated despite it being syllable-initial and not part of a complex onset. Khan (1976) points out that the difference between these two realizations of /p/ is the location of the stressed syllable, arguing that because the second syllable in *happy,* [pi], is unstressed, the consonant is actually considered to be part of both syllables simultaneously (as cited in Kenstowicz & Kisseberth, 1979, pp. 256–258). To control for syllable stress, a common format for tokens when measuring VOT is a disyllabic word with the voiceless stop placed word-initially (Keating et al., 1981; MacKay et al., 2001; Nagle, 2018). Other formats include using monosyllabic words, real or nonce, in which the syllable can only be stressed (Abramson & Lisker, 1973; MacKay et al., 2001; Repp & Lin, 1990).

2.5.2 Comparing English and Spanish Voice Onset Time

It is generally accepted that English primarily has aspirated voiceless stops. Considering the aforementioned spectrum of aspiration, in contrast to Spanish voiceless stops, English voiceless stops are at the opposite end of the spectrum, ranging from 90 to 100 percent aspirated (Ladefoged, 1975, pp. 261, 270). Thus, English voiceless stops are classified as "long-lag", and are typically considered to have VOT values of 30 milliseconds minimally, though often these values are much higher (Flege & Hillenbrand, 1984; Kissling, 2013; Nagle, 2018; Zampini, 2014), and vary by place of articulation, with the bilabial being the shortest and the velar the longest (Cho & Ladefoged, 1999; Zampini, 2014). Aspiration of English voiceless stops occurs in stressed syllables (Abramson & Lisker, 1973). An example of the longer length of the English voiceless aspirated bilabial stop [p^h] can be seen in Figure 5. In addition to aspirated stops, English also contains the contrasting unaspirated voiceless stops, but only under certain conditions. The main condition in which unaspirated voiceless stops can be found in English is in consonant clusters, particularly after /s/ (Kenstowicz & Kisseberth, 1979; Ladefoged, 1975). As was mentioned previously in Kahn (1976), another condition in which unaspirated voiceless stops can be found in English is in unstressed syllables, even outside of a consonant cluster (as cited in Kenstowicz & Kisseberth, 1979), as in *breaking* [ˈbɹeɪ.kɪŋ] and *echo* [ˈɛ.koʊ].

Figure 5

Spectrogram of English pour[31]

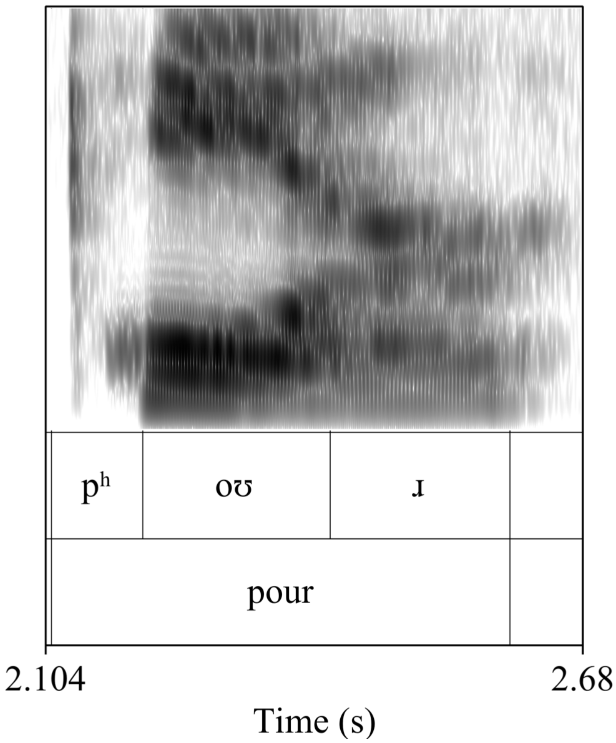
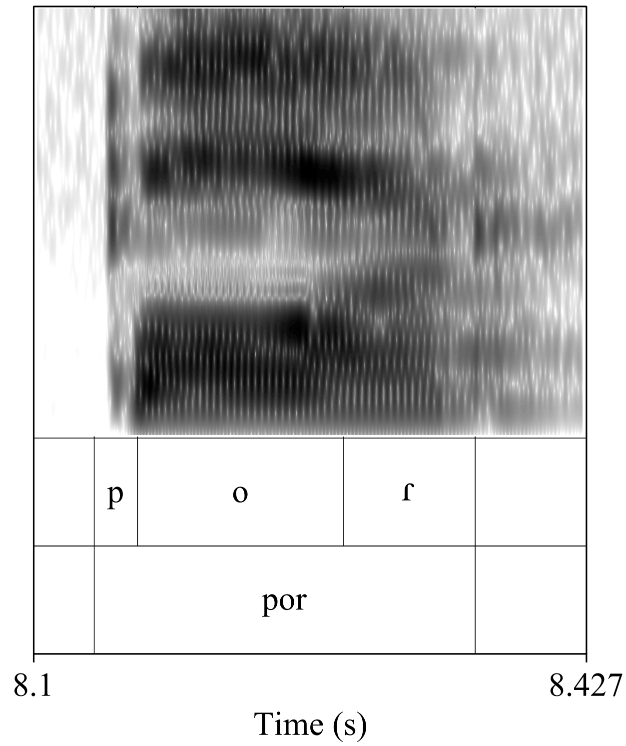


Figure 6

Spectrogram of Spanish por *'for'*[32]



The voiceless subset of stop consonants in Spanish (/p t k/) are considered to be unaspirated (Ladefoged, 1975). On the spectrum of aspiration, where an aspiration percentage of zero means that the sound is constantly voiced, Spanish voiceless stops are in the middle of the spectrum at approximately 50 percent of the total possible aspiration, which classifies them as unaspirated (Ladefoged, 1975, pp. 261, 270). These voiceless stops are considered to be "short-lag", with VOT values ranging from 4 to 29 milliseconds (Flege et al., 1994; Kissling, 2013; Lisker & Abramson, 1964; Nagle, 2018). Short-lag VOT, exemplified by the voiceless unaspirated bilabial stop [p], can be seen in Figure 6. The variation in the VOT values is due to place of articulation, with the bilabial stop having the shortest values and the velar stop having

[31] Figure 5 depicts long-lag (e.g., Lisker & Abramson, 1964)
[32] Figure 6 depicts short-lag (Ibid.)

the longest values (Cho & Ladefoged, 1999; Zampini, 2014). The VOT of Spanish voiceless stops is not affected by them being part of a consonant cluster as they are already unaspirated. Moreover, these stops are phonotactically restricted from occurring after /s/ in a consonant cluster in Spanish; when they appear after /s/, the /s/ becomes the coda of the preceding syllable, as in *España* [es.ˈpa.ɲa] 'Spain'.

Understanding the similarities and differences between languages is crucial when investigating the acquisition of VOT, as there is typically some interaction between the phonologies of the languages in question, be it a phonetic drift in which the L2 influences the L1 (Chang, 2012, 2013; Ulbrich & Ordin, 2014), or learners assigning separate sounds to the same phonetic category and no longer being able to differentiate them (Flege & Hillenbrand, 1984, pp. 716–717). Although English was not explicitly investigated in the present study, influence or interference from the L1 (English) was expected due to the participants being beginning learners of Spanish.

Both articulatory and acoustic phonetics contribute to the analysis and interpretation of the acquisition process of L2 phonology. Combined with graphemic analysis, these elements serve to explain and predict potential erroneous L2 pronunciations, particularly at beginning levels of acquisition. Due to adult SLA often taking place in a classroom and making use of a textbook and other written materials, learners are exposed to the written form of the language from the start; because these learners (typically) have little knowledge of the phonology of the language they aim to acquire, the orthographic representation has the potential to influence their pronunciation choices.

As has been mentioned previously, technological advancements have been increasingly incorporated into SLA generally, as well as for specific purposes such as the improvement of pronunciation. However, the judges who determine the utility of technology are not the creators, but the users. The present study utilized a questionnaire to investigate the opinions of the participants in reference to the technology they interacted with over the course of a semester. While there are numerous ways to interpret this kind of data, the present study elected to perform a psychometric analysis in order to investigate "social and psychological states" (Pennebaker, 2013, p. 3) in addition to the patterns and trends found in the language itself. The subsequent section describes the functionality, utility, and applications of the psychometric software, Linguistic Inquiry and Word Count (LIWC).

2.6 Linguistic Inquiry and Word Count (LIWC)

According to Aitchison (2003), "[w]ords cannot be treated as if they were a swarm of bees – a bundle of separate items attached to one another in a fairly random way. They are clearly interdependent" (p. 75). Beyond the strictly linguistic connections between words (e.g., lemmas and their derivations, closed classes, among others), life experiences can prompt more personal connections or associations between words. Thus, "[w]ords [. . .] can be thought of as powerful tools to excavate people's thoughts, feelings, motivations, and connections with others" (Pennebaker, 2013, p. xi). Given Pennebaker's (2013) explanation, that impressions of political parties (Rúas-Araújo et al., 2016; Vilares et al., 2015), product reviews (Chiranjeevi et al., 2018; del Pilar Salas-Zárate et al., 2016), identification of satire or authorship (Salas-Zárate et al., 2017; Isbister et al., 2017, respectively), and the detection of deception (Egnoto et al., 2018) are among the multitude of areas where psychometric analysis has been applied is less surprising.

LIWC is a psychometric word processing software created in order to "provide an efficient and effective method for studying the various emotional, cognitive, and structural components present in individuals' verbal and written speech samples" (Pennebaker, Boyd, et al., 2015, p. 1). Texts are uploaded and the words therein are compared to built-in dictionaries. The entries in these dictionaries are tagged with semantic meaning or weight, as well as part of speech, word function (auxiliary, negation, etc.), type (analytic, clout, authentic, tone), and category (perception, health, etc.), for example (e.g., Pennebaker, 2015; Tausczik & Pennebaker, 2010). When the words are run through the database of dictionaries, they may ping one or more categories. LIWC keeps a record of how many words correspond to each category, and the output received is both qualitative and quantitative in nature: it is qualitative in the sense that the words are categorized semantically and/or psychologically, allowing for interpretation of mood, mental state, etc. However, it is also quantitative in that every category is assigned a percentage representing how much of the text falls into that category; percentages are used due to ease of comparison across multiple texts, as there is variation in length (i.e., number of words) in any given sample (Tov et al., 2013). This is a unique feature in that it facilitates quantitative analysis of qualitative data. Due to the numerical representations, statistical analyses can be run to compare various texts. With the ability to quantify attitudes, emotion, and opinions, which are often analyzed qualitatively due to their collection methods (surveys, essays, interviews, etc.), LIWC facilitates analyses such as trends and change over time (Sell & Farreras, 2017), as well as consideration of social variables such as age, gender, and status (Pennebaker, 2017). While the majority of research using LIWC analyzes English, LIWC dictionaries have been created in other languages as well (Arroju et al., 2015; Isbister et al., 2017; Pennebaker et al., 2007).

The following section begins with an explanation of how LIWC functions in terms of analysis, as well as the interpretation of responses and trends. Subsequently, this section introduces studies exemplifying the application of LIWC in educational settings generally. It then presents LIWC research on experiences in virtual education and second/foreign language education. Finally, relevance to the current study is discussed.

2.6.1 Function and Interpretation

Before delving into the analyses LIWC can enhance, it is important to understand how LIWC works. Tausczik and Pennebaker (2010) explain in detail how LIWC functions, including generally accepted interpretations of the output. They explain that LIWC, at its core, is a word-processing program; at the most basic level, it opens files, analyzes each word in the file one at a time, and counts how many times each word appears. According to Tausczik and Pennebaker (2010), "[their] goal was to create a program that simply looked for and counted words in psychology-relevant categories across multiple text files" (p. 27). What sets LIWC apart is that it uses natural language processing (NLP)[33] to parse text files and tag the contents (e.g., words, punctuation, etc.) with a specific category (sometimes more than one) for any word found in its dictionary, either internal[34] or customized. Simultaneously, the software also counts the frequency of words pertaining to those categories. In this case, dictionary is defined as "the collection of words that define a particular category" (Tausczik & Pennebaker, 2010, p. 27). These categories are either objective, as is the case with parts of speech, or subjective; the subjective categories, such as positive and negative emotions, are those that required human judgments to determine how to categorize the words.

In order to investigate the psychological impact of words, Tausczik and Pennebaker (2010) separate words into two categories: content words and style words. Content words are those that carry meaning. Typically included in this category are open classes of words (i.e.,

[33] (Pennebaker, Boyd, et al., 2015)
[34] The LIWC2015 internal dictionary contains approximately 6,400 entries, including words, word stems, some emoticons, and punctuation (Pennebaker, Boyd, et al., 2015).

those parts of speech which allow new words to be added), whereas style words, commonly referred to as function words, are typically closed classes of words such as pronouns, prepositions, articles, and conjunctions. To further explain the rationale for investigating function words, the authors refer to Miller (1995), who states that:

> although we tend to have almost 100,000 English words in our vocabulary, only about 500 (or 0.05%) are style words. Nevertheless, style words make up about 55% of all the words we speak, hear, and read. Furthermore, content and style words tend to be processed in the brain very differently. (as cited in Tausczik & Pennebaker, 2010, p. 29)

Since style words comprise such a large amount of the words people use and encounter, it is unsurprising that they could be used to identify characteristics, trends, and implied meaning in language; Tausczik and Pennebaker (2010) state that style words provide insight into the manner in which people communicate while content words convey meaning.

In their article describing uses of LIWC, Tausczik and Pennebaker (2010) mainly focus on style (or function) words, though they often refer to content words in the context of their interaction with or correspondence to style words. Based on trends that have been confirmed in their own research and supported by studies conducted by others, the authors were able to associate psychological attributes with word type, thereby providing typical interpretations or conclusions as to the psychological meaning behind the linguistic behavior. The following table summarizes a variety of psychological elements, the word categories that correspond to them, as well as the significance or interpretation of these elements based on the word category associated with them, as described in Tausczik and Pennebaker (2010).

Table 1

Interpretation of Psychological Attributes Based on Word Type or Category[35]

Category	Type	Interpretation
Attentional focus	Pronouns	Demonstrate priorities, intentions, thoughts
	Verb tense	Temporal focus; present, future (positive), past (negative)
Emotionality	Positive correlation with pronouns, auxiliary verbs, negation	Positive emotion words indicate positive viewpoint of event.
	Negative correlation with articles, prepositions, relativity words	Negative emotion words indicate negative viewpoint of event.
Status, dominance, social hierarchy	Pronouns (first person plural)	Higher frequency correlates to higher rank
	Question marks	Fewer question marks correlate to higher rank
Honesty and deception	Larger number of exclusive words	Greater complexity Telling the truth
	More negative emotion and motion words. Fewer exclusion words. Less first person singular	Lying
Close relationships	Pronouns (first person plural)	Unrelated to higher relationship quality
	Pronouns (second person)	High usage relates to more criticism. Hostility, potential for confrontation
Thinking styles	Exclusion words	Higher usage relates to honesty
	Conjunctions	Complexity of thought (joining thoughts together) Create coherent narratives
Language complexity	Prepositions	High usage indicates communication of complex or concrete information
	Cognitive mechanisms	Complex language
	Words (> 6 letters)	
Cognitive mechanisms	Causal words	Active reappraisal of past event Organize thoughts
	Insight words	
Language use	Tentative language	Uncertainty, insecurity regarding topic
	Fillers	

As can be seen in Table 1, the majority of the word types fall into the style (functional) category; those that refer to classes of content words are often referenced in conjunction with style words.

[35] information drawn from Tausczik and Pennebaker (2010), pp. 30-37.

From a psychological standpoint, of particular interest to the current study are the first three elements: (1) attentional focus, (2) emotionality, and (3) status, dominance, and social hierarchy. As shown in Table 1, attentional focus can be demonstrated through *pronoun* usage as well as *verb tense*, and emotionality can also be gleaned from those categories (as well as others listed in Table 1) The study of political tweets conducted by Rúas-Araújo et al. (2016) exemplifies an analysis based on the associations in Table 1. Their aim was to attempt to identify strategies and/or patterns in the parties' and candidates' tweets[36] in a campaign for regional elections in Galicia. They found that there were significant differences between the Twitter usage of both the candidates and the political parties, as well as their relative success or failure in the campaign, though corresponding to different linguistic elements.

Rúas-Araújo et al. (2016) found that the same three elements showed statistically significant differences in usage within both the pool of candidates and within the assemblage of political parties: use of first person plural pronouns (e.g., *nosotros* 'we' and *nos* 'us'), verbs conjugated in present tense, and inclusive language (e.g., *con* 'with'; *y* 'and'; *incluyendo* 'including'). They associated the use of first person plural pronouns with group inclusivity (which also applies to inclusive language) and high status of the speaker. Significant differences in language use were discovered between the tweets coming from the candidates that did not occur between the tweets from the political parties, and vice-versa (Rúas-Araújo et al., 2016). The candidates differed from each other in terms of their use of third person, negations, words related to cognitive processes, tentative language, certainty, words related to social processes, and exclusive language. Among these, they associated tentative language with insecurity regarding a given topic. In terms of differences found between the parties (but not the

[36] Messages sent using the social media site/app Twitter.

candidates), Rúas-Araújo et al. (2016) found that the political parties differed in their use of negative emotion, words related to anxiety, and past tense verbs. They related low frequency of negative emotion words with positivity; words related to anxiety were also associated with insecurity, and higher use of past tense verbs was associated with putting distance between the speaker and the topic, as well as more negative feelings.

Finally, when considering the parties whose candidates were the most successful in the election, Rúas-Araújo et al. (2016) found that there was a negative correlation between the success of the candidate or political party (measured in number of votes) and the amount of tweets sent (i.e., the party that tweeted the most received the fewest votes and vice-versa). The most successful candidate/party pairing differed in their use of six-letter words and words related to social processes, which Rúas-Araújo et al. (2016) attribute to more complexity and elaboration in the candidate's speech, although the party used more concrete language, indicated by a higher use of articles. Both the party and the candidate frequently used inclusive language. The second most successful candidate/party pairing differed in their use of verbs pertaining to time orientation, with the candidate using more present tense verbs (which are associated more with positivity) while the party used more past tense verbs (which are associated with negativity). However, both the party and candidate used similar amounts of inclusive language.

Another variable Pennebaker (2013) mentions is status, as in a person's position within a hierarchy. Specifically pertaining to *I-words*, he found that people with higher status (e.g., a boss, higher military ranks) use fewer *I-words* than those with lower status (e.g., employees, lower military ranks). Pennebaker (2013) concluded that this is due to a particular factor: attention. Previously, Tausczik and Pennebaker (2010) explained that pronouns and verb tenses are strong indicators of attentional focus, and that attention sheds light on people's intentions, thoughts, and priorities. Returning to Pennebaker (2013), it is clear that this explanation fits the disparities in *I-word* usage pertaining to status because he concluded that the difference between the two is that those of higher status are paying more attention to others (such as the group they oversee). In contrast, those of lower status are more concerned with themselves, and may be more self-aware, but also more insecure, than those who have higher status.

Other situations pertaining to imbalance of power have also been shown to affect language use. Zhu et al. (2019) found that, in online course-related discussion forums, different patterns in language use emerged based on the role in the course (e.g., student, student facilitator, instructor). They found that, in comparison with the students, the student acting as facilitator used more *positive emotion* words, had a more positive *emotional tone*, and were more authoritative (akin to the instructor) than the rest of the students. Students (not in the facilitator role) showed a higher use of first person singular pronouns (pp. 215–216). These findings seem to fit with Pennebaker (2013)'s interpretation of the use of *I* pertaining to status; the students appeared to be more self-aware, as their attentional focus was aligned toward themselves.

2.6.2 LIWC Analyses of Experiences with Educational Technology

LIWC is useful as an investigation and analysis tool in situations in which there is text to analyze, especially in instances where emotion (or other sentiments) could be expressed. One situation where this often occurs is online, and particularly in more informal, quasi-conversational settings. With technology continuing to advance, research continually attempts to determine the effectiveness and the impact of the content being provided online and/or by artificial intelligence systems. Computer tutors were investigated by D'Mello and Graesser (2012), who aimed to identify student emotions during chat-style tutoring interactions. They included many LIWC categories in their investigation, but found correlations between some

LIWC categories and feelings expressed by the students about their experiences with the computer tutor. Among these, and of particular interest to the present study, was a correlation between feelings of frustration (from the student) but responses indicating uncertainty (e.g., potentially using *tentative language*) from the computer tutor. They conclude that "it takes a combination of psychological, linguistic, and cohesion features to predict student affect" (D'Mello & Graesser, 2012, p. 315), though their study did reveal that some inference could be made based on interchanges between student and computer tutor.

Recently there has been a dramatic increase in the number of online courses available, causing not only students but also teachers, administrations, and publishers to need to adapt to this new teaching platform. Andrews et al. (2009) investigated student opinions of the learning management system (LMS), Blackboard, particularly related to reactions to technological issues, finding significant differences in the uses of both *positive* and *negative emotion* in the evaluations over time. To explain an increase in both categories of emotion, they claimed that "students mainly in integrated courses with greater Blackboard experience are active adaptors but feel negative emotions in having to deal with system uncertainty" (Andrews et al., 2009, p. 52). Additionally, they found that average use of *negative emotion* words and past experiences with systems such as Blackboard were significant predictors of behavior when required to adapt due to technical issues (Andrews et al., 2009, p. 51).

Studies have also utilized LIWC to evaluate experiences in online classes. One type of class that is popular, at least in part, because the classes are free, is Massive Online Open Courses (MOOCs). MOOCs have been investigated from the standpoint of CALL, but CALL investigates the manner in which the content is delivered, rather than the language used in the delivery. Much of the sentiment analysis research conducted on MOOCs has focused on the content of the discussion boards (Almatrafi et al., 2018; Zhu et al., 2019). Almatrafi et al. (2018) analyzed written content posted on discussion boards in MOOCs in terms of the urgency of the communication. They found that urgent posts shared similar textual elements. These posts tended to contain more question marks and numbers than non-urgent posts; they also contained tentative and authentic language. Conversely, they found that there was a negative correlation between urgency when compared with clout, tone, and social language (p. 7). Additionally, Zhu et al. (2019) found a difference in language use depending on the role of the contributor; the variation they found centered around the LIWC category *positive emotion*, and summary variable *emotional tone*[37], "which is calculated as the difference between the positive emotion and negative emotion dictionaries that is then standardized" (Rocklage et al., 2018, p. 1337). Zhu et al. (2019) interpreted the more frequent use of *social* and *positive emotion* words as indicating a higher likelihood of being satisfied with the progression of the discussion (p. 221).

Related to investigations of certain features of MOOCs, Geng et al. (2020) investigated student opinions of MOOCs as a whole using the SC-LIWC dictionary (i.e., simplified Chinese); they attempted to find a connection between language use and the rates at which students actually completed their chosen MOOC given that "[d]espite the popularity and potential associated with MOOCs, the completion rates are dismal compared to that of traditional online course[s] [. . .] less than 10% on average" (Geng et al., 2020, p. 1786). The student evaluations they considered included comments regarding teaching, the platform hosting the MOOC, and the content of the course. While they claim that the overall opinions of the students were positive, *past tense*, *anxiety*, and *sadness* were among the categories that occurred most within individual reviews (Geng et al., 2020, p.1798). They interpreted the high frequency of words in these

[37] (Pennebaker, Boyd, et al., 2015

categories as areas or topics to which the students might allot more attention, but studies have shown that these features correlate with negative sentiment (e.g., Tausczik & Pennebaker, 2010), which could explain negative sentiment comprising approximately 33% of the total reviews they considered (Geng et al., 2020, pp. 1794, 1798).

In addition to students, faculty opinions regarding online instruction have been investigated. Glass (2017) explored attitudes toward online teaching via a series of interviews. Though not using LIWC directly, Glass (2017) utilized "a list of 915 affective processes developed by Tausczik and Pennebaker (2010) to identify segments of text where [faculty members] communicated emotion during the interviews" (p. 244). Glass (2017) found that emotion played a role in differences in opinion regarding online instruction, potentially being affected by "the *quality of [the participants'] own experiences* expressing subject matter and performing valued social roles in their online courses" (pp. 249–250). Since instruction in general is typically student-focused and student-oriented, analysis of faculty attitudes toward and perceptions of online instruction has the potential to improve instruction by addressing issues from the instructor's standpoint in addition to that of the students.

2.6.3 LIWC Analyses of Language (Learning) Experiences

In addition to educational technology, researchers have utilized LIWC to gain insight into students' feelings about their educational experiences, typically attempting to discover connections between language use and performance. For example, LIWC has also been used to analyze feedback. Though not specifically in an educational (or L2) setting, Bond and Pennebaker (2012) investigated feedback and pronoun usage in writing. More specifically, each time after completing a timed writing task (e.g., responding to a prompt asking about a negative experience), some were given no information regarding their use of pronouns, some were given usage statistics, and others were given detailed information about their pronoun usage and directed to focus on something or someone else during the next writing task (pp. 1015–1016). While the treatment did cause shifts in pronoun usage, it “could have changed the ways people wrote about the events but not how they were truly thinking about them” (p. 1018), meaning that their emotions relating to the events they described were not significantly impacted.

Pertaining more to educational settings, Nemec and Dintzner (2016) analyzed feedback on writing assignments by type, comparing (transcribed) audio and textual feedback, finding that the written (textual) feedback contained many more words associated with *negative emotion*, but overall affect and words associated with *positive emotion* did not show much variation (p. 157). Relatedly, Smith-Keiling et al. (2018) utilized LIWC to inform assessment, particularly with English Language Learners (ELLs), as actual knowledge and understanding of academic subjects can be obscured by language errors. They found that it was possible to use sentiment analysis (via LIWC) to read L2 writing without bias and, when applicable, give credit for the content of the written response, in spite of grammatical errors (Smith-Keiling et al., 2018).

Pertaining specifically to L2 experiences, LIWC has been used to analyze study abroad experiences, in attempts to link word usage and performance, as well as in evaluating pedagogical methods. Savicki and Price (2015) analyzed student written reflections about academic expectations, cultural expectations, and psychological issues at three intervals during study abroad: before, during, and after (p. 595). They found that, among many categories, when comparing use over time (i.e., before-during-after), differences in word usage were often significant, but varied by topic. For example, *positive emotions* and *tentative language* were among the significant categories for academic expectations, but for cultural expectations, the categories included *anger* and *word count*. For psychological issues, categories such as *positive* *emotion* and *negative emotion* (as well as *sadness* and *anxiety*) differed significantly in terms of word usage (Savicki & Price, 2015, pp. 592–593). They then compared the writing topics to each other during each time period, finding that before studying abroad, there was a significant difference in use of words in categories such as *negative emotion*, *positive emotion*, *tentative language, sadness* and *anxiety*. During study abroad, *positive emotion* and *negative emotion* (including *sadness*) were still significantly different, but usage had changed pertaining to *tentative language* and *anxiety*. After study abroad, the results were similar to those during study abroad in terms of emotion, but *tentative language* reappeared as being used significantly differently between the topics (Savicki & Price, 2015, pp. 596–597). They concluded that "each content area exhibited a different profile of cognitive processing and affective language, although the psychological issues content area showed the most distinctive configuration" (Savicki & Price, 2015, p. 598).

Also focusing on L2 experiences, Polat (2013) analyzed oral interviews (transcribed) from L2 learners studying French, investigating individual differences that could affect language acquisition. Dividing participants by course performance, Polat (2013) interpreted language use trends from a qualitative perspective. For example, the participants who performed better in class "expressed their love of French or their enjoyment of studying language" (p. 77), while the lower performing students did not; the higher achieving students also focused on specific grammar-related topics in their interviews, but the lower performers referenced learning more generally.

Continuing in this line of research, Polat (2014) added a quantitative dimension to the analysis of L2 learning experiences. This study investigated English Language Learners (ELLs), and aimed to find correlations between groups of students (categorized by their most frequent use of words) and performance on the Test of English as a Foreign Language (TOEFL). Polat (2014) designated the groupings of students as "doing" (i.e., words related to actions, events), "thinking" (i.e., words related to learning, analyzing, conditional situations), and "feeling" (i.e., words related to liking, anxiety, learning) (p. 86). When comparing these groups to the learners' respective TOEFL scores, Polat (2014) found significant differences between the "doing" group and the "feeling" group (i.e., the group that used more emotionally impactful words), with the latter performing better.

In addition to French and English, LIWC has been used to analyze opinions of Spanish language learners at the university level. Zapata and Ribota (2021) conducted an investigation into the opinions of university students regarding their experiences in university-required foreign language classes, specifically beginning and intermediate Spanish classes. The particular pedagogical methodology they aimed to evaluate utilized *identity texts*, "which can be written, spoken, visual, musical, dramatic, or multimodal combinations [of texts, i.e., products], [and] are positive statements that students make about themselves" (Cummins et al., 2005, p. 40; as cited in Zapata & Ribota, 2021, p. 5). Identity texts are typically included in the Learn-by-Design multiliteracies pedagogical approach (Zapata & Ribota, 2021). Students completed a survey containing free-response style questions after completing their identity texts; the responses were processed using LIWC to analyze the emotional content of the survey responses pertaining specifically to the students' experiences with foreign language instruction. Based on the scores calculated by LIWC, they concluded that:

> the participants' views on L2 instruction after working with identity texts were overwhelmingly positive and exhibited similarities between the [beginning] and

[intermediate] participants. The emotional tone category registered a percentage of 96.85% for [beginning] students, and 95.39% for their [intermediate] counterparts. (Zapata & Ribota, 2021, p. 11)

While their study was more qualitative in nature (i.e., they did not compare the student groups directly to each other), they did indicate that they had a large participant pool, which supports their interpretation of their findings since words indexed as *positive emotion* would have to have been consistently frequent in order for the scores to remain at such high percentages. As was mentioned previously, the score for *emotional tone* compares the sample against LIWC's dictionaries rather than against the total words in the sample (Rocklage et al., 2018; Smith-Keiling & Hyun, 2019). Zapata and Ribota (2021) also found that "the majority of students in both groups (68% NH and 75% IM)[38] expressed their wish to continue studying Spanish beyond the requirement" (p. 15), which they claim is a complete reversal from findings in previous research. While these statements came from a more qualitative analysis of their data, they do support the interpretation based on the *emotional tone* scores populated by LIWC.

Delving into the meaning encoded in the words chosen to describe or explain something provides insight into thought processes and emotions, for example, which are key pieces of information that can be used in analysis. The present study investigates participants' attitudes toward two pronunciation-related technologies (and one technology used for grammar practice only) in order to not only shed light on the use of the technologies for pronunciation training, but also to determine how the technologies were perceived by those who could potentially benefit from them.

[38] ACTFL proficiency levels Novice High (NH) and Intermediate Mid (IM) (Zapata & Ribota, 2021, p. 9)

2.7 Applications to the Present Study

2.7.1 Grapheme-Phoneme Mismatches

The possibility for interference from grapheme-phoneme mismatches indicates that the written form of the language is a significant factor in the present study, especially because participants are asked to read sentences aloud. The manner in which the stimuli were presented highlights an important difference between L1 and L2 acquisition (especially for adults): literacy. While languages exist that have a spoken form but not a written one, no language has a written form but not a spoken one (Gass & Selinker, 2008, p. 90). However, the point at which the written form of the language is learned differs between L1 and L2 acquisition, especially adult L2 acquisition, which represents the population the present study draws from. Bassetti (2008) points out two main differences between child and adult language acquisition, the first stating that "children learn to speak before they learn to read [. . .] [but] in instructed L2 learners spoken and written language can emerge at the same time" (p. 8). Thus, by the time children are beginning to read, they already have good command of the pronunciation of the language; conversely, the L2 learner not only has to contend with a new phonological system, but also the potential differences in the manner in which phonemes relate to graphemes in each language. For example, Porter (2010) found that L2 learners in second-semester Spanish classes were likely to apply English reading rules to Spanish-like nonce words when the words ended in <e>. Relatedly, Bassetti (2008) also stated that children are taught how to read, but L2 learners, particularly adults, are typically not explicitly taught how to read and write in the L2, and recommends that research further investigate the effects of orthography, especially pertaining to phonological transparency (i.e., orthographic transparency). She states that:

> [i]t is likely that native users of phonologically transparent writing systems rely on L2 orthographic input more than native users of phonologically opaque writing systems, and that learners of second languages that have phonologically transparent writing systems rely on L2 orthographic input more than learners of languages that have an opaque writing system. (Bassetti, 2008, p. 9)

Phonological, or orthographic, transparency refers to a continuum on which languages can be placed depending on the manner in which their phonemes correspond to their graphemes. To have complete transparency, every phoneme in a language would have to correspond to one and only one grapheme, and vice versa. While no language exists that has complete transparency, languages such as Spanish are considered more transparent due to having a higher number of one-to-one correspondences between grapheme and phoneme (Carreiras et al., 2014; Hualde, 2014). English, conversely, is considered more opaque as there are many instances in which one phoneme corresponds to more than one grapheme, and vice versa (Timmer & Schiller, 2012). Pertaining to the present study, because the participants, who have a more orthographically opaque L1, have been asked to read aloud in the more transparent L2, it is possible that the presence of the written form of the language during the participants' practice activities could have had an effect on accuracy over time.

Due to the potential for L1 interference, contrastive analysis greatly influenced the creation of the methodology and the selection of the stimuli; care was taken to provide instances where the L2 learner's knowledge of their L1 could transfer into the L2, causing mispronunciations or incorrect associations between the graphemes and the phonemes. The graphemes selected for investigation were chosen because they contrast between the two languages, such as the lack of pronunciation of <h> in Spanish (but the reverse in English).

Other contrasts were more subtle, such as those pertaining to <c> and <g>, since there is variation in their pronunciation in the L1 as well. The idea of contrasting two languages, especially through their respective orthographic representations, is central to the present study.

2.7.2 Acquisition of Voice Onset Time

Since voice onset time (VOT) comprises one part of the investigation in the present study, it is important to understand the typical assumptions in acquisition of L2 phonology as well as the accepted values for VOT in both English and Spanish. As has been mentioned in a previous section, VOT refers to the amount of time between the closure of the stop consonant and the beginning of the voicing of the following sound (e.g., Ladefoged, 1975). Regarding the voiceless stops /p t k/, this is an area in which English and Spanish are mutually exclusive. According to Zampini (2008), the English and Spanish voiceless stops are categorized differently according to their voice onset time. She names the English voiceless stops as long-lag, meaning that their average voice onset time is longer than 35 milliseconds , and the Spanish voiceless stops as short-lag, having an average voice onset time of 0 to 35 milliseconds (Keating, 1984; Lisker & Abramson, 1964).

As the participants in the current study were in first-semester Spanish courses, the likelihood that their VOT values will reduce enough to cross into Spanish-like values is low; however, because research has concluded that L2 learners can perceive the differences between Spanish unaspirated voiceless stops and English aspirated voiceless stops (L1 Spanish: Aliaga-García, 2007; Alves & Luchini, 2017, 2020, L1 English: Kissling, 2015; Offerman, 2020; Zampini, 1998) signifies that the L2 learners' VOT values could reduce over time, eventually producing the voiceless stops with duration somewhere between English and Spanish (Zampini, 2014; Nagle, 2019).

Additionally, the timing of the acquisition of L2 lexical stress in the acquisition process has the potential to assist in explaining patterns pertaining to the participants' VOT values. During stimuli creation, much attention was paid to ensuring that /p t k/ occurred in stressed syllables in order to ensure consistency in measurement between English and Spanish due to English VOT length differing depending on whether the consonant is in a stressed or unstressed syllable (Kenstowicz & Kisseberth, 1979). Additionally, instances where /p t k/ occurred in unstressed syllables were included in the stimuli (but not measured).

In spite of these measures, negative transfer of L1 stress patterns is expected, especially in the beginning stages of the L2 acquisition process. Thus, although every effort was made to control for stress, participants did show influence from L1 stress patterns occasionally shifting the intended stress to a different syllable, which (depending on the word) may or may not have contained a voiceless stop. Although this knowledge may not have much direct bearing on the conclusions that may be drawn from the findings of the present study, it was taken into account during the coding and analysis of the data. The original purpose of controlling so strictly for lexical stress was to ensure an intra-group comparison, in that stressed syllable VOT was only compared to other instances of stressed syllable VOT (and vice versa for the unstressed VOT). However, when coding the data, not only was it be noted whether or not the token in question is in a stressed syllable in Spanish, but also whether or not the participant stressed that syllable in their production of the word. That way, those productions of VOT pertaining to stressed syllables that ended up unstressed did not affect the data from stressed syllables, and vice versa if the syllable the stress moved to contains /p t k/ in the desired conditions (no preceding stop, etc.).

Pertaining to instruction, while previous studies have utilized explicit instruction to attempt to modify the students' VOT values (González López & Counselman, 2013; Kissling, 2013; among others), the only instruction given to the participants in the present study was either a) in the form of binary correct/incorrect feedback from the speech recognition feature (i.e., they are instructed to try again due to making an error), or b) in the form of instructions prior to beginning an online pronunciation activity through the participants' homework site for their Spanish classes. If explicit instruction is not the only method useful in improving VOT, it might be more palatable to embrace phonetics and pronunciation as part of the language acquisition process, and therefore foreign language classroom, since there would be less threat to (or incursion into) class time.

2.7.3 Linguistic Inquiry and Word Count

The present study utilized LIWC to assist in determining whether there were differences in the language used by participants who interacted with different technologies over the course of a semester. Some of the aforementioned studies are directly relatable to the present study. For example, the findings of Andrews et al. (2009) provide insight into student opinions of Blackboard. As the Control Group (CG) in the present study completed their grammar-related activities through this platform, evidence that both *positive* and *negative emotion* were important factors in the evaluation of Blackboard is highly relevant.

Additionally, the interpretation of the investigation into the link between TOEFL scores and language use from Polat (2014) could be applicable to the present study. While the present study did not consider course grades or other scholastic measurements, it was designed to measure improvement over time. For example, given that the coding of the grapheme-phoneme mismatches was binary, that data could be interpreted as a "test" of sorts, particularly if all the graphemes were analyzed jointly per participant. Additionally, Polat (2014) found that language use, and particularly emotional language use, positively impacted the TOEFL scores of the participants. Emotional language (i.e., *positive emotion, negative emotion, affect, emotional tone*) was investigated in the present study. Finally, the present study implemented a related method of analysis that Polat (2014) mentioned as a possible future direction. According to Polat (2014), "multiple regression analysis could be used to see which semantic categories are most predictive of L2 performance" (p. 103). The present study utilized linear regression to analyze the VOT data, as well as a related type of regression (logistic) that interprets findings based on odds rather than by making predictions for the grapheme-phoneme mismatch data analysis.

Finally, the investigation by Zapata and Ribota (2021) into attitudes toward L2 instruction utilizing a specific pedagogical approach has clear implications for the present study. First, the population from which participants were recruited (adult, university students in Spanish courses) was identical. Second, while not precisely a new or different pedagogical approach, the present study did make use of new and/or under-utilized technology to facilitate pronunciation practice and potential improvement. Finally, both studies analyzed responses to open-ended survey questions using LIWC. Given all the similarities between the two studies, the conclusions drawn in the present study added to existing research on the utility of evaluating L2 opinions of instruction using a sentiment analysis tool such as LIWC.

METHODOLOGY

This study takes a primarily clinical approach to pronunciation acquisition and improvement, focusing on phonological differences between English and Spanish, and the interference that not only can occur therein but also can be compounded by the presence of orthographic representation. The idea of improvement is analyzed through the lens of pronunciation training, specifically focusing on the use of technology. In effect, the ideas of *training* and *practice* are closely related, given that repetition of an activity (or similar activities), which would constitute *practice*, with regularity can *train* learners in whatever topic they practice. *Instruction*, on the other hand, differs greatly from *training.* Olson (2014) highlights a key difference; he states that "repetition and modeling [form] the bulk of pronunciation instruction, while speech analysis technology is entirely absent. Such a system puts the burden of pronunciation training entirely on the instructor, while denying students the tools to analyze their own productions" (p. 59). *Instruction*, according to Olson (2014), is entirely teacher centered, while *training* places the onus on students to participate actively in improving their pronunciation by taking advantage of available resources.

Technological advances, such as speech recognition and speech analysis, have made it possible for students to *train* on their own. It is this idea that the present study addresses: *can students' pronunciation improve over the course of a semester if they autonomously train using speech recognition software that does provide some corrective feedback? How will this compare to the pronunciation of students who train using more typical* listen-and-repeat *style activities (which do not provide corrective feedback)?* To accompany this investigation, the final element of the study is the evaluation of and attitudes toward the technology itself. Conclusions regarding learners' attitudes toward technology vary, but learner comments have indicated that the use of technology provided a sense of independence or autonomy (Golonka et al., 2014; McCrocklin, 2016), as well as improved confidence and motivation (Golonka et al., 2014). All of these affect how learners view the technology, shaping their opinions of it. The present study aims to analyze these attitudes from a psychometric standpoint.

This chapter presents the methodology of the present study. It begins with a description of the phonemes (i.e., sounds) and graphemes (i.e., letters) under investigation, also mentioning rationales for excluding certain graphemes, after which the voice onset time (VOT) for English and Spanish voiceless stops is discussed. Then, the method of conducting the experiment is described, including stimuli, instruments, equipment, participants, and experimental procedure. Subsequently, the coding, marking, and measuring for the phonetic data is explained, along with rationales for exclusion criteria and the use of canonical pronunciation. The data extraction and methodological concerns pertaining to the Technology Questionnaire are also included.

3.1 Mismatches: Investigated Graphemes

Eight graphemes (i.e., letters), designated here by open angle brackets, were selected as the focus of this study due to differences in pronunciation between English and Spanish (when applicable): <c>, <g>, <h>, <ll> <ñ>, <j>, <ch>, and <z>. However, beyond differences in pronunciation, the presence of many of these graphemes often misinforms pronunciation, something that can compound the interference between the L1 and the L2. Consider, for example, Table 2, which exemplifies the overabundance of English sounds that correspond to the graphemes investigated in the present study[39].

[39] Excluding <p> and <t>, which were investigated specifically for the analysis of voice onset time.

Table 2

Mapping of Graphemes to Phonemes and Allophones in English

Grapheme	Phoneme	Allophone(s)	Example	Transcription
<ca>			‘cat’	[ˈkʰæt]
<co>	/k/	[k]	‘cold’	[ˈkʰoʊld]
<cu>			‘curt’	[ˈkʰɚt]
<ci>	/k/	[k]	‘arcing’	[ˈɑɹ.kʰɪŋ][40]
	/s/	[s]	‘cinch’	[ˈsɪnt͡ʃ]
	/ʃ/	[ʃ]	‘official’	[ə.ˈfɪ.ʃl]
	/t͡ʃ/	[t͡ʃ]	‘provincial’	[pɹə.ˈvɪn.t͡ʃəl][41]
<ce>	/k/	[k]	‘arced’	[ˈɑɹkt][42]
	/s/	[s]	‘cell’	[ˈsɛl]
	/ʃ/	[ʃ]	‘ocean’	[ˈoʊ.ʃən]
	/t͡ʃ/	[t͡ʃ]	‘cello’	[ˈt͡ʃɛl.loʊ]
<ga>			‘gain’	[ˈɡeɪn]
<go>	/ɡ/	[ɡ]	‘ago’	[ə.ˈɡoʊ]
<gu>			‘gum’	[ˈɡʌm]
<gi>	/ɡ/	[ɡ]	‘give’	[ˈɡɪv]
	/d͡ʒ/	[d͡ʒ]	‘giant’	[ˈd͡ʒɑɪ.ənt]
	/ʒ/	[ʒ]	‘regime’	[ɹə.ˈʒim]
<ge>	/ɡ/	[ɡ]	‘get’	[ˈɡɛt] / [ˈɡɪt]
	/d͡ʒ/	[d͡ʒ]	‘gel’	[ˈd͡ʒɛl]
	/ʒ/	[ʒ]	‘genre’	[ˈʒɑn.ɹə]
<h>	/h/	[h]	‘high’	[ˈhɑɪ]
	Ø	Ø	‘honor’	[ˈɑ.nɚ]
<ll>	/l/	[l]	‘spell’	[ˈspɛl]
<ñ>	n/a	[n.j]	‘jalapeño’	[ˌhæ.lə.ˈpʰeɪn.joʊ]
<j>	/d͡ʒ/	[d͡ʒ]	‘just’	[ˈd͡ʒʌst]
	/h/	[h]	‘jalapeño’	[ˌhæ.lə.ˈpʰeɪn.joʊ]
	/j/	[j]	‘hallelujah	[ˌhæ.lə.ˈlu.jə]
<ch>	/t͡ʃ/	[t͡ʃ]	‘choose’	[ˈt͡ʃuz]
	/ʃ/	[ʃ]	‘machine’	[mə.ˈʃin]
	/k/	[k]	‘tech’	[ˈtɛk]
	Ø	Ø	‘yacht’	[ˈjɑt]
<z>	/z/	[z]	‘zoo’	[ˈzu]
	/ʒ/	[ʒ]	‘azure’	[ˈæ.ʒɚ]

[40] Accepted pronunciation for inflections on *arc* (Oxford University Press, n.d.-b)
[41] Accepted pronunciation for *provincial* (Oxford University Press, n.d.-k)
[42] Accepted pronunciation for inflections on *arc*. (Oxford University Press, n.d.-a)

Though some of the allophones[43] in Table 1 exist as options due to historical borrowing from other languages, and may be rather infrequent, there are still many potential options for almost every consonant. Spanish, in contrast, has closer to a one-to-one correspondence between graphemes and phonemes, sometimes even collapsing multiple graphemes into referring only to one phoneme or allophone (e.g., <z> and <c{i, e}> both corresponding to /s/), shown in Table 3.

Table 3

Mapping of Graphemes to Phonemes and Canonical Allophones in Spanish

Grapheme	Phoneme	Allophone(s)	Example	Translation	Transcription
<ca>			‘cartas’	*letters*	[ˈkaɾ.tas]
<co>	/k/	[k]	‘coro’	*chorus*	[ˈko.ɾo]
<cu>			‘documentos’	*documents*	[do.ku.ˈmen̪.tos]
<ci>	/s/	[s]	‘cinco’	*five*	[ˈsiŋ.ko]
<ce>			‘cerca’	*near*	[ˈseɾ.ka]
<ga>			‘amiga’	*friend (fem.)*	[a.ˈmi.ɣa]
<go>	/g/	[ɡ] / [ɣ]	‘agosto’	*August*	[a.ˈɣos.to]
<gu>			‘gusto’	*pleasure*	[ˈɡus.to]
<gi>	/x/	[x]	‘girasoles’	*sunflowers*	[xi.ɾa.ˈso.les]
<ge>			‘gente’	*people*	[ˈxen̪.te]
<h>	∅	∅	‘hotel’	*hotel*	[o.ˈtel]
<ll>	/ʝ/	[ʝ]	‘talla’	*size*	[ˈta.ʝa]
<ñ>	/ɲ/	[ɲ]	‘español’	*Spanish*	[es.pa.ˈɲol]
<j>	/x/	[x]	‘hija’	*daughter*	[ˈi.xa]
<ch>	/t͡ʃ/	[t͡ʃ]	‘techo’	*roof*	[ˈte.t͡ʃo]
<z>	/s/	[z]	‘zapatos’	*shoes*	[sa.ˈpa.tos]

As can be seen in Table 3, not only are there fewer choices of allophones that canonically correspond to these graphemes in Spanish, but many compound the issue for L1 English-speakers because the sounds do not exist in English. Due to the potential for confusion regarding the mapping of phonemes to graphemes in the L1 as compared to the L2, this section describes the grapheme-phoneme mismatches investigated in the present study based on their similarities

[43] i.e., the sound(s) associated with a given phoneme, such as the change in pronunciation of the /s/ between *cats* [ˈk^{h}æts] and *dogs* [ˈdɑgz]

and/or differences when comparing English and Spanish. It also includes explanations for the exclusion of certain consonants from investigation in the present study.

3.1.1 The Graphemes <c> and <g>

The graphemes <c> and <g> represent particular pronunciation issues for native speakers of English because of a difference in phonological association between English and Spanish. For example, consider the following words: *caldo* 'broth' [ˈkal̪.ðo], *cesto* 'basket' [ˈses.to], *cinco* 'five' [ˈsiŋ.ko], *corro* 'I run' [ˈko.ro], and *curandera* 'healer' [ku.ɾan̪.ˈde.ɾa]. To begin, *caldo, corro*, and *curandera* contain the surface realization [k] of the phoneme /k/ (which corresponds to the grapheme <c>), but *cesto* and *cinco* contain the surface realization [s] of the phoneme /s/, which also corresponds to the grapheme <c>[44]. The deciding factor between the two phonemes is the subsequent vowel: the vowels /a/, /o/, and /u/ correspond to /k/, while the vowels /i/ and /e/ correspond to /s/. One sequence of graphemes likely to be affected by the L1 (English) is <ci> when it occurred prior to <on>. Due to its similarity to the *–tion* suffix in English, the <ción> sequence could provide evidence of L1 interference, and inhibit the learners' ability to recognize that <ci> corresponds to /s/ in this context.

A similar alternation (to that of <c>) between stops and fricatives can be seen in Spanish when looking at the grapheme <g>, as in words such as: *gato* 'cat' [ˈga.to], *gemelo* 'twin' [xe.ˈme.lo], *girasoles* 'sunflowers' [xi. ɾa.ˈso.les], *gol* 'goal' [ˈgol], and *igual* 'equal' [i.ˈ ɣu̯al]. That is, /a/, /o/, and /u/ correspond to the stop /g/ (and allophones [g] and the approximant [ɣ]), while /i/ and /e/ correspond to the fricative, in this case the fricative /x/ (and its allophone [x]). The pronunciation of <ge> as [ge] exemplifies the erroneous pronunciations (often attributed to L1 transfer) that the present study investigates. Specifically, it is attempting to

[44] Given the participant population and location, a (Latin) American pronunciation is assumed.

determine if using ASR and feedback can ameliorate this kind of error. In essence, the present study aims to realign grapheme-phoneme mapping away from the L1 patterns so that <ge> and <gi> are at the very least associated with a fricative ([h] is accepted as English does not contain the voiceless velar fricative [x] (Odden, 2005, p. 148)) while <g{a, o, u}>, which includes <gue> and <gui>, are mapped to [ɡ] (or [ɣ] intervocalically and in other leniting contexts).

3.1.2 The Graphemes <h>, <ll>, <ñ>, <j>, <ch>, and <z>

The other graphemes included in this study, <h>, <ll>, <ñ>, <j>, <ch>, and <z>[45], are somewhat less complicated; the surrounding sounds do not affect them in the same manner (unlike <c> and <g>). To begin, the grapheme <h> is left as a reminder of a process of loss of aspiration (*la pérdida de la aspiración*) that occurred in the evolution of the Spanish language; in Modern Spanish, <h> (when not forming part of a digraph) is not typically pronounced even though it is written (Hualde et al., 2010, p. 293)[46]. Within English itself, <h> is more frequently pronounced [h], but there are words in which <h> is not pronounced. Nevertheless, the cross-linguistic differences in pronunciation likely create a grapheme-phoneme mismatch.

Similar to <h>, <ll> is also treated differently between English and Spanish. The differences in the phonetic referents for the grapheme <ll> can be exemplified by a comparison of the words *call* [ˈkɑɫ] and *calle* 'street' [ˈka.ʝe], in which the English pronunciation of <ll> is typically the velarized alveolar lateral approximant [ɫ] due to it occurring in coda position (Turton, 2014), while in Spanish, the more canonical pronunciation of <ll> is [ʝ][47]. However, recognition of these kinds of differences between the L1 and the L2 is unlikely to be immediately

[45] The grapheme-phoneme mismatches are presented and/or discussed in this order throughout.
[46] There are some instances of regional variation in which <h> may be pronounced /h/, though this can cause confusion with <j> (Hualde et al., 2010, p. 293).
[47] This is due to being in onset position in the syllable, as palatal sounds are restricted in word-final position in Spanish (e.g., *don* 'Mr.' [ˈdon] versus *doña* 'Mrs.' [ˈdo.ɲa]) (e.g., Hualde, 2014, p. 174).

salient to first-semester language learners; thus, it is important to the present study to note that the English and Spanish realizations of <ll> contrast due to the latter containing a palatal feature when considering a canonical interpretation of its pronunciation (i.e., acknowledging but ignoring regional variation)[48].

Considering another palatal, the grapheme <ñ> (corresponding to the palatal phoneme /ɲ/ and the allophone [ɲ]) is an unusual case, as strictly speaking it does not occur in English. However, because English frequently borrows words from other languages, there are words deemed acceptable in English that contain this grapheme (such as *jalapeño*). Additionally, depending on the area of the United States under consideration, exposure to this grapheme can be quite high in other ways (e.g., last names, street signs). The palatal quality of <ñ> has been shown to occur even when the grapheme itself is not present, but the word in question has been borrowed from Spanish (i.e., hyperforeignism (Simonović, 2015; Moqin & Salmons, 2020), see Section 2.4.1.5 for further explanation), as in the common mispronunciation of *habanero* with the palatal [j] inserted erroneously after the nasal [n] *[ha.bə**n**.ˈ**j**eɪ.ɹoʊ]; this is incorrect as this word is derived from the city name La Habana, in which the second word is pronounced [a.ˈβa.na], not *[a.ˈβa.ɲa], in Spanish.

Another grapheme susceptible to L1 interference is <j>. In Spanish, <j> corresponds to [x] (as do <ge> and <gi>), as in the words *jamón* ‘ham’ [xa.ˈmon], *gesto* ‘gesture’ [ˈxes.to], and *girasol* ‘sunflower’ [xi.ɾa.ˈsol], respectively. In contrast, in English <j> corresponds to [d͡ʒ] (as do <ge> and <gi> in some cases) as in *just* [ˈd͡ʒʌst], *gender* [ˈd͡ʒɛn.dɚ], and *giant* [ˈd͡ʒɑɪ.ənt]. Pertaining to the voiceless velar fricative, /x/, it is important to note that this sound is not part of

[48] There is a palatal lateral pronunciation in some varieties of Spanish (e.g., Hualde, 2014; Lipski, 1994), but as the participants in the present study would more than likely not have been exposed to this sound, this variation is not a focus of this study.

the phonemic inventory of the target population's L1. The current study did not investigate the acquisition of non-native sounds, but rather whether or not participants were able to identify that the graphemes <j> (as well as <gi> and <ge> mentioned previously) correspond to a fricative, instead of the stop or affricate associated with them in the L1. Utterances were considered to be pronounced acceptably if the voiceless glottal fricative, [h], was produced in response to the grapheme <j> (as well as <ge> and <gi>).

Additionally, the digraph <ch>, which this study analyzed as a whole due to its correspondence to a single sound [t͡ʃ] (and not a combination of potential sounds associated with both graphemes individually), was included in the study due to responses elicited in a previous study conducted by the investigator. In the previous study, some participants pronounced words containing the grapheme <ch> as [k]; because this pronunciation appears to stem from an interference from English (which contains words such as *echo* [ˈɛ.koʊ] and *ache* [ˈeɪk]), the grapheme remained in the study. Another acknowledged pronunciation of <ch> in English is [ʃ], as in *Chicago* [ʃə.ˈkʰɑ.goʊ], *machine* [mə.ˈʃin], and *parachute* [ˈpʰeɪ.ɹə.ˌʃut]. The grapheme <ch> can also be pronounced [ʃ] (among other pronunciations not pertinent to the present study as those sounds do not occur in English) in certain varieties of Spanish, depending on the region (Hualde, 2014; Lipski, 1994), sociolinguistic variables, or simply the lenition of the affricate to a fricative in general (Hualde, 2014). However, for the purposes of this study, the canonical pronunciation [t͡ʃ] is considered the target pronunciation.

Finally, the grapheme <z>, in most varieties of Spanish, represents the phoneme /s/, and for the purposes of this study is considered to be pronounced as [s], though the present study does acknowledge that other phonetic representations of this grapheme and corresponding phoneme exist in Spanish (e.g., Hualde, 2014). However, ignoring whole borrowings, <z>

typically does not correspond to /s/ in English; as was seen in the decision map for *ajedrez* 'chess' in Figures 3 and 4 (see Chapter 1), to accommodate the L2, the learner has to add a grapheme-phoneme association instead of being able to rely on one from the L1, despite /s/ being a phoneme in English (Odden, 2005, p. 148).

3.1.3 Consonants and/or Sequences of Graphemes Not Analyzed in the Present Study

The premise of the present study is to investigate mismatches. However, some graphemes (or grapheme sequences) appear in the stimuli but were not analyzed in the present study. Specifically, these are: <gui>, <gue>, <qui>, <que>, <k>, and <rr>. For the purposes of the present study, <gui> and <gue> were not distinguished from <gu> because all three map to /g/; <g{a, o, u}> itself was included in the present study as a contrast to <g{i, e}>[49]. Since the purpose of the present study was to investigate whether students can learn to differentiate between sequences that correspond to /g/ (<ga>, <go>, <gu>) versus /x/ (which would be adapted as [h]; <gi>, <ge>), an analysis at this level of complexity could be left for future investigation. For example, based on participant responses in the present study, <gui> and <gue> may be an area to investigate in future studies from the standpoint of loanword phonology, given that [gw] was a pronunciation that often appeared in response to some of the words (e.g., *guerra* 'war') but not all (e.g., *portugués* 'Portuguese').

The grapheme sequences <qui> and <que> were also not analyzed in the present study due to difficulties in distinguishing between L1 (in the present study, English) influence and L2 pronunciation improvement. Historically, English has been influenced greatly by other languages including (but not limited to) Latin, Old French, Old Norse, Celtic, and Scandinavian (Baugh & Cable, 2013), and in Modern English there are many words containing <qui> or <que>

[49] <c{a, o, u}> was included for the same reason: to contrast with <c{i, e}>

that have been derived from Latin or Old French, but also some that were borrowed as a whole from a Latin-derived language (e.g., Spanish, French). Those borrowed from French vary in their pronunciation, as some are pronounced with [k] (e.g., *technique*[50], *clique*[51], *quiche*[52], *mannequin*[53]) but others with [kw] (e.g., *equipage*[54], *sequin*[55], *questionnaire*[56]); many whole borrowings from Spanish keep the pronunciation [k] (e.g., *mosquito*[57], *tequila*[58], *quesadilla*[59], *queso*[60]), with some exceptions pronounced as [kw] (e.g., *quixotic*[61]). Therefore, it could be difficult to determine the cause of (mis)pronunciations of <qui> and <que> because it may not be clear whether influence from loanwords in the L1 or acquisition of pronunciation in the L2 resulted in the students' pronunciation. As was the case for <gui> and <gue>, the methodology of the present study was not designed to investigate this issue, though it could be of interest to future investigations. Additionally, while <qui> and <que>, when in stressed word-initial and word-medial positions, could have been measured for the VOT analysis, there was an abundance of instances of <c{a, o, u}> in stressed word-initial and word-medial positions, and was therefore deemed unnecessary.

Similar to <qui> and <que>, the grapheme <k> (corresponding to the phoneme /k/) was not considered in the present study for reasons pertaining to loanwords. Porter (2010) points out that the grapheme <k> almost exclusively occurs in words that have been borrowed into Spanish

[50] (Oxford University Press, n.d.-s)
[51] (Oxford University Press, n.d.-d)
[52] (Oxford University Press, n.d.-o)
[53] (Oxford University Press, n.d.-g)
[54] (Oxford University Press, n.d.-e)
[55] (Oxford University Press, n.d.-q)
[56] American English pronunciation (Oxford University Press, n.d.-n)
[57] (Oxford University Press, n.d.-h)
[58] (Oxford University Press, n.d.-t)
[59] (Oxford University Press, n.d.-l)
[60] Accepted in 2017 (Oxford University Press, n.d.-m)
[61] (Oxford University Press, n.d.-p)

from another language that uses this grapheme (pp. 25–26). In addition to borrowings, <k> can also be found in technical words (Hualde et al., 2010). It is also possible that some of these loanwords are undergoing changes to make their orthographic representation more canonical; *kiosco* 'kiosk', for example, can also be found as *quiosco*, and other words derived from this noun also show this potential alternation (*kiosquero / quiosquero*) (Underwood et al., 2012, p. 233). Due to these complications, the phoneme /k/ was analyzed when it corresponded to <c{a, o, u}> in the present study.

The grapheme <rr> also appears in the stimuli but was not analyzed for two reasons. First, the sequence <rr> is canonically pronounced as the voiced alveolar trill [r] in Spanish (e.g., *guerra* 'war' [ˈge.ra]) but as the voiced alveolar approximant [ɹ] in English (e.g., *mirror*); both sounds share the same place of articulation (Pullum & Ladusaw, 1996). Therefore, it is reasonable to assume that <rr> in a non-native word would be adapted to [ɹ], given that the trill [r] does not exist in English and the adaptation to [ɹ] would take the least amount of steps to accomplish (Paradis & LaCharité, 1997). Thus, it is difficult to consider it to be a mismatch between grapheme and phoneme (especially in contrast to those included in the present study). Second, the purpose of the present study is not to determine whether students can acquire the non-native sound (i.e., the trill [r]) using pronunciation training, though that idea could be useful to investigate in the future.

3.2 Voiceless Stops (/p t k/)

The final class of sounds investigated in this study differ from the rest as there is no grapheme-phoneme mismatch present. Instead, the difference between the English and Spanish pronunciations of these sounds is found in their aspiration. Unlike the aforementioned graphemes, the voiceless stops /p t k/ are often not the focus of targeted pronunciation training, as their mispronunciation does not typically interfere with communication (i.e., it is not an error in the traditional sense), but rather is only a mark of a non-native accent. The accepted VOT values of Spanish voiceless unaspirated stops and English voiceless aspirated stops are mutually exclusive (when they occur in stressed syllables and outside of consonant clusters). Typically accepted Spanish /p t k/ VOT values range from 0 to 30 milliseconds (Benkí, 2005; Kissling 2003; Nagle, 2018; Zampini, 2014) or 35 milliseconds (Keating 1984; Lisker & Abramson, 1984; Zampini, 2008), while the accepted English VOT /p t k/ values have been described as 30+ milliseconds (Benkí, 2005; Zampini, 2014), between 58 and 80 milliseconds (Kissling, 2013; Lisker & Abramson, 1964) or between 60 and 90 milliseconds (Nagle, 2018). Due to the clear distinction in VOT duration between English and Spanish, this study also aims to investigate whether pronunciation training affects the acquisition of a sound that already exists in the native inventory but contains a different feature (i.e., aspiration).

Subsequent sections describe the procedure used to conduct the present study. First, the stimuli and their presentation are explained. Next, the instruments used by the participants are explored, followed by a description of the location in which the recording sessions took place. Then, the participant pool is described, along with the procedure for the study over the course of one semester. Finally, the manner in which the data were coded, marked, and/or measured is presented, as well as the methods for data extraction and analysis.

3.3 Stimuli

The participants were asked to read sentences shown on an iMac screen in a PowerPoint presentation. The sentences were adopted, and sometimes adapted, from a textbook designed for first-semester Spanish classes, *Vistas*, published by Vista Higher Learning (Blanco & Horwitz, 2001); the textbook used in first-semester Spanish classes at this particular southeastern

American university is published by the same entity. This book was used so that the sentences being read sounded natural and were mostly at an appropriate level for the students, especially at the beginning of the experiment. The only instances in which the sentences were modified occurred when it was necessary to add or change words to augment the number of tokens for a grapheme. The breakdown of graphemes, the corresponding numbers of tokens, and the target pronunciations can be found in Table A1 in Appendix A. Each grapheme-phoneme pairing under consideration was repeated a minimum of 30 times within the stimuli. Words containing the investigated graphemes were spread throughout the sentences that the participants read aloud.

The stimuli containing the graphemes <p>, <t>, and <c{a, o, u}> were strictly controlled in order to create similar if not identical situations for the measurement of VOT. Stimuli containing these graphemes in word-initial position were disyllabic with penultimate stress, such as *para* 'for' [ˈpa.ɾa], *tomo* 'I take' (1SG-PRS) [ˈto.mo], and *cada* 'each' [ˈka.ða]; none of those measured was followed by another consonant, nor did the previous word end with a consonant. Stimuli containing these graphemes in word-medial position, such as *repita* 'repeats' (1SG-PRS) [re.ˈpi.ta] and *apuntes* 'notes' [a.ˈpun̪.tes], were constrained in similar ways; the voiceless stop was required to be in the stressed syllable, and most of the words had primary stress on the penultimate syllable[62]. Additionally, the coda of the previous syllable could not be a consonant in order to measure VOT in environments where it was most likely to occur without the interference of other consonants; however, it is acknowledged that it is impossible to control for all interference from the participants' L1. The other graphemes investigated did not have restrictions placed upon them as, in general, their placement within words did not affect their

---

[62] For word-medial VOT, some tokens were accepted for measurement that contained primary stress on the ultimate syllable.

pronunciation. The complete list of tokens, accompanied by their translations, transcriptions, and notations regarding the grapheme(s) analyzed can be found in Appendix B.

3.4 Instruments

Three separate instruments were used in this study. The first two instruments contained pronunciation elements. The mobile application, Spanish SOLO (Distant Tribe LLC, 2016), corresponded to Experimental Group 1 (henceforth, EG1); the online (textbook) provided activities were completed by Experimental Group 2 (EG2). The Control Group (abbreviated CG) completed activities through the university's online course management platform, Blackboard Learn; these activities were grammatical in nature and contained no pronunciation element. All participants were required to interact with their assigned technology for a minimum of two hours per week for the duration of the study. Each instrument is described in detail in the following sections, and all three are summarized in Table 4.

Table 4

Summary of Instrument Features and Specifications

	EG1	EG2	CG
Software	Spanish SOLO mobile app	Online companion to course textbook	Blackboard
Operating system	iOS only	Any	Any
Accessibility	Mobile technology (e.g., iPhone, iPad) only	Smartphone/tablet browser, desktop/laptop browser	Smartphone/tablet browser, desktop/laptop browser
Cost	Free to try; earn or buy app currency to unlock more activities[63]	Online companion purchased with course textbook	N/A
Offline	Yes, if downloaded first[64]	No	No
Required technology	Microphone, speakers	Microphone, speakers	No

[63] Participants were provided $17.59 on an Apple App Store gift card to purchase sufficient app currency
[64] Distant Tribe, LLC (2016)

Recommended technology	Headphones	Headphones	No
Activity type(s)	"instructional videos, vocab lessons, multiple choice questions, conversation practice, fill-in-the-blanks, cultural information"[65]	Mainly listen-and-repeat, some listen-and-answer or listen-and-transform	Mainly cloze, some matching, occasional multiple choice
Speech modeling	Some activities	All pronunciation activities	No
Voice only completion	Most activities	All pronunciation activities	No
Focus	Fairly balanced between form and meaning	Mostly form, sometimes also meaning	Form
Pronunciation practice	Yes	Yes	No
ASR	Yes	No	No
Feedback	Yes: binary correct/incorrect with check mark or 'X' and accompanying sounds	Yes: listen to self, compare with native speaker	Yes: Correct answer provided

Included in Table 4, *speech modeling* refers to whether or not the participants received aural input from the technology that they could model their own speech after. All of the activities completed by participants in EG2 contained recordings of native speakers to which learners could compare their own speech. In contrast, the activities from the app (EG1) modeled speech for the participants in some of the activities. For example, in the vocabulary lesson, clicking on a word or phrase would play a recording that modeled the utterance for learners; subsequently, they could use the speech recognition function to pronounce the word or phrase themselves and have their speech evaluated. However, speech modeling was not available in all activities completed by participants in EG1. For instance, there was an activity that simulated a conversation; while learners were given words or phrases to choose from, they received no audio

[65] (Matte, 2019, par. 2)

input prior to being evaluated. The following sections provide more detail regarding the three instruments used in the present study (summarized in Table 4).

3.4.1 Mobile Application

The instrument for EG1 was the mobile application (henceforth referred to as *app*) Spanish SOLO. During (and prior to) the Fall 2018 semester, this app was available for free from the Apple Inc. App Store, but only for iOS devices. The app was released in 2016 and was developed by Distant Tribe LLC and was a completely self-contained app, not connected to any textbook or publisher. The app was intended to assist with solo language learning. As of approximately Summer 2019, the app had been removed from the App Store, and its website had been taken down.

The app provided a wide range of features to assist with the language learning process, though not all of them aligned with some of the current theories in second language acquisition. The app provided a coach, Sofia, who guided users through the activities, parts of the storyline, and lessons on history and culture. Sofia mainly spoke English, at least in the lessons that were available in the Beginner and Beginner II levels. In fact, the entire app was bilingual, with the elements of most activities being presented in both English and Spanish.

The app provided a variety of lessons, beginning with greetings, numbers, and the alphabet, and eventually moving the learner toward more advanced concepts, such as verbs of being (*ser* and *estar*), telling time, subject-verb agreement, and conjugation in present tense. Within every lesson, there was a variety of activities. The app began with an explanation of the concept that was the focus of that lesson, and then provided vocabulary that could be useful when communicating using the concept. If the lesson focused on grammar, the concept was explicitly taught and explained, accompanied by drill-like activities to practice the concept presented. Next, the app included activities that helped the student to further practice the concepts that were explained and to use the provided vocabulary. These were also drill-like in nature. After the learner successfully completed these activities, the app provided activities in which the focus was on communication, such as ones in which the learner had to provide the missing parts of the conversation.

A special feature included in the app was the speech recognition feature. For the majority of the activities, it was possible to practice (in the case of vocabulary) and respond using this feature. Some exceptions were word order and multiple choice activities, for example. The app's version of this feature included two elements that are expected, and one that is uncommon: the recording and playback features could be found in many available activities, but few provided feedback on the pronunciation. The feedback was binary in nature; it deemed the input it was given as either acceptable or unacceptable. However, it provided something that many of spoken practice activities lack, which illustrates the difference between recorded and speech recognition pronunciation activities: the recorded activities place the responsibility on the participant to figure out the errors, whereas the app's speech recognition function evaluated and provided feedback to the participant. The feedback was only of the correct/incorrect variety, but it allowed participants to be aware of issues with their pronunciation instead of simply guessing. In the present study, participants were instructed to attempt the utterance again if they received feedback that it was incorrect, and there was no limit to the number of times they could do so.

3.4.2 Online (Textbook-Provided) Pronunciation Activities

The instrument for EG2 was a group of pronunciation activities available to the participants through the online homework site that accompanies their course textbook, *Protagonistas* (Underwood et al., 2012). The pronunciation activities included topics such as the pronunciation of the alphabet; the letters <h>, <j>, and <g>; <ch> and <p>; <d> and <t>; <c> (before a consonant) and <q> (Underwood et al., 2012). Each chapter in the textbook was divided into sections; typically, each section contained two to three pronunciation practice activities. They were mainly "listen-and-repeat" activities; some focused on specific words, while others required repetition of a phrase or sentence; a few required a response to a question, or for the student to transform what they heard in some way. To complete a given activity, participants followed these steps after opening the activity: first, they clicked on an icon to listen to a recording of a native Spanish-speaker pronouncing the target word or phrase; next, they audio-recorded themselves reading the word or phrase aloud, attempting to replicate what they just heard. Some of the activities had additional steps (icons to click on) in which the recording played again, followed by the participant's recording; others simply stated in the instructions that the students should play both recordings and compare them (implying use of the existing icons). There was no feedback given; participants were expected to notice the differences between themselves and the native speaker and attempt to correct their speech by listening and comparing the recordings.

3.4.3 Blackboard Learn Activities

The participants in CG completed activities online weekly. These activities were completed on the university's LMS, Blackboard Learn,[66] and mirrored those that the students already were completing as part of the homework for their course. The activities they completed for this study served as extra practice, pertaining to topics covered in the activities followed along with the grammar assigned to the units in the *Protagonistas* textbook, and were adapted or

---

[66] Blackboard, when mentioned, refers to the LMS currently in use; while participant commentary may reference Blackboard specifically, these same types of issues could easily arise in another current (or future) LMS with the same functionality.

adopted from two textbooks at the same level as *Protagonistas*: *Portales* (Blanco, 2016) and *Vistas* (Blanco & Horwitz, 2001). The format of the activities varied, but made use of typical question types for which specific correct and incorrect answers could be provided. Multiple choice activities were infrequent; more often the participants completed cloze activities, where they were asked to fill in a blank in a sentence. In these cases, often the participants were given the verb they were to use and asked to follow certain directions, such as conjugating the verb in a particular tense, or they were asked to select the best verb for the sentence and then conjugate it. Additionally, some questions required the participants to match answers to each other. Regardless of the content, the activities were drill-like in nature, giving the participants extra practice with grammatical concepts. It is important to note that the activities provided to the participants in CG had no pronunciation elements attached to them.

3.5 Recording Space, Equipment, and Technology

All recordings for this experiment were conducted in a soundproof recording studio on the university campus and used the iMac present in the studio. Audacity® (iOS)[67] was used to record the participants and export the sound files in WAV signed 16-bit PCM format. The study utilized a head-worn condenser microphone, commonly used in phonetics research (Antoniou et al., 2011; Guion et al., 2000; Nagle, 2018; Olson, 2013; Simonet, 2014b; i.a.) because it controls the distance between the microphone and the speaker's mouth; the specific microphone, the AKG Pro Audio C520 L, has been used in recent phonetic studies (Chang, 2012, 2013; Huensch & Tremblay, 2015). A pop filter was added to minimize (if not eliminate) the popping sound that often accompanies stop sounds (e.g., /p t k/) in recordings. The adapter connecting the microphone to the interface was the AKG MPA VL Phantom Power adapter which had a built-in

[67] (Audacity Team, 2008)

base roll off to eliminate audible noise. A Motu 828 series Thunderball interface converted the signal from analog to digital and amplified the signal. Participants in EG1 were required to use Apple products (iPhone or iPad) in order to interact with the mobile application[68]. The online homework site provided by Vista Higher Learning in conjunction with the course textbook, *Protagonistas*, served as the platform for participants in EG2. Headphones were recommended while practicing, but neither required nor provided. Blackboard Learn, accessed through the university's system, served as the platform for CG; this group neither received audio input nor practiced pronunciation.

3.6 Participants

The participants in the experiment were recruited out of Spanish 101 (first-semester Spanish) classes at a southeastern American university during the Fall 2018 semester, beginning immediately after the deadline to add or drop classes had passed. Per the placement requirements of the Spanish Basic Language Program at this university, these participants should have studied Spanish for one or fewer years within the past four years, or never have studied Spanish at all (The University of Alabama: Department of Modern Languages and Classics, n.d.). Students in these classes were given until the end of the third week of classes to complete the Baseline Recording process (which enrolled them in the study) and were barred from enrolling in the study after this point in order to limit the amount of exposure to Spanish the participants had when recording what was considered to be their starting level of pronunciation. A total of 56 students from this set of Spanish 101 classes elected to take part in the study.

[68] Participants who did not own an Apple product but whose class section was assigned to use the Spanish SOLO app were added to CG (see Section 3.7.2 for further explanation).

3.6.1 Compensation

All participants were offered extra credit for participating in the study. The amount of extra credit offered for participation in the study consisted of an 18% increase in the students' homework score, plus a 5% bonus applied to the students' highest quiz grade if the student successfully completed all the elements of the study. Both amounts could be split over multiple scores to ensure the students received the full amount they earned. The students were allowed to earn partial credit, which was awarded based on the number of tasks completed.

3.6.2 Guidelines for Instructors

The investigator met with every instructor ($N = 8$) teaching at least one section of Spanish 101 ($N = 18$) at the beginning of the Fall 2018 semester. During this meeting, the instructors were given explicit direction regarding the manner in which pronunciation should be addressed during the semester. In an attempt to minimize the impact of differences in teaching styles, the instructors were asked not to assign recording activities (designated with a microphone symbol) as homework, nor complete them during class. Additionally, the instructors were told that while error correction (e.g., recasting[69]) was acceptable, the investigator instructed them to refrain from teaching pronunciation or focusing extra attention on pronunciation.

3.7 Procedure

3.7.1 Timeline

After the period in which students could add or drop classes had ended, recruitment for the study began. Enrollment in the study consisted of completion of the Baseline Recording (and its accompanying elements); students were allowed to enroll in the study until the end of Week 3. Recording 1 took place between three and four weeks after the Baseline Recording was

[69] (Lee & VanPatten, 2003; Lightbown & Spada, 2013)

completed (during Weeks 6 and 7). Recording 2 occurred three to four weeks after Recording 1 (during Weeks 10 and 11), and Recording 3 (and its accompanying elements) took place during Weeks 14 and 15 of the semester.

3.7.2 Baseline Recording (Pre-Test)

After setting up an appointment using the online calendar link provided to their instructors and located on their Spanish course's Blackboard site, participants arrived at the media center in the library for their appointment, where they read and signed the consent form and extra credit addendum, acknowledging that they consented to participate and fully understood how the extra credit worked. At that point, participants from 18 sections were assigned to a technology on a rotating basis, beginning with the app, then the textbook companion activities, then Blackboard Learn. This was done in order to ensure that both experimental technologies were never assigned to students in the same class section. The class section of the first participant to meet with the investigator was assigned the app (if the student did not own Apple technology, the section was still assigned to the app even though that particular participant was placed in CG). The class section of the next participant to meet with the investigator was assigned the textbook companion activities[70]. The class section of the third participant to meet with the investigator was assigned Blackboard Learn. This cycle was repeated until all class sections had been assigned a technology. However, because it was crucial that the numbers of participants in the two experimental groups were relatively equal, an escape loop was implemented; this escape loop stated that if the number of participants using the app was too low, within three enrollment days (days in which participants could be recorded) of the cutoff for participation, the cycle could be skipped in order to assign the app to more class

[70] unless the participant's class section had already been assigned a technology, in which case the participant was assigned that technology

sections. In the present study, this was done three times and resulted in very similar numbers of participants in both experimental groups. The breakdown of class sections by their assigned technology, with the number of students from each section, can be found in Appendix A, Table A2. The end result was $N = 22$ (app), $N = 20$ (textbook companion), and $N = 14$ (Blackboard).

After being assigned a technology, participants were assigned an anonymous code that they were instructed to use whenever accessing a survey for the duration of the study; this code designated their technology and their participant number to ensure anonymity while granting the investigator the ability to group survey responses by participant and/or technology. Each code began with the technology designation (SOL, VHL, BBL)[71], then the experiment abbreviation (in this case, E1), and finally assigned a number to each participant (e.g., 001, 002, 003). For the investigator's use, a code specifying the recording time period was added at the end: (B(aseline), R(ecording)1, R2, R3). During this appointment, participants were asked to fill out a background questionnaire (via the Qualtrics online survey site) designed to assess their origin, exposure to Spanish, and comfort level with technology, which can be found in Appendix E.

Once they had completed the background questionnaire, the investigator took them into the recording studio to complete the Baseline Recording, assisting them with the microphone and starting the recording in Audacity before leaving the room. This process consisted of being audio-recorded while reading Spanish sentences aloud (see Appendix C for recording instructions, Appendix D for stimuli) to the best of their ability at that point, as they had just begun learning Spanish. The sentences were presented using a PowerPoint presentation that the participants navigated through by pressing the forward arrow button on the keyboard or clicking

[71] SOL: Spanish SOLO app (EG1) | VHL: Vista Higher Learning, publisher of *Protagonistas* textbook, whose online companion provided the listen-and-repeat style activities for EG2 | BBL: Blackboard Learn (CG)

the mouse. The investigator was not present in the recording studio during this or any subsequent recording session.

After finishing the recording, participants in EG1 were shown a demonstration of how to use the Spanish SOLO app, particularly the speech recognition function, and asked to create an account. Once the account was created, the investigator provided each participant with an Apple App Store gift card in the amount of $17.59 ($15.99 plus 10% sales tax) and assisted the participants in buying *solos*, the app currency, so that the participants could move through the activities in the app at the speed the study required, as well as follow along with their Spanish class. For participants placed in EG2, the investigator demonstrated where to find and how to complete the pronunciation activities through their online homework site, emphasizing that in order to successfully complete the activities, a functional microphone and speakers were required. If they were placed in CG, they were enrolled in the Blackboard course created for the research study and then shown where to find the site and where the grammar activities were located. All were reminded that their success (or lack thereof) in no way impacted their grade in their course, but to receive the extra credit, a minimum of two hours of interaction with their assigned technology per week was required.

At this point, the participants in the two experimental groups were provided with information to assist them in successful completion of the study. All participants in these two groups were emailed a link to the weekly check-in survey, via Qualtrics (Appendix F), which would act as a journal; they were instructed to log (preferably on Saturdays or Sundays) the activities they had completed that week and reflect on their pronunciation. Each group was also provided a targeted list of activities (Appendices G and H), selected specifically to correspond to the topics being covered in their Spanish classes that week; they were not required to follow the list, but rather were advised to use it as a guide if they were not sure what to do. It is important to note that the logging of activities was not required of the participants in CG, as the built-in functionality of Blackboard Learn monitored the time spent completing activities. Finally, they were reminded that the investigator would contact them in a few weeks to set up an appointment for the subsequent recording session.

3.7.3 Recordings 1 and 2

All the participants, regardless of their group placement, were required to return for the first recording between Weeks 6 and 7 of the semester; they returned for the second recording between Weeks 10 and 11. The meetings took place in the recording studio in the media center in the library; the participants were audio-recorded reading the same set of sentences aloud in Spanish that they had read aloud for the Baseline Recording. This procedure was followed in order to maintain identical numbers of sounds, as well as identical placement. These shorter meetings also gave the investigator an opportunity to check in with the participants and answer any questions or address any concerns they had.

3.7.4 Recording 3 (Post-Test)

All the participants in the study were required to return for the third and final recording. This recording session took place between Weeks 14 and 15 in the recording studio in the media center in the library. As before, the participants were audio-recorded reading the same set of Spanish sentences aloud. Once the recording was completed, the participants were directed to a computer to complete the post-participation Technology Questionnaire (Appendix G). This brief questionnaire prompted participants to evaluate the technology they had used throughout the semester.

While the participants were completing the questionnaire, the investigator checked and confirmed their completion of the study, considering their completion of the weekly check-in surveys, as well as information from the technology they used. For those who used the app, the activity timeline (which displayed the dates on which activities were completed) was checked and logged; for those who completed the online pronunciation activities through the course's homework site, the gradebook provided information regarding their completion of the activities. Similarly, the study's Blackboard site provided information regarding the participation of those in CG. At this time, the extra credit the participants would receive was confirmed (based on what they had completed) and explained to them; the amount of extra credit each participant earned was communicated to the class instructor after the last day of classes.

3.8 Coding, Marking, and Measuring of Data

3.8.1 Rationale for Use of Canonical Pronunciation

This section explains what is meant by *canonical pronunciation* and briefly discusses consonantal allophonic variation in both languages. The term *canonical*, when used to describe pronunciation, references a pronunciation that is: (1) more aligned with the phonemic representation of the sound (Adda-Decker & Lamel, 1999; Greenberg, 1999; Saraçlar, 2001; Tsai & Lee, 2003; Wang & Lee, 2012); (2) normative, most common, most likely, or standard (Fosler-Lussier & Morgan, 1998; Greenberg, 1999; Liu et al., 2000); or (3) the form from which alternative or dialectal pronunciations are derived (Fukada & Sagisaka, 1997; Fukada et al., 1998; Goronzy et al., 2004; Liu et al., 2000; van Santen & Sproat, 1999). Only the canonical pronunciations associated with the graphemes investigated were considered for analysis because the present study recruited participants from beginning (first-semester) Spanish classes. Learners at this level would not have been exposed to much Spanish in general (Department of Modern

Languages and Classics, n.d.; Office for Academic Affairs – Testing Services, 2018), let alone to allophones that vary by country/region. More importantly, this restriction was in place because consistent use of dialect-specific allophones when reading aloud in Spanish in the Baseline Recording would have resulted in the participant's data being excluded from analysis, per the exclusion criteria. Despite the focus on canonical pronunciation of consonants, the present study acknowledges that allophonic variation exists in Spanish, represented in multiple regional varieties, and moreover is considered perfectly acceptable. Table 5 summarizes the graphemes investigated in the present study, as well as the phonemes and allophones they correspond to in both English and Spanish, considering only canonical pronunciation.

Table 5

Investigated Graphemes with Their Phonemes and Canonical Allophones in English and Spanish[72]

Grapheme	English Phoneme/s	Allophone/s	Spanish Phoneme/s	Allophone/s
<c>	/k/ ; /s/ ; /ʃ/ ; [t͡ʃ]	[k^{h}] ; [s] ; [ʃ] ; [t͡ʃ]	/k/ ; /s/	[k] ; [s]
<g>	/g/ ; /d͡ʒ/ ; /ʒ/	[ɡ] ; [d͡ʒ] ; [ʒ]	/g/ ; /x/	[ɡ], [ɣ] ; [x]
<h>	/h/ ; ∅	[h] ; ∅	∅	∅
<ll>	/l/	[l]	/ʝ/	[j]
<j>	/d͡ʒ/ ; /h/ ; /j/	[d͡ʒ] ; [h] ; [j]	/x/	[x]
<ñ>[73]	n/a		/ɲ/	[ɲ]
<ch>	/t͡ʃ/ ; /ʃ/ ; /k/ ; ∅	[t͡ʃ] ; [ʃ] ; [k] ; ∅	/t͡ʃ/	[t͡ʃ]
<z>	/z/ ; /ʒ/	[z]	/s/	[s]

As can be seen in Table 5, there are two graphemes included for which there is (or can be) no phonemic representation, as either the sound does not exist in the language, or the grapheme still exists but has no corresponding phoneme (and therefore allophone) as it is not pronounced.

[72] excluding cases of assimilation

[73] The phoneme /ɲ/, and therefore its allophone [ɲ], are not part of the English phonological inventory (Stefanich & Cabrelli, 2020, p. 2). However, due to borrowings from Spanish containing <ñ> such as *jalapeño, piña colada, piñata,* and the song, *Malagueña*, the pronunciation of <ñ> may be recognizable. For this reason, and in spite of the absence of a phoneme, the present study considers the adaptation of /ɲ/ as [n.j] to be the target pronunciation because it conserves the most features (Paradis & LaCharité, 1997) from the Spanish phoneme.

Table 5 serves to present the differences in transparency seen when comparing English orthography to Spanish. However, it also highlights the instances in which the graphemes have pronunciations in common, illustrating the decisions an L1 English speaker might make when encountering a novel Spanish word for all the investigated graphemes, as was shown in the decision trees in Figures 1 through 4.

3.8.2 Exclusion Criteria

In order to accurately compare the improvement of the participants over the course of the semester, it was important that all the participants begin the experiment with similar levels of knowledge of Spanish (ideally, quite low). Thus, it was necessary to establish criteria which would be used if it were necessary to make a determination regarding whether or not to exclude a given participant's recordings from analysis. Table A3 in Appendix A provides examples of Spanish regional consonantal allophonic variation (and the graphemes to which the allophones correspond) that served as potential exclusion criteria. Considering the Baseline Recording (i.e., pre-test), if participants had been able to consistently produce a) consonantal allophonic variation and/or b) any Spanish sound not found in the English phonological inventory, their recordings would have been excluded from the study as they most likely were too advanced for beginner Spanish. However, it was not necessary to exclude any participant from analysis due to regional variation[74].

3.8.3 Coding and Marking Mismatched Tokens

After the conclusion of the experiment, the grapheme-phoneme mismatches were coded using a binary system according to the more canonical Spanish pronunciation, 1 representing

[74] One participant was excluded for interacting with the Spanish SOLO app in addition to the Blackboard Learn activities throughout the semester; the participant reported this use of both sets of materials in one of the responses to the Technology Questionnaire.

“target” (i.e., “acceptable”) pronunciation and 0 representing “non-target” (i.e., “unacceptable”) pronunciation. For example, responding to the grapheme <h> by pronouncing *hablo* ‘I speak’ (1SG-PRS) as *[ˈha.blo] would be coded as non-target since <h> is not pronounced in Spanish; pronouncing *llamo* ‘I call’ (1SG-PRS) as *[ˈla.mo] would be coded as non-target because recognition of the palatal feature is required for the pronunciation to be coded as target. In contrast to errors such as these, a pronunciation considered “target” in the present study is either a) the canonical Spanish pronunciation or b) an adaptation of a non-English sound that preserves the most features from the original sound (Paradis & LaCharité, 1997). The word *cine* ‘theater’ [ˈsi.ne] is an example of (a); [s] is the canonical Spanish pronunciation of <ci>, and no adaptation is necessary because [s] exists in the English phonological system. In contrast, producing [ˈnin.jo] in response to the word *niño* ‘boy’ cannot be considered as an example of (a) because the voiced palatal nasal /ɲ/ is not a sound that can be found in both languages. Instead, the pronunciation [ˈnin.jo] would fall into the (b) classification because it preserves the most information from the original sound (the *voiced* alveolar *nasal* [n] plus the *voiced palatal* approximant [j]). Table 6 lists these and the other investigated graphemes, their canonical pronunciations, and the target pronunciations in this scoring system, which are considered “acceptable” because they fall into one of the aforementioned classifications.

Table 6

Canonical and Target Pronunciations of Investigated Graphemes used for Coding

Graphemes	Canonical Pronunciation	Target Pronunciation
<p>	[p]	[p], [p^{h}][75]
<t>	[t]	[t], [t^{h}]
<c> + {/a/, /o/, /u/}	[k]	[k], [k^{h}]
<c> + {/i/, /e/}	[s]	[s]
<g> + {/a/, /o/, /u/}	[g]	[g]
<g> + {/i/, /e/}	[x]	[h]
<h>	∅	∅
<ll>	[ɟ]	[j]
<ñ>	[ɲ]	[n.j]
<j>	[x]	[h]
<ch>	[t͡ʃ]	[t͡ʃ]
<z>	[s]	[s]

As can be seen in Table 6, in most cases, the canonical and target pronunciations are the same or very similar. However, there are three instances in which an adaptation was required. In the case of <ñ>, while the English phonological inventory does not contain the palatal nasal /ɲ/ (Odden, 2005, p. 148), /n/ is common to both English and Spanish, and English also contains the palatal glide /j/ (Odden, 2005, p. 35), which when accompanying /n/ provides the palatal element found in /ɲ/; the recognition of the palatal nature of /ɲ/ is important, as /n/ and /ɲ/ form minimal pairs in Spanish (e.g., *una* ‘a/one’ (F.SG) [ˈu.na] and *uña* ‘(finger/toe)nail’ [ˈu.ɲa]) and thus it is unlikely that a mispronunciation of *uña* as *una* would be comprehensible. In contrast, the case of <g{i, e}> and <j> is somewhat more complicated. The voiceless velar fricative /x/, which is the canonical pronunciation of both <j> as well as <g> when it is followed by /i/ or /e/, is not part of the phonological inventory of American English[76] (Odden, 2005, p. 148), causing the need for

[75] Producing aspiration when pronouncing the voiceless stops /p t k/ is also considered the ‘target’ pronunciation because it would be unusual for beginning learners of Spanish to fully acquire the unaspirated variant in the time allowed.
[76] This study recognizes that there are varieties of English (e.g., Scots (English)) in which the voiceless velar fricative exists (Odden, 2005, p. 26).

adaptation. Thus, assigning /x/ to the closest American English sound (Paradis, 2006; Peperkamp & Dupoux, 2003) and/or preserving the most phonological information (Paradis & LaCharité, 1997) would produce either /s/ or /h/ as options. Since /s/ is a sibilant (Ladefoged, 1975, p. 146) but /x/ and /h/ are not, /h/ would be the adaptation that preserves the most elements from the original sound, causing [h] to be considered the "target" pronunciation for [x].

To mark target or non-target pronunciation for each recording, point tiers in the acoustical analysis software PRAAT (Boersma & Weenink, 2018) were used to designate the location of the grapheme and its pronunciation; the binary code was input in the top tier of a textgrid within the program. A third point tier was used to note the participants' pronunciation when it differed from the accepted pronunciation (target or canonical), or to note anything interesting, and a fourth identified the token (i.e., word) containing the grapheme.

3.8.4 Measuring Voice Onset Time

Voice onset time (VOT) for /p t k/ was marked and measured in PRAAT (Boersma & Weenink, 2018), from the left edge of the release to "the start of periodicity in the vowel" (Foulkes et al., 2011, p. 67). VOT was measured for voiceless stops positioned both word-initially and word-medially following these guidelines. However, instances in which the voiceless stops were part of a consonant cluster, either within the word or created by the previous word ending in a consonant, were not considered in the analysis, as this has been shown to dramatically affect VOT length in the participants' L1 (Ito & Strange, 2009; Kenstowicz & Kisseberth, 1979; Ladefoged, 1975; Repp & Lin, 1991). Additionally, voiceless stops were not measured if they did not occur in the stressed syllable of the word in question due to syllable stress having an effect on the aspiration of voiceless stops in English (Khan, 1976, as cited in Kenstowicz & Kisseberth, 1979).

3.9 Data Extraction and Analysis

Out of the 56 participants who began the study, only 32 completed all four recordings. Thus, to obtain a clear picture of change over time, as well as to allow for the most time to pass between measurements, the recordings analyzed for interpretation in the present study were limited to those from participants who completed a) the Baseline Recording (and Background Questionnaire), b) Recording 3, and c) the Technology Questionnaire. The culling of participants in this manner produced the following: 13 participants in EG1, 12 participants in EG2, and seven participants in CG. However, a response from one of the CG participants written in the Technology Questionnaire revealed that this participant had also downloaded the app (used by EG1) and interacted with it in addition to the grammar activities throughout the semester; for this reason, this participant's recordings and questionnaire responses were excluded, leaving six participants in CG, and 31 in total.

Prior to scoring the responses, all the recordings were mixed together, randomized, and given unrecognizable file names in order to obscure not only which participant it was, but also which recording (Baseline, R1, R2, or R3) it was. To accomplish this, all the recordings were recoded with a random selection of words. These words were then alphabetized and assigned a number based on their alphabetic position. Last, a randomly scrambled list of numbers (Mohr, 2004) was generated pertaining to the total number of recordings; the recordings were analyzed in the order their number appeared in this randomized list. This was done in order to avoid bias in scoring, as, if the recordings were scored in chronological order, there could be an unconscious bias toward obtaining the desired result of the experiment when the investigator rated the utterances.

Python code was then utilized to extract the information from the PRAAT (Boersma & Weenink, 2018) textgrids. This was specifically accomplished using the python library praatIO (Mahrt, 2016) along with various other standard python libraries. Using commands from this library, the data for each participant were exported into Comma Separated Values (CSV) files; similar code was written for the extraction of the grapheme-phoneme data and the voice onset time data, though in the case of the latter, an extra step measured the duration of the voiceless stops. Subsequently, python code was also written to aid the analysis of the data, including (but not limited to) the compilation of data for participants by technology type, the sorting of data by consonant, and the organization of the VOT data by token position (word-initial or word-medial). Particularly in the case of the VOT data, some analysis was then completed by hand in Excel due to variation in the number of scoreable responses per recording. Statistical tests for all three analyses (i.e., grapheme-phoneme mismatches, VOT, and psychometric) were performed using R (R Core Team, 2018); R was also used to produce the graphics in the figures included in Chapters 4 and 5.

3.9.1 Technology Questionnaire

The responses to the Technology Questionnaire were exported from Qualtrics into CSV files. Due to the small number of responses, the textual responses were extracted by hand and compiled into files based on the type of technology utilized by the participants. All questions within each participant group were evaluated together (i.e., as separate entries in the same document) in LIWC due to the small number of responses. As described previously, LIWC is text analysis software (Pennebaker, Booth, et al., 2015) that was created to parse text and count the number of words that fall into various pre-defined categories. In other words, "it reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech" (Pennebaker Conglomerates, 2015, par. 1). It does this by comparing the words in any text processed through it "with a group of built-in dictionaries [. . .] [which] identifies which words are associated with which psychologically-relevant categories" (Pennebaker Conglomerates, 2015, par. 3). To gain insight into the mindsets and attitudes of the participants, their responses to the Technology Questionnaire were compiled and then tagged using LIWC.

One of the special features of LIWC is that it represents the frequency of any of its elements in the form of percentages, which allows quantitative analyses to be performed (Pennebaker et al., 2015). As such, the output from LIWC was limited to five categories (plus any sub-categories): *positive* and *negative emotions*, *temporal focus* (*focus past*, *focus present*, *focus future*), *pronouns*, and *tentative language*. *Emotional tone* and *affect* scores were also included due to the categories' relationship with the *positive* and *negative emotions* categories. Due to the small sample size, Kruskall-Wallis tests were performed on each LIWC category investigated, as well as ad hoc Dunn tests to determine where (i.e., between which participant groups) differences in psychologically-relevant word usage could be found.

Additionally, a separate analysis of technology-related words was conducted utilizing AntConc (Anthony, 2018). After ensuring that the responses of all three groups were shown concurrently, a list of technology-related words was compiled using AntConc's "Word List" tab. Then, the investigator reviewed every distinct token in the word list to determine whether it related to technology. Some words were clearly related (e.g., 'glitch,' 'app,' 'VHL,'[77] 'Blackboard') and automatically included; however, any token for which it was unclear if there was an association with technology (e.g., 'voice,' 'session,' 'access,' 'accessible') , the

[77] This is the name by which most (if not all) students and instructors refer to the online companion website to the course textbook.

investigator selected the token and was taken to AntConc's "Concordance" tab in order to view where it occurred in the response, thus determining its membership in the technology-related word category through context. Chi-Square Goodness of Fit tests were performed on the frequencies of use for the technology-related words between all three participant groups, as well as for each pairing of participant groups (i.e., EG1-EG2; EG1-CG; EG2-CG). Disparities in the use of words relating specifically to problems with technology were also discussed.

A more qualitative analysis of the questionnaire responses was also conducted. This portion of the analysis focused on various themes found throughout the responses. Specifically, these themes were: positive impressions of technology; impressions of feedback from technology; technological issues; suggestions for improvement (which subsumed negative impressions of technology); impact on performance; and rationales behind recommendations. The compiled responses from all participants in each group were read individually; any response that directly stated or implied something that fit any of the themes was pasted into a separate list.

3.9.1.1 Methodological Concerns. The aforementioned LIWC categories were selected prior to the completion of any analysis, and even before the study had been completed, in order to attempt to avoid statistical error. There is a higher likelihood of a statistically significant result being a type 1 or type 2 error because of the sheer number of variables contained in LIWC (80 language categories[78]). Too broad of an analysis risks finding statistical significance that does not represent a meaningful comparison (a type 1 error), while too narrow of a comparison may obscure significance that actually exists (a type 2 error). The risk of type 1 or 2 errors was compounded by the small sample size in the present study. Therefore, careful consideration was undertaken not only to ensure that the categories included in the analysis were feasible based on

[78] (Tausczik & Pennebaker, 2010)

the types of questions asked, but also to limit the investigation to more meaningful categories based on the situation and atmosphere in which the data were collected (a survey taken on a computer in the media center of a university library) and supported by previous research.

PHONETICS FINDINGS AND DISCUSSION

The following section describes the articulatory and acoustic phonetics findings of the present experiment, along with its corresponding discussion section. First, the articulatory phonetics findings are presented (i.e., the analyses of the grapheme-phoneme mismatches), including statistical analyses pertaining to change over time, the relationship between the dependent and independent variables, and comparisons between the three participant groups. This section is followed by the acoustic phonetics findings (i.e., the analysis of voice onset time), including statistical analyses of change over time, inter-group comparisons, and population inference from sample data. Finally, the discussion section incorporates both articulatory and acoustic phonetics, focusing on the implications for L2 acquisition (of pronunciation) and the potential impact on current technological implementation and future technological development and implementation.

Among the graphics utilized to illustrate data trends in this chapter are beanplots; specifically, pre- and post-test intra-group comparisons in Sections 4.1.1, as well as 4.2.2 through 4.2.4 are presented in this manner. According to Kampstra (2008):

> a beanplot is an alternative to the boxplot for visual comparison of univariate data between groups. In a beanplot, the individual observations are shown [. . .] in a one-dimensional scatterplot. Next to that, the estimated density of the distributions is visible and the average is shown. (p. 1)

Beanplots provide a unique way of illustrating paired data, as each side can represent one group in the pair, creating an asymmetric beanplot (Kampstra, 2008, p. 3). Additionally, a horizontal bar on each side of the beanplot represents the mean. Figures 7 and 9 through 14 utilize the asymmetric feature to provide a direct comparison between the pre-test and post-test data. However, in contrast to the description provided by Kampstra (2008), the present study does not make use of the scatterplot feature because showing all of the data points would overwhelm and obscure the beanplot.

4.1 Articulatory Phonetics Findings

The grapheme-phoneme mismatch tokens were coded as target (1) or non-target (0); the binary nature of this coding fit well with specific statistical tests. First, each participant's responses were treated as an individual test and presented as the total frequency of target pronunciations out of the total possible tokens ($N = 596$) (see Figure 7; see Appendix A, Table A4 for the corresponding table). Then, the frequency of target pronunciations was separated by consonant for both the pre-test (Baseline Recording; Table 10) and post-test (Recording 3; Table 11) responses, which allowed paired t-tests to be run in order to determine whether time was a factor in the participants' improvement overall, and which of the consonants specifically showed improvement (or lack thereof) over time. Figure 8 presents a visual of the change over time using the means for each grapheme. As time was generally found to be a factor in improvement, it is not included as part of the model in subsequent statistical analyses.

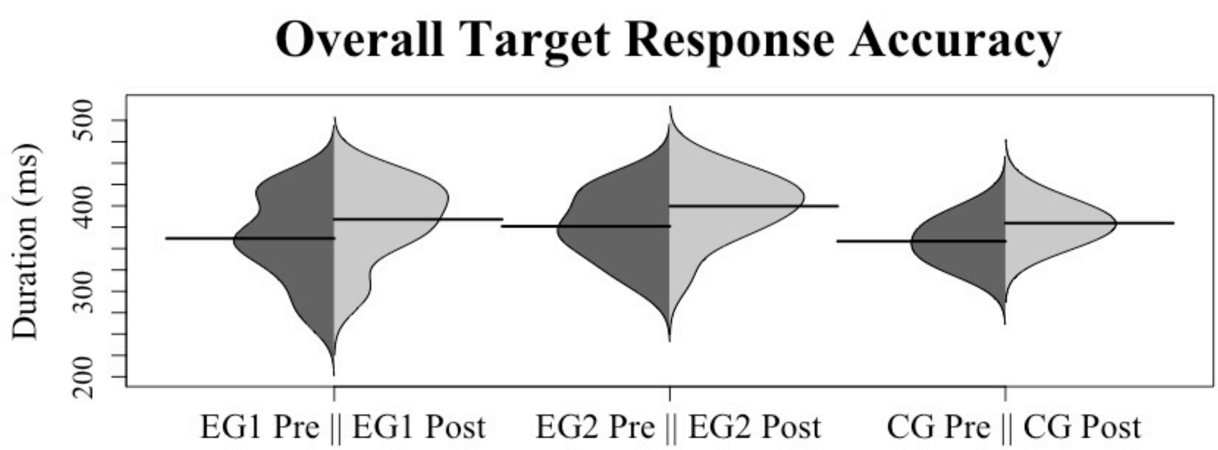
Logistic regression was selected as the statistical analysis method most fitting for the binary nature of the grapheme-phoneme mismatch data. However, prior to performing these regressions, Chi-Square tests were performed in order to determine that a relationship existed between the dependent and independent variables. Once these relationships were confirmed, logistic regressions were performed, first on each of the data from each grapheme individually. Subsequent logistic regressions were then performed on the data compiled into various groupings based on commonalities between the phonemes associated with the graphemes. Some of these comparisons were based on the Spanish phonemes, while others were based in the theory of contrastive analysis and either compared the L1 (English) to the L2 (Spanish) phonological inventories, or approached the graphemes from the phonemic perspective of the L1.

4.1.1 General Analysis over Time

Prior to other analyses, it was important to discover whether time (i.e., interacting with the technology over the course of one (1) semester) was a factor in improvement in accuracy. To accomplish this, the pre-test and post-test recordings were first treated as if the entire recording were a test; the responses to all graphemes were considered together, providing the number of target responses produced in the each recording as a whole for each participant. The aggregate pre-test and post-test accuracy totals for each participant group are represented in Figure 7.

Figure 7

Total Accuracy Distribution per Participant Group, Pre-Test to Post-Test



Paired t-tests were conducted on the accuracy raw frequency totals to determine whether time (pre-test to post-test) had an effect on accurately pronouncing problematic graphemes. Statistically significant differences were found within all three conditions, indicating general improvement in accuracy over time when comparing each participant's pre-test accuracy to their post-test accuracy. This effect was highly statistically significant for EG1[79] ($t(12) = -5.65$, $p = 1.07^{-4}$), as well as for EG2[80] ($t(11) = -4.60$, $p = 7.67^{-4}$). CG[81] also showed statistically significant improvement from pre-test to post-test, though the p-value was larger than that of the other two groups ($t(5) = -3.38$, $p = .012$).

4.1.2 Analysis by Grapheme over Time

After establishing that time was a significant factor generally, the overall accuracy scores were separated and analyzed by grapheme, or in the cases of <c> and <g>, by pattern (e.g., <c{a, o, u}> because all three are pronounced as [k]) in order to determine which graphemes in particular were affected by treatment (or grammar practice) over time. The participants' responses in the Baseline Recording (i.e., pre-test, see Table 7) were compared to those from Recording 3 (i.e., post-test, see Table 8). Then, paired t-tests were performed as the analysis compared scores from the same participants within each participant group (EG1, EG2, and Control, abbreviated CG). Table 7 presents the frequency of target responses (in comparison to the total number of tokens) for each participant group and each grapheme from the Baseline (or pre-test) recording. Table 8 presents the frequency of target responses by group and grapheme for Recording 3 (or the post-test recording).

---

[79] EG1 utilized the Spanish SOLO app that had speech recognition functionality.
[80] EG2 utilized the online listen-and-repeat style activities that accompanied their course textbook.
[81] CG did not interact with pronunciation technology, but rather completed grammar activities through Blackboard Learn course management system.

Table 7

Pre-Test Accuracy Totals, by Participant Group, by Grapheme

	<c>		<g>								
	{a,o,u}	{i,e}	{a,o,u}	{i,e}	<h>	<ll>	<ñ>	<j>	<ch>	<z>	Total
N	*1820*	*572*	*559*	*390*	*494*	*416*	*390*	*507*	*429*	*442*	*6019*
EG1	1739	521	513	121	256	285	295	423	399	169	4721
N	*1680*	*528*	*516*	*360*	*456*	*384*	*360*	*468*	*396*	*408*	*5556*
EG2	1610	489	477	120	291	310	296	394	367	159	4513
N	*840*	*264*	*258*	*180*	*228*	*192*	*180*	*234*	*198*	*204*	*2778*
CG	822	235	241	51	66	142	145	196	186	67	2151

Table 8

Post-Test Accuracy Totals, by Participant Group, by Grapheme

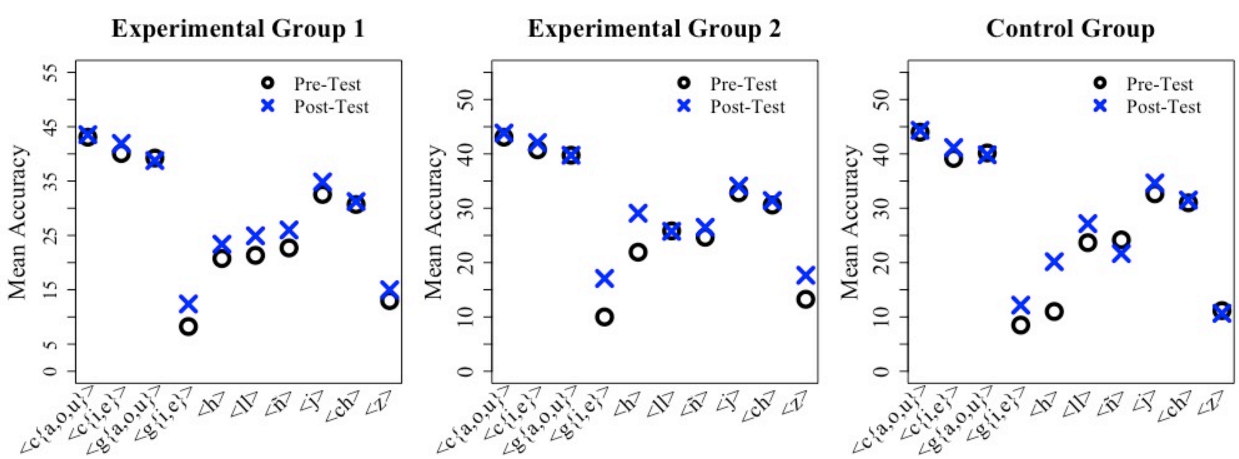
	<c>		<g>								
	{a,o,u}	{i,e}	{a,o,u}	{i,e}	<h>	<ll>	<ñ>	<j>	<ch>	<z>	Total
N	*1820*	*572*	*559*	*390*	*494*	*416*	*390*	*507*	*429*	*442*	*6019*
EG1	1763	545	505	161	305	325	338	452	408	195	4997
N	*1680*	*528*	*516*	*360*	*456*	*384*	*360*	*468*	*396*	*408*	*5556*
EG2	1620	503	471	215	352	310	319	409	378	205	4782
N	*840*	*264*	*258*	*180*	*228*	*192*	*180*	*234*	*198*	*204*	*2778*
CG	814	242	235	82	127	168	148	205	190	81	2292

The raw frequency of target responses to each grapheme were compiled for all the participants in each group. Then, paired t-tests were performed, comparing each participant's pre-test accuracy to their own post-test accuracy within each participant group to determine which graphemes showed statistically significant differences in accuracy over time. The results of the paired t-tests indicated that the pre-test and post-test accuracy was statistically significantly different for certain graphemes within certain groups. When the participants' pre-test accuracy was compared to their own post-test accuracy, EG1 demonstrated statistically significant differences over time for five graphemes: <c{i, e}> ($t = -2.84$, $df = 12$, $p = .015$), <g{i, e}> ($t = -2.65$, $df = 12$, $p =$

.021), <h> ($t = –2.21$, $df = 12$, $p = .047$), <ll> ($t = –2.51$, $df = 12$, $p = .027$), and <ñ> ($t = –3.01$, $df = 12$, $p = .011$). For EG2, there were only three graphemes that showed statistically significant differences over time: <g{i, e}> ($t = –2.99$, $df = 11$, $p = .012$), <h> ($t = –2.37$, $df = 11$, $p = .037$), and <ñ> ($t = –2.22$, $df = 11$, $p = .048$). CG, however, only showed statistically significant differences in accuracy over time for two graphemes: <c{i, e}> ($t = –2.74$, $df = 5$, $p = .041$) and <h> ($t = –4.04$, $df = 5$, $p = 9.97^{–3}$). Overall, three graphemes approached significance, which were <c{i, e}> ($t = –2.11$, $df = 11$, $p = .058$) in EG2, as well as <z> in both EG1 ($t = –1.92$, $df = 12$, $p = .080$) and EG2 ($t = –1.92$, $df = 11$, $p = .081$). Figure 8 presents a visual representation of change over time for each participant group (represented by average frequency for each grapheme).

Figure 8

Average Accuracy for Individual Graphemes, Pre-Test to Post-Test, by Participant Group[82]



As can be seen in Figure 8, for graphemes such as <j>, <ch>, and <z>, which showed no statistically significant difference over time, there is nearly a complete overlap between the black ‘o’ and blue ‘x’ symbols representing the average frequency of a target response. In contrast,

[82] Data for <c{a, o, u}> were normalized to $N = 45$ in this figure. In total, the number of tokens for <c{a, o, u}> was $N = 140$ (see Table for the number of tokens per grapheme in the stimuli)

<h> (for all three participant groups) depicts a statistically significant difference over time, illustrated by the sizeable gap between the blue and black symbols. Table 9 contains a summary of the statistical significance (or lack thereof) for the paired t-tests.

Table 9

Summary of Paired t-Test Significance by Grapheme over Time, by Participant Group

	<c>		<g>							
	<{a,o,u}>	<{i,e}>	<{a,o,u}>	<{i,e}>	<h>	<ll>	<ñ>	<j>	<ch>	<z>
EG1	nsd	*	nsd	*	*	*	*	nsd	nsd	(*)
EG2	nsd	(*)	nsd	*	*	nsd	*	nsd	nsd	(*)
CG	nsd	*	nsd	nsd	**	nsd	nsd	nsd	nsd	nsd

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

As the combination of the statistical analyses for the pre- and post-test recordings as a whole (see Section 4.1.1) as well as the analyses summarized in Table 9 generally indicate change in the participants' accuracy over time, subsequent analyses remove *time* as a variable and focus on the post-test data.

4.1.3 Relationship between Variables

Chi-Square tests were performed on the post-test data as a whole (see Table 9) to determine if there was a relationship between the categorical variables: group membership and response type. There was a highly statistically significant relationship between group membership and giving a target or non-target answer ($\chi^2(2, N = 14{,}353) = 26.64, p < 1.00^{-5}$). More specifically, the results of the Chi-Square test indicate that, overall, giving a target or non-target answer is dependent upon the treatment experienced by the participant. Subsequent Chi-Square tests were then run on data for each grapheme individually in order to determine whether the dependency between response type and treatment still exists when looking at specific graphemes. Table 10 summarizes the statistical significance of the Chi-Square tests calculated for each grapheme individually.

Table 10

Summary of Significance by Grapheme for Dependency Chi-Square Tests

	<c{a,o,u}>	<c{i,e}>	<g{a,o,u}>	<g{i,e}>	<h>	<ll>	<ñ>	<j>	<ch>	<z>
Sig.	nsd	(*)	nsd	***	***	*	nsd	nsd	nsd	*

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

As can be seen in Table 10, a dependency between treatment and response type was found for only some of the graphemes. Specifically, the graphemes <g{i, e}> ($\chi^2(2, N = 930) = 26.69$, $p < 1.00^{-5}$) and <h> ($\chi^2(2, N = 1178) = 40.39$, $p < 1.00^{-5}$) showed a highly statistically significant relationship between the two variables; additionally, a statistically significant relationship was found for the graphemes <ll> ($\chi^2(2, N = 992) = 7.51$, $p = .023$) and <z> ($\chi^2(2, N = 1054) = 6.80$, $p = .033$). The relationship approached significance for the grapheme <c{i,e}> ($\chi^2(2, N = 1364) = 5.40$, $p = .067$), while the other five graphemes showed no relationship between treatment and accuracy of response type.

Thus, the analysis considering the graphemes individually also demonstrates a statistically significant impact of time on the participants' responses, though for different graphemes in different participant groups, and not for all graphemes. Additionally, the findings regarding the relationship between the independent and dependent variables (treatment and response type, respectively) help inform the subsequent analyses comparing one participant group to the other, and each treatment group to the control.

4.1.4 Inter-Group Comparisons of Grapheme-Phoneme Mismatch Accuracy

In order to determine which graphemes were affected by which treatments, the present study utilized the logistic regression statistical test, "which examines the relationship of a binary (or dichotomous) outcome [. . .] with one or more predictors which may be either categorical or continuous" (Ranganathan et al., 2017, p. 148). The interpretation of logistic regression statistical analyses focuses on the odds of a particular outcome based on the factor(s) provided (Ranganathan et al., 2017). In the present study, the logistic regression provides the odds of a response type (target or non-target) to a particular grapheme when group membership (i.e., EG1, EG2, CG) is taken into account. These analyses were calculated using the statistical analysis software *R*. It is important to note that regression analyses compare one or more categories to a reference category, and "[b]y default R uses the alpha-numerically first category as the reference category (e.g. "a" with letters, "0" with numbers)" (Starkweather, 2018, par.1). Thus, with the coding of zero (0) and one (1), in which zero represents a non-target response and one represents a target response, zero was used as the reference category. Therefore, the results can be interpreted as higher log odds of a target response for positive values and lower log odds of a target response for negative values. Similarly, between CG and the two experimental groups, *control* occurs first alphabetically, and was therefore automatically taken to be the reference category to which each of the experimental groups is compared.

First, the regression for each grapheme was calculated separately; therefore, each regression evaluated the participants' responses to the selected grapheme (target or non-target accuracy) across the three participant groups. The output produced by these calculations contains various elements. Table 11 provides an example of the output reported for logistic regression using the output from the analysis of <c{a, o, u}>.

Table 11

Sample Logistic Regression Output

							95% C.I. for *Exp(B)*	
	B	*S.E.*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	3.44	0.20	17.29	1	$< 2^{-16}$	31.31	21.67	47.46
EG1	−0.01	0.24	-0.05	1	.960	0.99	0.61	1.57
EG2	−0.15	0.24	-0.62	1	.535	0.86	0.53	1.36

Most crucial to understanding and interpreting logistic regression output are: *Int., B, Exp(B)* and *Sig.* The abbreviation *Int.* represents the intercept, which represents the reference category to which the other groups were compared (in this case, CG); underneath it, the other independent variables are listed (in this case, the other two participant groups). *B*, also referred to as the estimate, provides "the values for the logistic regression equation for predicting the dependent variable from the independent variable [. . .] in log-odds units" (Logistic Regression, n.d., par. 36). *Exp(B)* relates to the values in the *B* column, as it contains "the odds ratios for the predictors" (Logistic Regression, n.d., par. 40) which are the actual values used to interpret the increase or decrease in likelihood, though the sign of the values in column *B* indicates the directionality of the change in odds. For example, in Table 11, based on the signs of the values in *B* and the values in *Exp(B)* for the experimental groups, the output indicates that membership in either experimental group decreased the odds of a target response, though only slightly (0.99 times less likely for EG1 and 0.86 times less likely for EG2). In other words, interaction with either pronunciation technology slightly decreased the odds of a target response in comparison to the participants in CG (who did not complete pronunciation activities). Finally, *Sig.* refers to the level of significance of the calculation; all calculations in this study assume a significance level of $p \leq .05$. Again considering Table 11, neither decrease in odds was statistically significant.

Other information provided in Table 11 includes: *SE, z value, df,* and *CI* First, the abbreviation *SE* refers to standard error, which represents how accurately the sample may represent the population as a whole (Rasinger, 2013, p. 145). Next, *z value* provides "the Wald statistic that tests the hypothesis that the estimate is zero" (Lillis, 2020, par. 20), and *df* is the abbreviation for degrees of freedom. Last, *CI* is the abbreviation for confidence interval, and both the lower and upper bounds of the interval are provided; these represent "an interval of scores that are very likely to contain the real [population] mean" (Eddington, 2015, p. 20).

After the logistic regression calculations, the Wald test was performed. The Wald test "is a way of testing the significance of particular explanatory variables in a statistical model" (Kyngäs & Rissanen, 2001, p. 774). In this case, the Wald test calculates the significance of both independent variables (i.e., the two experimental groups) together[83]. Then, in order to determine the overall effect size, a McFadden's Pseudo R^2 and a p-value based on that effect size were then calculated for any regression demonstrating significance. The R^2 value represents "how much of variance in [the dependent variable] would be accounted for if the model had been derived from the population from which the sample was taken" (Field, 2000, p. 130; as cited in Rasinger, 2013, p. 180). More simplistically and specifically, "r^2 is the fraction of the total variation in y that is explained by variation in x [. . .] [and] [a] value of $r^2 = 1$ means that all of the variation in the response variable is explained by variation in the explanatory variable" (Crawley, 2013, p. 456–457). The new p-value illustrates the significance of the effect size.

Subsequent logistic regressions (and calculations to determine effect size) were then performed on data in which the graphemes were grouped by various unifying characteristics. First, the graphemes were grouped by common features, either place or manner of articulation of their associated phonemes in Spanish; specifically, they fit into three groups: fricatives, stops, and palatals. One grapheme, <h>, was omitted from this analysis due to its target pronunciation being an absence of sound. Second, the graphemes were grouped by whether or not the L2 phoneme exists in the L1 inventory. The final set of groupings was based on the number of grapheme-phoneme associations each grapheme has in the L1. The graphemes in this study have

[83] The significance levels in the logistic regression tables are the result of the Wald test on each (non-reference category) independent variable separately.

up to four associations in English; these associations are explored individually (1 through 4), and a comparison of one association to more than one association is also explored. Table 12 contains the descriptive statistics for the ten graphemes in this study.

Table 12

Descriptive Statistics for Mismatch Accuracy, by Grapheme, by Participant Group

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
<c{a,o,u}>	1820	0.97	0.17	1680	0.96	0.19	840	0.97	0.17
<c{i,e}>	572	0.95	0.21	528	0.95	0.21	264	0.92	0.28
<g{a,o,u}>	559	0.90	0.30	516	0.91	0.28	258	0.91	0.29
<g{i,e}>	390	0.41	0.49	360	0.60	0.49	180	0.46	0.50
<h>	494	0.62	0.49	456	0.77	0.42	228	0.56	0.50
<ll>	416	0.78	0.41	384	0.81	0.39	192	0.88	0.33
<ñ>	390	0.87	0.34	360	0.89	0.32	180	0.82	0.38
<j>	507	0.89	0.31	468	0.87	0.33	234	0.88	0.33
<ch>	429	0.95	0.22	396	0.95	0.21	198	0.96	0.20
<z>	442	0.44	0.50	408	0.50	0.50	204	0.40	0.49

It is important to note that due to the binary coding (i.e., either 1 or 0) of the responses to the graphemes, the average (mean) values in column *M* in Table 12 (and subsequent descriptive statistics tables) also represents the ratio (equivalent to the percentage) of target responses. The variation (or lack thereof) in the averages of the responses for EG1 and EG2 appears to reflect whether the treatment affects the odds of a target response. The subsequent logistic regression tables include the log-odds values (*B*), which assist in interpreting the direction of the change in odds, along with the *p*-values for the independent variables (*Sig.*) and the odds ratios for each treatment group (*Exp(B)*). Alpha levels for all statistical tests are ($p = .05$).

Table 13

Logistic Regression Output for <c{a, o, u}>

	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	95% CI for *Exp(B)* *Lower*	*Upper*
(Int.)	3.44	0.20	17.29	1	$< 2^{-16}$	31.31	21.67	47.46
EG1	−0.01	0.24	-0.05	1	.960	0.99	0.61	1.57
EG2	−0.15	0.24	-0.62	1	.535	0.86	0.53	1.36

The Wald test performed on the logistic regression model (i.e., considering all non-reference category independent variables together) for <c{a, o, u}> was not statistically significant ($\chi^2(2)$ = 0.66, *p* = .720), nor were the differences between CG and either experimental group (Table 13; previously shown as sample data in Table 11). However, of interest to the present study is the evidence that the odds of a target response for both EG1 and EG2 slightly decreased in comparison to CG (0.99 times less likely for EG1 and 0.86 times less likely for EG2).

Table 14

Logistic Regression Output for <c{i, e}>

	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	95% CI for *Exp(B)* *Lower*	*Upper*
(Int.)	2.40	0.22	10.77	1	$< 2^{-16}$	11.00	7.29	17.52
EG1	0.61	0.30	2.04	1	.041	1.84	1.02	3.28
EG2	0.60	0.30	1.995	1	.046	1.83	1.00	3.31

The Wald test performed on the logistic regression model for <c{i, e}>, shown in Table 14, approached significance ($\chi^2(2)$ = 5.30, *p* = .072). Despite this lack of significance when both independent variables were considered jointly, the likelihood of a target response increased significantly for both experimental groups. A target response was 1.84 times more likely after receiving the treatment given to EG1 (*p* = .041) and 1.83 times more likely after receiving the treatment for EG2 (*p* = .046). Though the increase in odds is similar across the experimental groups, this is one of the few instances in which the treatment for EG1 outperformed that of

EG2. The McFadden's Pseudo R^2 calculation revealed that the independent variables in the model accounted for approximately 0.84% of the variation in the data, but that variation approaches significance (R^2 = 0.0084, p = .088).

Table 15

Logistic Regression Output for <g{a, o, u}>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	2.32	0.22	10.64	1	$< 2^{-16}$	10.22	6.81	16.11
EG1	−0.09	0.26	−0.34	1	.735	0.92	0.54	1.51
EG2	0.02	0.27	0.09	1	.928	1.02	0.60	1.72

Presented in Table 15, the Wald test performed on the logistic regression model for <g{a, o, u}> was not statistically significant ($\chi^2(2)$ = 0.31, p = .860). Additionally, neither of the experimental groups produced statistically significantly different odds for a target response when compared to CG. It is interesting to note, however, that the response odds for the two experimental groups did not change in the same direction; the odds of a target response decreased slightly for EG1 (0.92), while the odds increased slightly for EG2 (1.02).

Table 16

Logistic Regression Output for <g{i, e}>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	−0.18	0.15	−1.19	1	.234	0.84	0.62	1.12
EG1	−0.17	0.18	−0.96	1	.338	0.84	0.59	1.20
EG2	0.57	0.18	3.11	1	.002	1.77	1.24	2.55

The Wald test performed on the logistic regression model for <g{i, e}> was highly statistically significant ($\chi^2(2)$ = 26.40, $p = 1.90^{-6}$). The difference in the odds of a target response, shown in Table 16, was also statistically significant when comparing EG2 to CG, EG2 being 1.77 times more likely to produce a target response (p = .002). For EG1, however, the odds of a target response were 0.84 times less likely, though this comparison lacked statistical significance. The McFadden's Pseudo R^2 revealed that the independent variables accounted in the regression model accounted for approximately 2.1% of the variation in the data; however, although the effect size was small, it was highly statistically significant ($R^2 = 0.021$, $p = 1.49^{-6}$).

Table 17

Logistic Regression Output for <h>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	0.23	0.13	1.72	1	.086	1.26	0.97	1.64
EG1	0.25	0.16	1.54	1	.124	1.28	0.93	1.76
EG2	0.99	0.17	5.70	1	1.23^{-8}	2.69	1.92	3.79

The Wald test performed on the logistic regression model for the grapheme <h> was highly statistically significant ($\chi^2(2) = 39.40$, $p = 2.80^{-9}$). The odds of a target response increased for both experimental groups (see Table 17), with EG1 1.28 times more likely to produce a target response, and EG2 2.69 times more likely, though only the latter was (highly) statistically significant ($p = .124$ and $p = 1.23^{-8}$, respectively). The McFadden's Pseudo R^2 indicated that the independent variables accounted for approximately 2.8% of the variation in the data; the effect was highly statistically significant despite the small effect size ($R^2 = 0.028$, $p = 1.05^{-9}$).

Table 18

Logistic Regression Output for <ll>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	1.95	0.22	8.92	1	$< 2^{-16}$	7.00	4.66	11.01
EG1	−0.67	0.25	−2.71	1	.007	0.51	0.31	0.82
EG2	−0.51	0.25	−2.02	1	.043	0.60	0.36	0.97

The Wald test performed on the logistic regression model for <ll> was statistically significant ($\chi^2(2) = 7.30$, $p = .025$). In contrast to the majority of the results from logistical regressions on individual graphemes, the odds of a target response decreased significantly in both experimental groups (see Table 18). A target response was 0.51 times less likely for EG1 (p = .007) and 0.6 times less likely for EG2 (p = .043). The McFadden's Pseudo R^2 revealed that the independent variables accounted for 0.83% of the variation in the data; however, despite the small percentage, the effect size is highly statistically significant (R^2 = 0.0083, p = .018).

Table 19

Logistic Regression Output for <ñ>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	1.53	0.20	7.86	1	3.98^{-15}	4.63	3.20	6.89
EG1	0.34	0.25	1.39	1	.165	1.41	0.86	2.26
EG2	0.52	0.26	2.03	1	.042	1.68	1.01	2.77

Shown in Table 19, the Wald test performed on the logistic regression model for <ñ> was not statistically significant as a whole ($\chi^2(2)$ = 4.20, p = .120). However, there was a statistically significant increase in the odds of a target response for EG2 (1.68 times more likely, p = .042). The odds of a target response also increased for EG1 (1.41 times more likely), though not statistically significantly (p = .165). The McFadden's Pseudo R^2 showed that the independent variables accounted for only approximately 0.55% of the variation in the data; despite one of the individual variables showing a significant effect on the odds of a target response, the overall effect size of the independent variables was not statistically significant (R^2 = 0.0055, p = .132).

Table 20

Logistic Regression Output for <j>

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	1.96	0.20	9.86	1	$< 2^{-16}$	7.07	4.88	10.64
EG1	0.15	0.24	0.62	1	.538	1.16	0.71	1.86
EG2	−0.02	0.24	−0.08	1	.936	0.98	0.60	1.56

The Wald test performed on the logistic regression for <j> was not statistically significant ($\chi^2(2) = 0.81$, $p = .670$). As can be seen in the difference in the sign for EG1 and EG2 in column *B* in Table 20, the likelihood of a target response increased for EG1 but decreased for EG2. A target response was 1.16 times more likely in EG1, but was 0.98 times less likely in EG2, though neither comparison was statistically significant.

Table 21

Logistic Regression Output for <ch>

	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	95% CI for *Exp(B)* *Lower*	*Upper*
(Int.)	3.17	0.36	8.78	1	$< 2^{-16}$	23.75	12.53	52.56
EG1	−0.20	0.42	−0.47	1	.636	0.82	0.34	1.81
EG2	−0.12	0.43	−0.28	1	.777	0.88	0.36	2.01

The Wald test performed on the logistic regression model for <ch> was not statistically significant ($\chi^2(2) = 0.23$, $p = .890$). As shown in Table 21, the odds of a target response decreased in both experimental groups by similar amounts (0.82 times less likely for EG1, 0.88 times less likely for EG2), though neither decrease was statistically significant.

Table 22

Logistic Regression Output for <z>

	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	95% CI for *Exp(B)* *Lower*	*Upper*
(Int.)	−0.42	0.14	−2.92	1	.004	0.66	0.50	0.87
EG1	0.18	0.17	1.05	1	.292	1.20	0.86	1.68
EG2	0.43	0.17	2.46	1	.014	1.53	1.09	2.16

The Wald test performed on the logistic regression model for <z> was statistically significant ($\chi^2(2) = 6.80$, $p = .034$). Looking at Table 22, the odds of a target response increased for both experimental groups. The likelihood of a target response was 1.2 times higher for EG1 ($p = .292$), and 1.53 times higher for EG2 ($p = .014$), though only the latter was statistically significant. The McFadden's Pseudo R^2 indicated that the independent variables accounted for approximately 0.47% of the variation in the data; however, despite the small effect size, the overall impact of the independent variables was statistically significant ($R^2 = 0.0047$, $p = .033$).

The logistic regression models for the individual graphemes indicate that the treatments (most often, the treatment given to EG2) have an impact on the odds of giving a target response to certain graphemes. Table 23 summarizes the logistic regression results for all ten graphemes, including the Wald test and the *p*-values from the McFadden's Pseudo R^2 calculation.

Table 23

Summary of Odds Change Direction and Significance for Investigated Graphemes, with Model and Effect Size, by Experimental Group

	Wald	Pseudo R^2	EG1		EG2	
	Sig.	*Sig.*	Direction	*Sig.*	Direction	*Sig.*
<c{a,o,u}>	nsd	nsd	decrease	nsd	decrease	nsd
<c{i,e}>	(*)	(*)	increase	*	increase	*
<g{a,o,u}>	nsd	nsd	decrease	nsd	increase	nsd
<g{i,e}>	***	***	decrease	nsd	increase	**
<h>	***	***	increase	nsd	increase	***
<ll>	*	*	decrease	**	decrease	*
<ñ>	nsd	nsd	increase	nsd	increase	*
<j>	nsd	nsd	increase	nsd	decrease	nsd
<ch>	nsd	nsd	decrease	nsd	decrease	nsd
<z>	*	*	increase	nsd	increase	*

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

As can be seen in Table 23, the treatment given to EG2 more frequently resulted in a significant change in the odds of a target response, typically an increase in the odds. Conversely, the treatment given to EG1 increased the odds of a target response for half the graphemes, but decreased it for the other half; for EG1, the majority of the changes in odds were not statistically significant. The Wald test pinpointed the graphemes for which the experimental treatments as a whole had a significant effect on the odds of a target response. Finally, the McFadden's Pseudo R^2 calculation and *p*-value indicated the graphemes where, despite small effect sizes generally, the independent variables played a significant role in explaining the variation in the data. These results can be utilized to influence the manner in which troublesome mismatches are presented to language learners (see Section 4.3), but there also may be larger trends that could assist in improving pronunciation accuracy; these are explored in Sections 4.1.5 and 4.1.6.

4.1.5 Analysis Based on Features

Following the regressions calculated on data from individual graphemes, the graphemes were grouped by common phonological features. Specifically, the graphemes were associated with their phonemes in the L2 (Spanish), and then the phonemes were grouped by common features. The phonemes associated with graphemes in this study are: /k/, /s/, /ɡ/, /x/, /ʝ/, /ɲ/, and /t͡ʃ/. Three of the graphemes are associated with the palatals /ʝ/, /ɲ/, and /t͡ʃ/. Two graphemes are associated with the stops /k/ and /ɡ/. Four graphemes are associated with the fricatives /s/ and /x/. The grapheme <h> was not included in this analysis because its pronunciation in Spanish is the absence of sound (i.e., *null* or *null phoneme*[84], represented by the symbol Ø[85]), which cannot be classified by place or manner of articulation (the regression for <h> can be found in Table 20). Table 24 presents the number of tokens, means, and standard deviations for the Place/Manner of Articulation groupings by participant group.

Table 24

Descriptive Statistics for Place/Manner of Articulation Groupings, by Participant Group

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Fricatives	1911	0.71	0.45	1764	0.76	0.43	882	0.69	0.46
Stops	2379	0.95	0.21	2196	0.95	0.21	10.98	0.96	0.21
Palatals	1235	0.99	0.34	1140	0.88	0.32	570	0.89	0.32

[84] In text-to-phoneme (TTP) models, used in speech recognition systems, "null phonemes are inserted to handle cases where a letter does not map to a phoneme" (Jensen & Riis, 2000, p. 318).
[85] "This symbol has no phonetic value, but is used by phonologists to notate a zero morpheme, or to indicate nothing" (Pullum & Ladusaw, 1996, p. 136).

The values for the average (*M*) of the responses, displayed in Table 24, provide insight into where treatment may have had an effect. Due to the binary nature of the data, the values in the average column also represent the percentage of target responses in each grouping (e.g., 0.71 = 71%). While the averages for Stops and Palatals are very similar across the three groups, EG2 shows a higher percentage of target responses to graphemes that are associated with fricatives than the other groups. Based on the similarity of the average values for Stops and Palatals, it appears unlikely that there would be a difference in the odds of a response; conversely, the difference between the average of responses between CG and EG2 has the potential for a significant change in the odds of a target response. Tables 25, 26, and 27 provide the logistical regression output for the place/manner of articulation groupings.

Table 25

Logistic Regression Output for Fricatives Grouping

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	0.81	0.07	11.08	1	$< 2^{-16}$	2.24	1.95	2.59
EG1	0.08	0.09	0.90	1	.378	1.08	0.91	1.29
EG2	0.32	0.09	3.48	1	5.06^{-4}	1.37	1.15	1.64

Table 26

Logistic Regression Output for Stops Grouping

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	3.06	0.15	20.97	1	$< 2^{-16}$	21.41	16.27	28.89
EG1	−0.05	0.18	−0.27	1	.790	0.95	0.67	1.34
EG2	−0.07	0.18	−0.41	1	.683	0.93	0.65	1.31

Table 27

Logistic Regression Output for Palatals Grouping

	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	95% CI for *Exp(B)* *Lower*	*Upper*
(Int.)	2.07	0.13	15.59	1	$< 2^{-16}$	7.91	6.15	10.35
EG1	−0.19	0.16	−1.22	1	.233	0.83	0.60	1.12
EG2	−0.04	0.16	−0.27	1	.789	0.96	0.69	1.31

Of the three groupings, only Fricatives (Table 25) demonstrated statistical significance from the Wald test performed on the logistic regression model ($\chi^2(2) = 15.50$, $p = 4.3^{-4}$). Additionally, a target response was 1.08 times more likely after the treatment given to EG1, and 1.37 times more likely after the treatment received by EG2, though only the latter was (highly) statistically significant ($p = .378$ and $p = 5.06^{-4}$, respectively). Due to the difference in the average of the responses for Fricatives between CG and EG2 (see Table 27), the odds being statistically significantly better for EG2 is logical. The McFadden's Pseudo R^2 revealed that the independent variables accounted for only approximately 0.29% of the variation in the data, but the effect size was highly statistically significant ($R^2 = 0.0029$, $p = 3.90^{-4}$).

In contrast, neither Wald test performed on the logistic regression models for the Stops ($\chi^2(2) = 0.17$, $p = .920$) nor the Palatals ($\chi^2(2) = 2.10$, $p = .350$) grouping demonstrated statistical significance (see Tables 26 and 27, respectively). For Stops, the odds of a target response slightly decreased in both of the experimental groups (0.95 times less likely for EG1, 0.93 times less likely for EG2) though neither was statistically significant. The Palatals grouping followed a similar pattern: both experimental treatments appeared to decrease the odds of a target response slightly (0.83 times less likely for EG1, 0.96 times less likely for EG2), though neither change in odds was statistically significant. Referring back to Table 24, these results are unsurprising because the percentages of target responses are very high in both groupings.

Additionally, looking at the averages between the three groups for the Stops grouping specifically, they are within approximately two percentage points of each other.

Table 28 summarizes the logistic regression results for the place/manner of articulation groupings of the graphemes in the present study. The Wald test and the McFadden's Pseudo R^2 significance values are also represented.

Table 28

Summary of Odds Change Direction and Significance for Place/Manner Groupings, with Model and Effect Size, by Experimental Group

	Wald	Pseudo R^2	EG1		EG2	
	Sig.	*Sig.*	Direction	*Sig.*	Direction	*Sig.*
Fricatives	***	***	increase	nsd	increase	***
Stops	nsd	nsd	decrease	nsd	decrease	nsd
Palatals	nsd	nsd	decrease	nsd	decrease	nsd

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

As can be seen in Table 28, Fricatives were the only grouping in which there was an increase in the odds of a target response in either experimental group. Moreover, it was the only grouping that showed statistically significant results. These results can be used to further guide acquisition of pronunciation (see Section 4.3). However, the analysis and any recommendations for pronunciation instruction is not complete unless the potential effects of the graphemic and phonological inventories of the L1 are considered.

4.1.6 Contrastive Analysis

The graphemes selected for investigation in the present study were chosen because their orthographic representations a) do not align with the expected phonological representation, or b) align to multiple phonological representations. What is left unstated is that these two rationales for inclusion are not based solely in the L2, but rather consider the potential difficulties of acquiring another language with a fully developed L1 already in place. The idea that grapheme-phoneme associations from the L1 can influence L2 learners' pronunciation of L2 graphemes intersects with contrastive analysis. Essentially, contrastive analysis can be explained as "predictions based on cross-linguistic comparisons" (Odlin, 2014, p. 28). The concept of contrastive analysis lends itself well to logistic regression analysis; both function on the premise of prediction, the former predicting L2 errors based on the L1, and the latter predicting odds based on comparison between an independent variable and the reference category. The two contrasts that are the focus of this section stem from the L1 phonological inventory and L1 grapheme-phoneme association(s).

The first contrastive analysis is based on whether the Spanish phoneme can be found in the phonological inventory of the ten investigated graphemes in the L1 (English). The graphemes whose associated phonemes in Spanish <u>are</u> found in the English phonological inventory are: <c{a,o,u}> /k/; <c{i,e}> /s/; <g{a, o, u}> /g/; <z> /s/; <ch> /t͡ʃ/; <h> Ø[86]. While not all of these sounds occur in the same way in both languages, they occur in both phonological systems and therefore ought to be less difficult to produce in the L2 (Cichoki et al., 1999). For example, <z> canonically correlates to /s/ in Spanish, though that is not the case for English; however, /s/ exists in the English phonological inventory (Odden, 2005, p. 148), meaning it is not a new, unfamiliar sound that learners must adapt to. Relatedly, <h> overwhelmingly corresponds to /h/ in English, but there are high-frequency words in which <h> is not pronounced, such as *hour*; this pronunciation is in line with the pronunciation of <h> in Spanish, which places <h> in this group.

Conversely, the present study also includes graphemes whose phonemes (when considering canonical pronunciation) in Spanish cannot be found in the phonological inventory

[86] See explanation in Section 4.1.5

of English (specifically American English). Two of the graphemes (<g{i, e}> and <j>) correspond to the voiceless velar fricative /x/. Additionally, the grapheme <ll> canonically corresponds to the palatal lateral /ʎ/. Finally, the grapheme <ñ> corresponds to the palatal nasal /ɲ/. None of these sounds is native to English. Table 29 contains descriptive statistics for *Yes* (i.e., the phoneme *is* in the L1 inventory) and *No* (i.e., the phoneme is *not* in the L1 inventory).

Table 29

Descriptive Statistics for L1 Inventory Presence, by Participant Group

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Yes	4316	0.90	0.34	3984	0.89	0.32	1992	0.85	0.36
No	1703	0.75	0.43	1572	0.80	0.40	786	0.78	0.42

As can be seen in Table 29, there is variation in the averages (*M*) for both *Yes* and *No* between the three participant groups. Column *M* also depicts the differences in the percentages of target responses for those groups because the data are coded in a binary fashion. Tables 30 and 31 contain the calculations from the logistic regressions.

Table: 30

Logistic Regression Output for Phonemes in the L1 Inventory ('Yes')

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	1.72	0.06	27.54	1	$< 2^{-16}$	5.57	4.94	6.31
EG1	0.12	0.08	1.51	1	.132	1.12	0.96	1.30
EG2	0.33	0.08	4.14	1	3.51^{-5}	1.39	1.19	1.63

Table 31

Logistic Regression Output for Phonemes Not in the L1 Inventory ('No')

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	1.19	0.08	14.13	1	$< 2^{-16}$	3.30	2.80	3.90
EG1	−0.10	0.10	−0.97	1	.334	0.91	0.74	1.10
EG2	0.18	0.11	1.67	1	.095	1.19	0.97	1.46

The Wald test performed on the logistic regression model for phonemes that *are* found in the L1 (English) inventory is highly statistically significant ($\chi^2(2) = 19.3$, $p = 6.50^{-5}$). As compared to CG (see Table 30), being a member of either experimental group increases a participant's odds of producing the target sound. Specifically, the odds of producing a target response were 1.12 times higher when receiving the treatment given to EG1, and 1.39 times higher for the treatment given to EG2, though only the latter was statistically significant ($p = .132$ and $p = 3.51^{-5}$, respectively). The McFadden's Pseudo R^2 revealed that the independent variables accounted for approximately 0.24% of the variation in the data; however, despite them accounting for such a small amount, their effect was highly statistically significant ($R^2 = 0.0024$, $p = 5.95^{-5}$).

In contrast to the analysis on Spanish sounds found in the L1, Table 31 illustrates that while the Wald test performed on the logistic regression model for graphemes whose phonemes (in Spanish) do not occur in English is statistically significant ($\chi^2(2) = 10.70$, $p = .005$), being a member of either experimental group did not statistically significantly affect the odds of a target response (EG1: $p = .334$; EG2: $p = .095$). Moreover, the treatment for EG1 appears to make the odds of a target response 0.91 times less likely for sounds *not* found in the L1, though the treatment for EG2 makes the odds of a target response 1.19 times more likely. The McFadden's Pseudo R^2 indicated that the independent variables accounted for approximately 0.25% of the variation in the data, though that small percentage of variation was statistically significant ($R^2 = 0.0025$, $p = .004$).

The second contrastive analysis pertains to the number of possible pronunciations of the grapheme in English (the L1)[87]. The grapheme-phoneme correspondence in English is opaque

---

[87] The grapheme <ñ> is not included in this analysis because it is not native to English.

(Ellis & Hooper, 2002; Timmer & Schiller, 2012) whereas the correspondence in Spanish is more transparent (Carreiras et al., 2014; Hualde, 2014; Meschyan & Hernandez, 2005)[88]. Therefore, there is the potential for interference for graphemes that, in the L1, correspond to multiple phonemes. The graphemes investigated in the present study are divided into four groupings: one (1), two (2), three (3), and four (4) L1 phonemes associations. Three of the graphemes have been classified as having a one-to-one correspondence between grapheme and phoneme in the L1[89]: <c{a,o,u}> /k/; <g{a, o, u}> /ɡ/; <ll> /l/. Two graphemes have two associations in the L1: <h> is associated with /h/ and ∅, and <z> is associated with /z/ and /ʒ/[90]. Two graphemes have three associations in the L1: <g{i, e}> is associated with /d͡ʒ/, /ʒ/, and /ɡ/, and <j> is associated with /d͡ʒ/, /j/, and /h/. Two graphemes have four associations in the L1; <c{i, e}> is associated with /s/, /ʃ/, /k/, and /t͡ʃ/, and <ch>[91] is associated with /t͡ʃ/, /ʃ/, /k/, and ∅. The results of this analysis are presented two ways. First, one (1) L1 option is compared to two or more options (2+: including two (2), three (3), and four (4) options); Table 32 presents descriptive statistics for this comparison.

Table 32

Descriptive Statistics for Number of L1 Options, 1 and 2+, by Participant Group

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
1 opt	2795	0.93	0.26	2580	0.93	0.25	1290	0.94	0.25
2+ opt	2834	0.73	0.44	2616	0.79	0.41	1308	0.71	0.45

[88] See Section 2.4 for further discussion.
[89] This excludes whole borrowings such as *Caesar, margarine, quesadilla,* and *bougainvillea*. It also excludes *llama*, but acknowledges that *llama* originated in Quechua (Hualde et al., 2010, p. 321); according to Hualde et al. (2010), "probablemente no es casualidad que una de las variedades del español latinoamericano donde se mantiene mejor el fonema palatal lateral /ʎ/ es precisamente el español andino, en contacto con el quechua y el aimara" [it probably is not a coincidence that one of the varieties of Latin American Spanish where the palatal lateral phoneme /ʎ/ is better maintained is precisely Andean Spanish, in contact with Quechua and Aimara] (p. 322).
[90] This excludes <z> as found in *schizophrenia* and *pizza.*
[91] This excludes whole borrowings from Hebrew such as *challah* and *chutzpah*.

Table 32 contains descriptive statistics for one (1) and two or more (2+) L1 options. As the coding for the data is binary, the values in the average (*M*) column are also the percentage of target responses. Based on the differences between these values, a significant change in the odds of a response appears more probable for the 2+ options group, as the averages for the one option group are similar across participant groups. Tables 33 and 34 present the logistic regression output for one (1) and two or more (2+) L1 options, respectively.

Table 33

Logistic Regression Output for Graphemes with One (1) L1 Option

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	2.81	0.12	23.35	1	$< 2^{-16}$	16.67	13.27	21.29
EG1	−0.26	0.14	-1.90	1	.064	0.77	0.58	1.01
EG2	−0.22	0.14	−1.52	1	.129	0.80	0.60	1.06

Table 34

Logistic Regression Output for Graphemes with Two or More (2+) L1 Options

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	0.89	0.06	14.61	1	$< 2^{-16}$	2.43	2.16	2.74
EG1	0.10	0.07	1.36	1	.175	1.11	0.96	1.28
EG2	0.43	0.08	5.49	1	3.99^{-8}	1.53	1.31	1.78

As a whole, the Wald test performed on the logistic regression model for graphemes with only one option in the L1 (English), shown in Table 33 was not statistically significant ($\chi^2(2) = 3.50$, $p = .170$). Despite there only being one option for the pronunciation of the grapheme in the L1, the information in Table 34 shows that the likelihood of a target response decreased in both experimental groups. For EG1, the likelihood of a target response became 0.78 times less likely, and for EG2, it became 0.8 times less likely, though only the former approached significance ($p = .064$ and $p = .130$, respectively).

In contrast to the model for graphemes with only one option, the Wald test performed on the logistic regression model for graphemes that map to at least two phonemes in the L1 was highly statistically significant ($\chi^2(2) = 39.40$, $p = 4.60^{-9}$). Additionally, in this case, the likelihood of a target response increased in both experimental groups, illustrated in Table 34. The odds of a target response were 1.11 times higher after receiving the treatment for EG1, and 1.53 times higher after receiving the treatment for EG2, though only the latter was (highly) statistically significant ($p = .175$ and $p = 3.99^{-8}$, respectively). The McFadden's Pseudo R^2 revealed that the independent variables accounted for approximately 0.51% of the variation in the data; however, the effect size was highly statistically significant despite its small size ($R^2 = 0.0051$, $p = 3.33^{-9}$).

As statistically significant results were found for 2+ options, the graphemes in this grouping were then divided into separate groups based on the number of options in the L1. These groups were then analyzed by themselves in order to determine whether specific numbers of L1 options affect accuracy in pronunciation. The following tables provide descriptive statistics (Table 35) for all four options, and regression outputs for two (2), three (3), and four (4) L1 options (Tables 36, 37, and 38, respectively); for comparison, the regression output for one (1) L1 option can be found in Table 33.

Table 35

Descriptive Statistics for Number of L1 Options, 1-4, by Participant Group

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
1 opt	2795	0.93	0.26	2580	0.93	0.25	1290	0.94	0.25
2 opt	936	0.53	0.50	864	0.64	0.48	432	0.48	0.50
3 opt	897	0.68	0.47	828	0.75	0.43	414	0.69	0.46
4 opt	1001	0.95	0.21	924	0.95	0.21	462	0.94	0.25

The differences in the averages (*M*) in Table 35 highlight the potential for a change in odds of a target response (comparing each experimental group with CG individually). As has been stated previously, since the coding for the data is binary, the averages also represent the percent of target responses for each L1 options grouping for each participant group. The statistics in Table 35 inform the subsequent logistic regressions.

Table 36

Logistic Regression Output for Graphemes with Two (2) L1 Options

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	−0.07	0.10	−0.77	1	.442	0.93	0.77	1.12
EG1	0.21	0.12	1.81	1	.070	1.24	0.98	1.55
EG2	0.67	0.12	5.60	1	2.19^{-8}	1.95	1.55	2.47

Table 37

Logistic Regression Output for Graphemes with Three (3) L1 Options

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	0.82	0.11	7.65	1	2.01^{-14}	2.26	1.84	2.79
EG1	−0.05	0.13	−0.36	1	.721	0.96	0.74	1.23
EG2	0.30	0.13	2.27	1	.024	1.35	1.04	1.76

Table 38

Logistic Regression Output for Graphemes with Four (4) L1 Options

							95% CI for *Exp(B)*	
	B	*SE*	*z value*	*df*	*Sig.*	*Exp(B)*	*Lower*	*Upper*
(Int.)	2.67	0.19	14.13	1	$< 2^{-16}$	14.4	10.13	21.3
EG1	0.32	0.24	1.34	1	.181	1.38	0.85	2.19
EG2	0.35	0.25	1.44	1	.150	1.42	0.87	2.29

The Wald test performed on the logistic regression model for graphemes with two (2) options in the L1 was highly statistically significant ($\chi^2(2) = 37.9$, $p = 6.00^{-9}$). Table 36 shows that the likelihood of producing a target response was 1.95 times higher for members of EG2 in comparison to CG, an effect which was highly statistically significant (p = 2.19^{-8}). Additionally, the likelihood of producing a target response was 1.24 times greater for members of EG1 (compared to CG), an effect which approached significance (p = .070). The McFadden's Pseudo R^2 indicated that the independent variables accounted for approximately 1.3% of the variation in the data, an effect size which is highly statistically significant (R^2 = 0.013, p = 4.48^{-9}).

The Wald test performed on the logistic regression model for graphemes that correspond to three (3) phonemes in the L1 (English) is also statistically significant ($\chi^2(2)$ = 11.20, p = .004). Additionally, from the output included in Table 37, the odds of a target response are 1.35 times more likely after receiving the treatment given to EG2 (p = .024). Conversely, it appears that receiving the treatment given to EG1 slightly decreases the odds of a target response (0.96 times less), though that comparison was not statistically significant. The McFadden's Pseudo R^2 revealed that the independent variables accounted for approximately 0.45% of the variation in the data; despite the small effect size, the effect is statistically significant (R^2 = 0.0045, p = .003). In contrast to two (2) and three (3) options, the Wald test performed on the logistic regression model did not show statistical significance for graphemes with four (4) L1 options ($\chi^2(2)$ = 2.40, p = .300) (see Table 38). Relatedly, though group membership did increase the odds of a target response (1.38 times more likely for EG1 and 1.42 times more likely for EG2), neither were statistically significant (p = .181 and p = .150, respectively).

Overall, graphemic and phonemic elements of the L1 (in this case, English) has a significant impact on the odds of a target response. Table 39 summarizes the results of the logistical regressions, Wald tests, and p-values associated with the McFadden Pseudo R^2 calculation for the contrastive analysis groupings.

Table 39

Summary of Odds Change Direction and Significance for Contrastive Analyses, with Model and Effect Size, by Experimental Group

	Wald	Pseudo R^2	EG1		EG2	
	Sig.	*Sig.*	Direction	*Sig.*	Direction	*Sig.*
L1 inventory (yes)	***	***	increase	nsd	increase	***
L1 inventory (no)	**	**	decrease	nsd	increase	nsd
1 L1 option	nsd	nsd	decrease	(*)	decrease	nsd
2+ L1 options	***	***	increase	nsd	increase	***
2 L1 options	***	***	increase	(*)	increase	***
3 L1 options	**	**	decrease	nsd	increase	*
4 L1 options	nsd	nsd	increase	nsd	increase	nsd

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

Based on the contrastive analyses summarized in Table 39, the graphemic and phonological inventories of the L1 appear to play a significant role in the odds of a target response, overall and specifically in response to the treatment given to EG2. The results from the contrastive analysis perspective, combined with the information from the place/manner of articulation groupings and the analysis of the graphemes individually, provide a more complete picture of the effects of two different types of pronunciation practice on grapheme-phoneme mismatches. The following section discusses these results and their impact in the study, as well as their impacts on the fields of second language acquisition and second language instruction.

4.2 Acoustic Phonetics Findings

In contrast to the more overt, more noticeable, and potentially more distracting grapheme-phoneme mismatches, this study also investigated the more implicit and subtle phenomenon of voice onset time (VOT). Unlike the mismatches discussed previously, even though the VOT values for English and Spanish are mutually exclusive (i.e., they do not overlap), they are not phonemically contrastive in either language. (As was illustrated in Figures 5 and 6 (see Section 2.5), pronouncing the Spanish *por* [p] 'for' as the English *pour* [p^h] may register as a non-native accent, but does not create a new lexical item.) These measurements were compiled for each of the three voiceless stops (/p t k/) in both of the word positions (initial and medial). Paired t-tests were then conducted within each experimental group in order to determine whether improvement had occurred over time; prior to performing the tests, the data were separated so that only tokens which had been successfully recorded and measured for both recordings were included. These t-tests compared pre-test VOT values to post-test VOT values, as well as word-initial VOT values to word-medial VOT values. Subsequently, an analysis of variance (ANOVA), with a post-hoc Tukey's HSD test, plus a linear regression were conducted. The ANOVA and Tukey's HSD were utilized to determine if the participant groups performed differently from each other at the end of the experiment (i.e., analyzing post-test data) and between which groups those differences occurred. The linear regression was performed in order to determine if there was an interaction between group membership (i.e., the three technological instruments), word location (i.e., word-initial or word-medial placement), and the duration (measurement) of VOT.

This section presents findings in the following order. First, changes in VOT values over time (i.e., pre-test versus post-test, word-initial versus word-medial) are illustrated in order to establish whether or not time plays a role in change in VOT overall. Then, shifting to focus on post-test data only, the results of the ANOVA and its post-hoc test, plus the linear regression are presented and interpreted. Finally, the implications of the results are discussed and positioned within previous research.

Thorough consideration was given to the treatment of outliers, which appear as circles in the boxplots found in this section. Initially, winsorization was considered, and appeared to be a useful way to maintain all data points while simultaneously counteracting any effect of the extreme values (i.e., outliers) by replacing them with "the highest and lowest remaining values before any calculations are performed. Typically, an equal number of high and low values are replaced, representing 10% to 25% of the total distribution" (American Psychological Association, 2020, par. 2). However, ultimately the decision was made to leave the outliers in the data. The rationale behind this decision first and foremost stems from the winsorizing procedure itself. As winsorizing is effectively the trimming of data, it must be done systematically; therefore, if a certain percent of the data are trimmed from one end of the spectrum, the same percent must also be trimmed from the other end (Glen, May 2016). However, trimming data from both ends of the spectrum would actually be misleading when looking at VOT, particularly when comparing two languages with mutually exclusive accepted values (Flege & Hillenbrand, 1984; Flege et al., 1994; Kissling, 2013; Ladefoged, 1975; Lisker & Abramson, 1964; Nagle, 2018; Zampini, 2014) and when improvement in L2 VOT acquisition would be indicated by values that decrease in duration. Additionally, despite the presence of outliers, no average VOT in any condition fell outside the accepted values for voiceless stops in stressed, onset positions in the L1 (English) (Ellis & Hooper, 2002; Timmer & Schiller, 2012). The final rationale for keeping the data whole and unmodified was that not all outliers could be accounted for by hesitancy, especially when a lack of hesitancy might be unusual for some students in a first semester foreign language class.

4.2.1 Descriptive Statistics for /p t k/ in All Participant Groups and Word Locations

Pertaining to the analysis of VOT in the present study, Tables 40 and 41 provide the numeric values that inform the majority of the statistical analyses in the remainder of this section. Table 40 contains the pre-test values, separated by phoneme (/p/, /t/, and /k/) and word location (word-initial (WI) and word-medial (WM)). It is important to note that there were not the same number of measurements for each phoneme for multiple reasons. First, there were different numbers of participants in all three of the groups (13 in EG1, 12 in EG2, and six (6) in CG). Additionally, while the experiment design was engineered so that all six phoneme-location pairings were recorded at least 30 times each (see Section 3.3), only tokens which were successfully produced in both recordings were considered for analysis, leading to the variation in the number of tokens measured for each phoneme-location pairing. For example, if a participant accidentally pronounced *esta* 'this' (F.SG) in *esta calle* 'this street' as *estas* 'these' (F.PL), then the /k/ that begins the word *calle* was not preceded by a vowel and was ineligible for VOT measurement per the criteria set forth in this study. Then, since that token was not measured, its counterpart in the other recording was also removed (if it had been successfully recorded).

Table 40

Descriptive Statistics for Pre-Test VOT Duration (in milliseconds)

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
/p/ WI	388	55.74	20.98	361	58.97	26.20	178	67.42	44.86
/p/ WM	334	42.58	19.19	323	45.38	25.64	151	48.11	32.73
/t/ WI	461	61.99	21.19	433	62.36	24.98	216	65.68	33.28
/t/ WM	348	55.06	23.61	334	54.30	28.44	163	66.44	37.52
/k/ WI	400	63.44	23.65	373	63.58	22.62	193	70.74	28.69
/k/ WM	313	50.43	19.87	291	48.79	22.23	156	56.75	31.15

Table 41

Descriptive Statistics for Post-Test VOT Duration (in milliseconds)

	EG1			EG2			CG		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
/p/ WI	388	53.34	25.07	361	49.25	20.01	178	56.41	28.78
/p/ WM	334	38.90	19.21	323	34.99	18.88	151	39.91	22.35
/t/ WI	461	59.49	25.50	433	51.96	24.65	216	56.71	26.06
/t/ WM	348	50.38	26.43	334	44.16	21.80	163	56.44	32.38
/k/ WI	400	61.62	21.66	373	55.54	20.23	193	61.69	30.47
/k/ WM	313	46.08	19.39	291	43.91	18.53	156	45.92	21.44

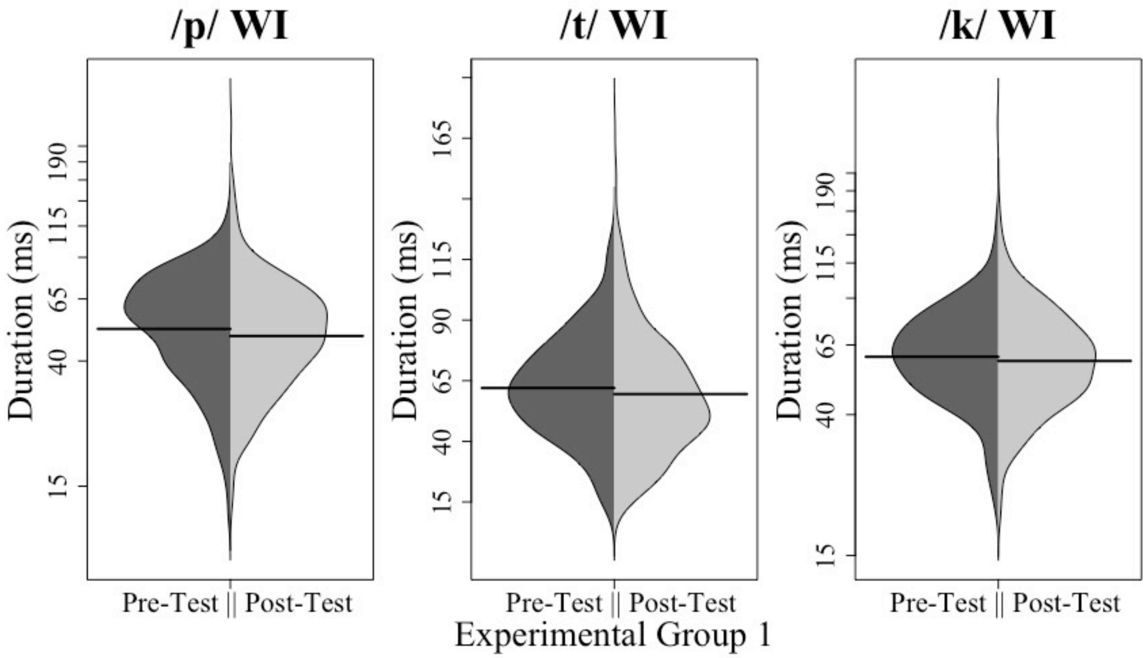
Table 41 contains the numeric values from the post-test recording, also showing variation in the number of tokens measured. However, it is important to note that those numbers (in the *N* columns of Tables 40 and 41) are identical. Due to the pre-test/post-test design of the experiment, the data are paired so that each participant's utterance in the pre-test is compared to their identical utterance in the post-test. At a glance, it is clear that the average VOT for each phoneme-location pairing decreased over time; the subsequent sections analyze those differences.

4.2.2 Experimental Group 1 Pre-Test and Post-Test Comparisons

Figures 9 and 10 illustrate the range of measurements for Experimental Group 1 (EG1), or the group that interacted with the mobile app that included a speech recognition feature (with a small amount of feedback). These measurements, and all subsequent references to calculations, utilize milliseconds (rather than a fraction of seconds) when describing the VOT measurements. Based on the differences between VOT values in English and Spanish, in order for the participants to demonstrate improvement in VOT, the length should decrease, as VOT values are smaller in Spanish than in English. Figure 9 illustrates VOT duration for /p t k/ in word-initial position, while Figure 10 depicts /p t k/ duration in word-medial position.

Figure 9

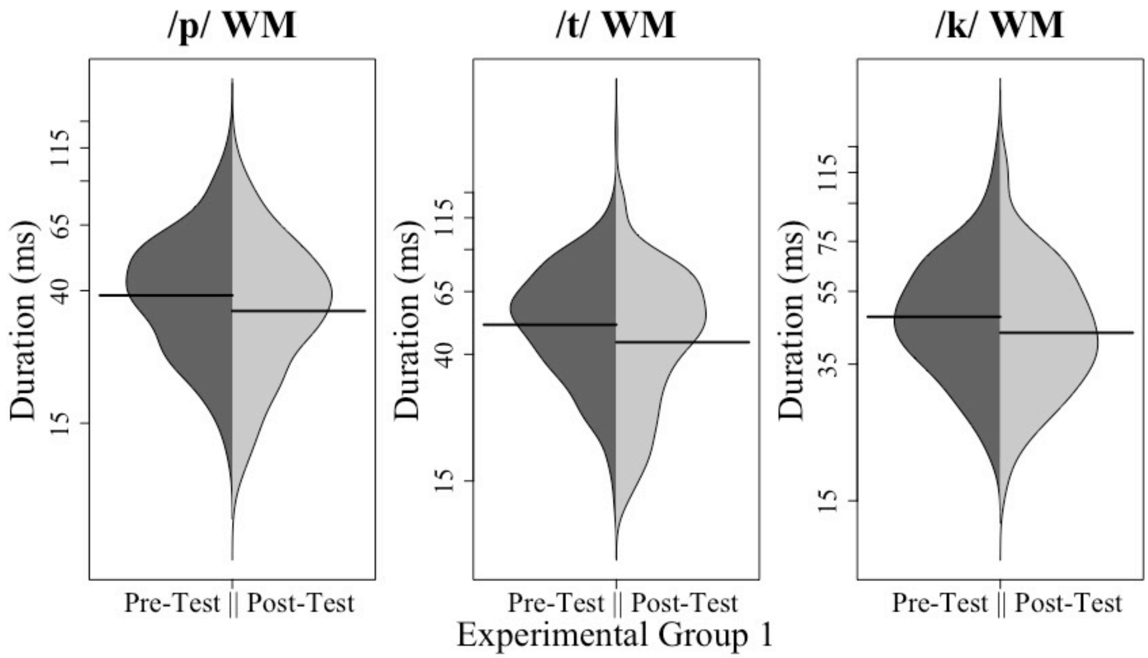
Beanplots of Word-Initial VOT for /p t k/, Pre- and Post-Test, for Experimental Group 1



The dark gray half of each bean represents the Baseline (or Pre-Test) recording, and the light gray half represents Recording 3 (or Post-Test). The lines bisecting each half of the beans represent the average VOT duration. As can be seen in Figure 9, the halves of each beanplot appear similar to each other for all three voiceless stops. Additionally, there is little separation between the average lines in any of the beanplots. While the density appears to be shifting downward when comparing the post-test to the pre-test, indicating that the VOT duration production is decreasing (i.e., moving toward a more Spanish-like production), due to the similarities in density and average duration, it is unsurprising that no statistical difference was found over time for /p/ ($t(387) = 1.63$, $p = .104$), /t/ ($t(460) = 1.91$, $p = .057$), or /k/ ($t(399) = 1.17$, $p = .243$) in word-initial position in EG1, though /t/ approached significance.

Figure 10

Beanplots of Word-Medial VOT for /p t k/, Pre- and Post-Test, for Experimental Group 1



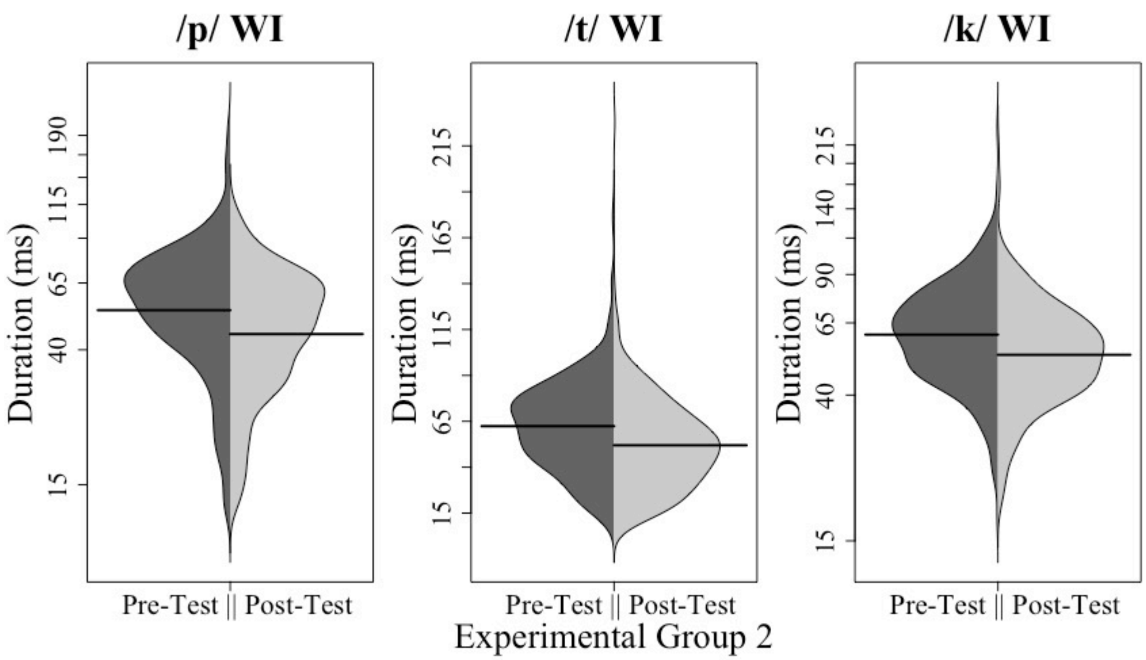
In contrast to Figure 9, the beanplots in Figure 10 all show a shift in density of VOT duration between the pre-test (dark gray) and the post-test (light gray) for EG1. Additionally, there is much more separation between the average bars on each side of the bean for all three voiceless stops. The paired t-tests revealed that differences between VOT lengths from the Baseline Recording (pre-test) to the final recording (post-test) were statistically significant for /p/ in word-medial stressed position ($t(333) = 2.85$, $p = .005$), /t/ in word-medial stressed position ($t(347) = 3.02$, $p = .003$), and were highly statistically significantly different for /k/ in word-medial stressed position ($t(312) = 3.22$, $p = 1.43^{-3}$).

4.2.3 Experimental Group 2 Pre-Test and Post-Test Comparisons

The results discussed in this section illustrate the VOT values from Experimental Group 2 (EG2); this group worked with the listen-and-repeat style activities included with their course textbook. The following figures present comparisons between the pre-test and post-test values for each stop. Figure 11 illustrates VOT duration for participants in EG2 over time for /p t k/ in word-initial position, while Figure 12 depicts /p t k / duration in word-medial position.

Figure 11

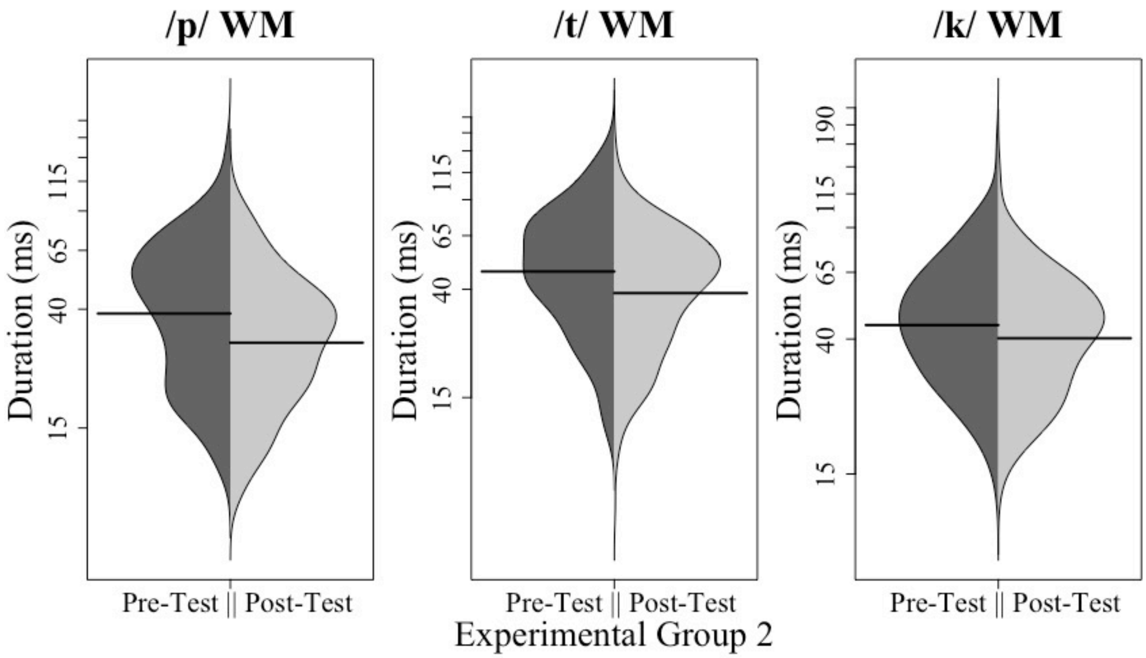
Beanplots of Word-Initial VOT for /p t k/, Pre- and Post-Test, for Experimental Group 2



As can be seen in Figure 11, there is a clear separation between the average VOT duration between pre-test and post-test performance for all three voiceless stops. Additionally, all three depict a downward shift in duration over time. The paired t-tests revealed that the decrease between the values for word-initial /p/ ($t(360) = 6.88$, $p = 2.62^{-11}$), /t/ ($t(432) = 8.92$, $p = 2.20^{-16}$), and /k/ ($t(372) = 5.81$, $p = 1.37^{-8}$) were all highly statistically significant.

Figure 12

Beanplots of Word-Medial VOT for /p t k/, Pre- and Post-Test, for Experimental Group 2



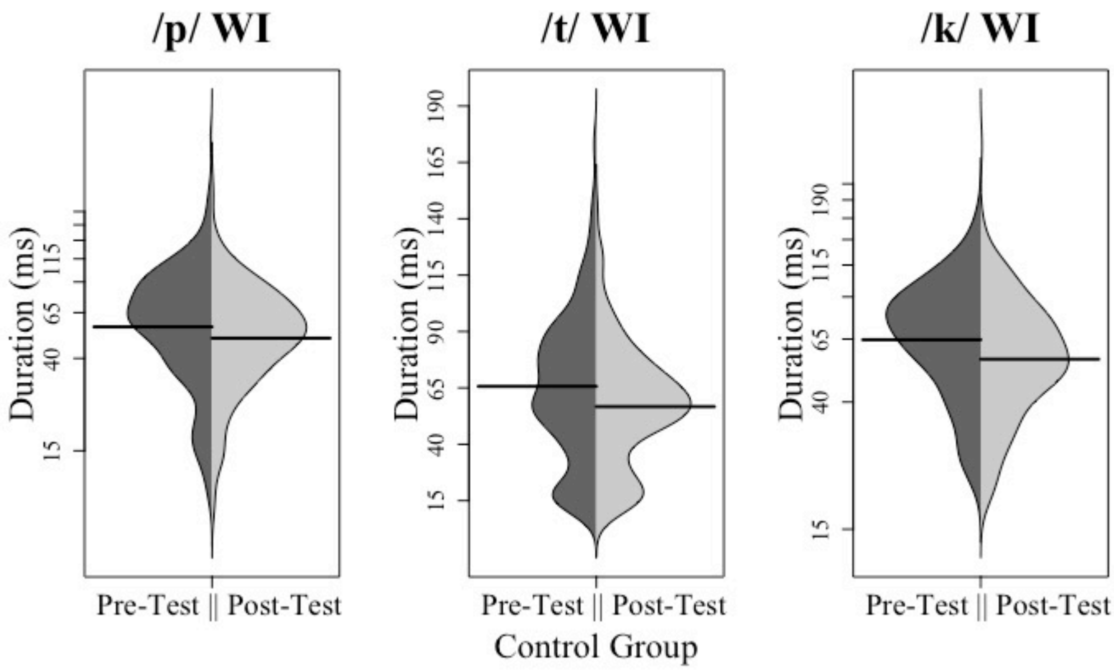
As was the case for the voiceless stops in word-initial position, VOT duration in word-medial position for EG2 also changed over time, as can be seen in Figure 12, with clear separation between the average bars for the pre- and post-test data for all three stops. This separation, combined with the downward shift in VOT duration density in the post-test data (light gray) as opposed to the pre-test data (dark gray), clearly show improvement over time. The paired t-tests indicated that the decreases between the word-medial values for /p/ ($t(322) = 7.04$, $p = 1.20^{-11}$), /t/ ($t(333) = 6.73$, $p = 7.66^{-11}$), and /k/ ($t(290) = 3.35$, $p = 9.22^{-4}$) values showed highly statistically significant differences over time.

4.2.4 Control Group Pre-Test and Post-Test Comparisons

In the present study, the Control Group (CG) performed grammar-related activities through Blackboard, which did not contain a pronunciation element. Thus, the expectation would likely be that the improvement between pre-test and post-test would be smaller or even negligible when compared to the two experimental groups. However, this was not the case. Figure 13 depicts VOT duration for CG over time in word-initial position, and Figure 14 illustrates /p t k/ duration in word-medial position.

Figure 13

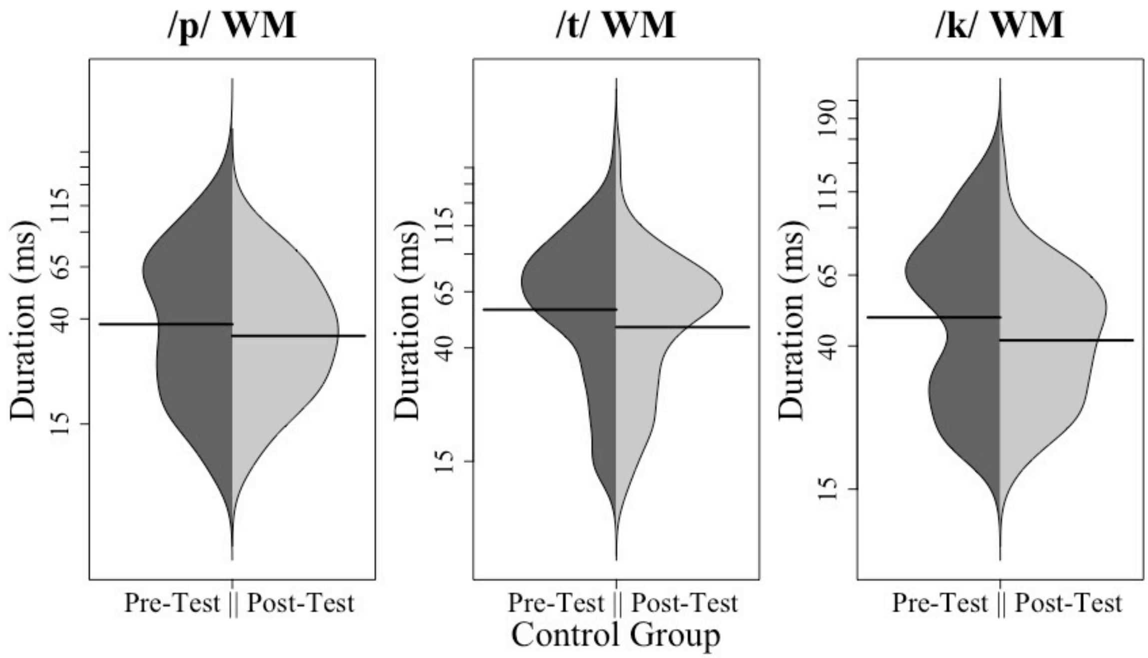
Beanplots of Word-Initial VOT for /p t k/, Pre- and Post-Test, for Control Group



Shown in Figure 13, the production of all three voiceless stops over time by participants in CG decreased over time, with a clear separation between the average VOT between the pre-test and post-test and a general downward trend in VOT duration. Paired t-tests revealed that the change over time for CG was highly statistically significant for word-initial /p/ ($t(177) = 3.57$, $p = 4.53^{-4}$), word-initial /t/ ($t(215) = 6.31$, $p = 1.61^{-9}$), and word-initial /k/ ($t(192) = 4.49$, $p = 1.23^{-5}$).

Figure 14

Beanplots of Word-Medial VOT for /p t k/, Pre- and Post-Test, for Control Group



As was the case for word-initial VOT for CG, Figure 14 also shows separation between the average VOT duration pre-test to post-test, as well as a shift in density toward smaller VOT duration values. The paired t-tests showed that the change in VOT production over time for participants in CG was highly statistically significant in word-medial position for /p/ ($t(150) = 3.51$, $p = 5.90^{-6}$), /t/ ($t(162) = 4.70$, $p = 5.42^{-6}$) and /k/ ($t(155) = 4.78$, $p = 4.06^{-6}$). Overall, the data show that the participants' VOT values did change significantly over time. Table 42 provides a summary of statistical significance for the t-tests presented in this section, while Figure 15 presents a visual representation of the average VOT durations, pre-test to post-test, by word location.

Table 42

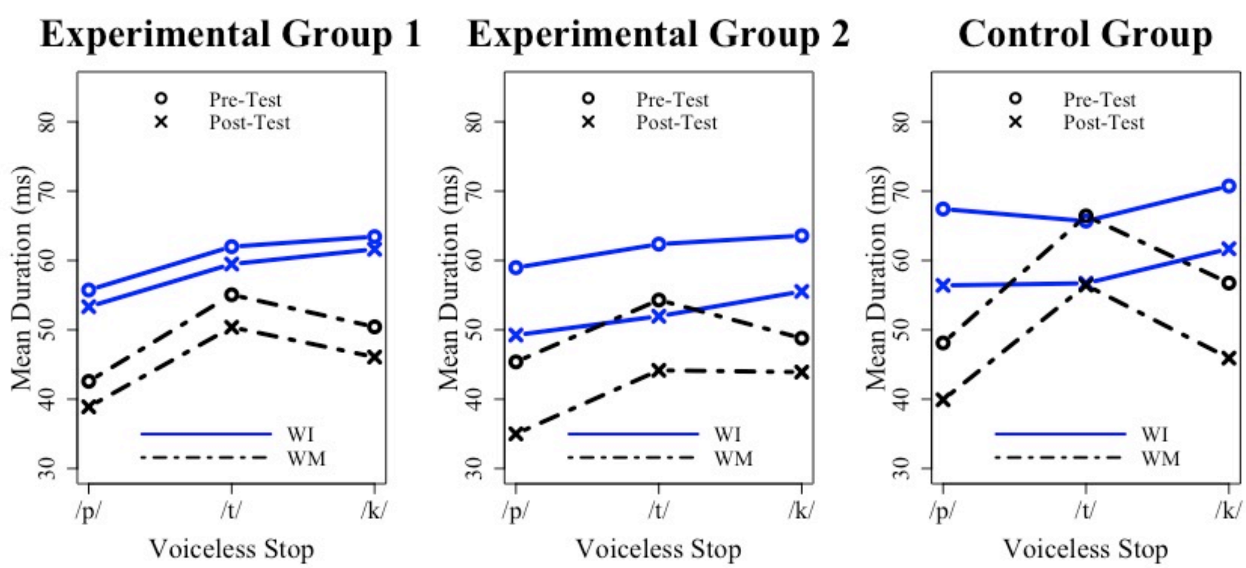
Summary of Paired t-Test Significance by Voiceless Stop and Word Location per Participant Group, Pre-Test to Post-Test

	/p/ WI	/p/ WM	/t/ WI	/t/ WM	/k/ WI	/k/ WM
EG1	nsd	**	nsd	**	nsd	**
EG2	***	***	***	***	***	***
CG	***	***	***	***	***	***

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

Figure 15

Average Voice Onset Time Duration, Pre-Test To Post-Test, by Word Location, by Participant Group



Since the majority of the stop+location combinations demonstrated statistical significance, all inter-group comparisons (Sections 4.2.5 through 4.2.7) consider post-test data only. It is interesting to note that EG1 showed no statistically significant differences for the VOT of any stop in word-initial position over time. This is best illustrated by the close proximity of the two solid blue lines in the word-initial graph in Figure 15, especially when compared to the more dramatic differences between the pre-test average values and the post-test average values for all the other stops and locations. The consistency of a lack of significant difference between all three stops in the same position prompted an investigation into the differences between the participants' VOT values between word-initial and word-medial positioning in order to establish whether or not the participants began the experiment being able to differentiate between the two word positions.

4.2.5 Pre-Test and Post-Test Comparisons between Word Locations for All Groups

A series of t-tests was performed between the word-initial and word-medial values of each of the three voiceless stops (/p/, /t/, /k/) for all three participant groups. Specifically, independent t-tests were utilized because, despite having pre-test and post-test data for each participant, the condition in which the VOT measurements were obtained was not identical between the tests in this case. Both the pre-test and post-test values were analyzed in order to verify whether or not the participants produced VOT differently depending on the location of the stressed syllable (word-initial or word-medial) when they began the semester. Table 46 summarizes the *p*-values for all the t-tests. Figure 15 can also be referenced for an illustration of the average VOT for each stop in each location, and Tables 40 and 41 (see Section 4.2.1) contain the means and standard deviations that informed these t-tests.

Table 43

Summary of Independent t-Test Significance for Intra-Group Comparisons of Word-Initial to Word-Medial VOT, by Participant Group

	EG1		EG2		CG	
	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
/p/ WI ~ /p/ WM	***	***	***	***	***	***
/t/ WI ~ /t/ WM	***	***	***	***	nsd	nsd
/k/ WI ~ /k/ WM	***	***	***	***	***	***

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

As can be seen in Table 43, there were highly statistically significant differences between the word-initial VOT values and the word-medial values for almost every stop at both moments in time. When comparing the values for /p/ between word-initial and word-medial position, highly statistically significantly differences were found in EG1, EG2, and CG in the pre-test (EG1: $t(717.34) = 8.81$, $p < 2.20^{-16}$; EG2: $t(676.54) = 6.85$, $p = 1.67^{-11}$; CG: $t(320.11) = 4.50$, $p = 9.41^{-6}$) and the post-test (EG1: $t(710.74) = 8.75$, $p = 2.20^{-16}$; EG2: $t(680.08) = 9.58$, $p < 2.20^{-16}$; CG: $t(324.69) = 5.85$, $p = 1.17^{-8}$). All three groups also showed highly statistically significantly different values for /k/ in the pre-test (EG1: $t(707.42) = 7.98$, $p = 6.01^{-15}$; EG2: $t(628.23) = 8.44$, $p < 2.20^{-16}$; CG: $t(319.23) = 4.32$, $p = 2.08^{-5}$) and the post-test (EG1: $t(698.14) = 10.09$, $p < 2.20^{-16}$; EG2: $t(645.18) = 7.71$, $p = 4.94^{-14}$; CG: $t(340.82) = 5.66$, $p = 3.18^{-8}$) when comparing word-initial and word-medial VOT. No statistically significant difference in VOT duration between word-initial and word-medial placement was found in CG for /t/, neither in the pre-test ($t(325.11) = -0.203$, $p = .839$) nor the post-test ($t(304.35) = 0.086$, $p = .931$). This lack of difference is represented visually in Figure 15; when looking at the rightmost graph, the middle points (the ‘o’ for /t/ pre-test and the ‘x’ for /t/ post-test) are in nearly identical positions. In contrast, highly statistically significant differences were found for /t/ in EG1 (pre-test: $t(701.59) = 4.32$, $p = 1.78^{-5}$; post-test: $t(733.07) = 4.93$, $p = 1.03^{-6}$) and EG2 (pre-test: $t(665.5) = 4.10$, $p = 4.60^{-5}$; post-test: $t(750.75) = 4.64$, $p = 4.08^{-6}$).

4.2.6 Inter-Group Comparisons (Post-Test Only)

In the previous sections, pre-test data has been compared to post-test data in order to determine whether treatment had an effect over time. Though not all results were statistically significant, there was enough evidence that time impacted the participants’ VOT production that subsequent analyses could remove time as a factor. Therefore, subsequent analyses in this section focus on post-test data only. In order to determine whether there was a difference in VOT duration between the three participant groups, one-way analyses of variance (ANOVAs) were run on the data for each stop in each word location. Subsequently, to discover which group pairings performed statistically significantly different from each other, the Tukey's HSD post-hoc test was calculated.

Five of the six one-way ANOVAs performed on the data indicated statistically significantly different productions of VOT between the three groups included in the calculation. The variance for all three stops in word-initial position between the three participant groups was found to be highly statistically significant (/p/: $F(2, 942) = 10.01$, $p = 5.02^{-5}$; /t/: $F(2, 1107) = 9.99$, $p = 5.01^{-5}$; /k/: $F(2, 963) = 7.94$, $p = 3.79^{-4}$). Conversely, the statistical significance between the groups for the stops in word-medial position was more varied, with /p/ being statistically significant ($F(2, 805) = 4.58$, $p = .011$), /t/ highly statistically significant ($F(2, 842) = 12.93$, $p = 2.96^{-6}$), and /k/ not significant ($F(2, 757) = 1.06$, $p = .346$).

However, despite indications of statistical significance for all but /k/ word-medially, the ANOVAs alone did not pinpoint where (between the three participant groups) there were statistically significant differences in VOT duration. Thus, a Tukey's HSD post-hoc analysis was conducted on the ANOVA data. Table 44 presents an overview of the statistical significance for each stop in each position by group pairing. For reference, Table 41 includes the token counts, means, and standard deviations for each stop in each position for the post-test data, separated by experimental group.

Table 44

Summary of Tukey's HSD Post-Hoc Significance (Post-Test Only) for All Participant Group Pairings

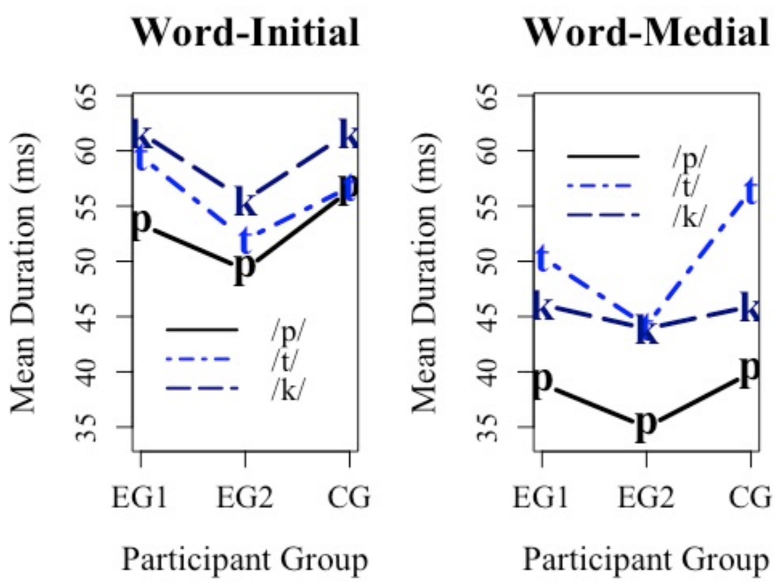
	/p/ WI	/p/ WM	/t/ WI	/t/ WM	/k/ WI	/k/ WM
EG1-EG2	***	*	***	**	***	nsd
EG1-CG	nsd	nsd	nsd	*	nsd	nsd
EG2-CG	**	*	(*)	***	**	nsd

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

The results from the Tukey's HSD post-hoc analysis indicate that statistically significant differences exist between the two experimental groups (EG1 and EG2) for every stop in both positions, with the exception of /k/ in word-medial position. In contrast, only one stop+location pairing showed a statistically significant difference in VOT duration between EG1 and CG. Based on the information in Table 44, there appears to be more separation between EG2 and CG. Figure 16 provides a visual representation of the average of the VOT values for each of the three stops and in both positions, for all three participant groups.

Figure 16

Average Voice Onset Time Values for /p t k/ for Each Participant Group, by Word Location



As can be seen in Figure 16, the VOT values for the three participant groups vary by word location. In word-initial position, the values follow the typical pattern in which VOT increases as the place of articulation moves backward in the mouth, though the averages for /p/ and /t/ in CG are close to each other. In word-medial location, /t/ shows longer average durations than /k/, though for EG2, the values are nearly identical. The differences and similarities depicted in Figure 16 provide a visual representation that allows for comparisons between the participant groups; the Tukey's HSD post-hoc analysis determined whether these differences were statistically significant.

For /p/ in word-initial position, the Tukey's HSD post-hoc analysis revealed that statistically significant differences in VOT production existed between EG1 (*M* = 55.74 ms, *SD* = 20.98) and EG2 (*M* = 49.25 ms, *SD* = 20.01) ($p = 2.18^{-4}$), as well as between EG2 and CG (*M* = 56.41 ms, *SD* = 28.71) ($p = 1.39^{-3}$). The same pattern held for /p/ in word-medial position, with statistically significant differences between the two experimental groups (EG1: *M* = 38.90 ms, *SD* = 19.21; EG2: *M* = 34.99 ms, *SD* = 18.88) (*p* = .031) and between EG2 and CG (*M* = 39.91 ms, *SD* = 22.35) (*p* = .031).

For /t/ in word-initial position, the post-hoc analysis revealed that the only statistically significant group pairing was between EG1 (*M* = 59.48 ms, *SD* = 25.50) and EG2 (*M* = 51.96 ms, *SD* = 24.65) ($p = 2.85^{-5}$). The results when comparing EG2 to CG (*M* = 56.71 ms, *SD* = 26.06) approached significance (*p* = .063). Conversely, statistically significant differences were found in all three group permutations for /t/ in word-medial position. The differences between the VOT productions for EG2 (*M* = 44.16 ms, *SD* = 21.80) and CG (*M* = 56.44 ms, *SD* = 32.38) were highly statistically significant ($p = 2.80^{-6}$). The comparison of EG1 (*M* = 50.38 ms, *SD* = 26.43) to EG2 was statistically significant (*p* = .005), though not as strong as the previous comparison. Finally, the comparison of EG1 and CG demonstrated significance (*p* = .038); for this particular group pairing, it was the only stop+position combination that did so.

For /k/ in word-initial position, the post-hoc analysis indicated that there were highly statistically significant differences between the VOT duration measurements between EG1 (*M* = 61.62 ms, *SD* = 21.66) and EG2 (*M* = 55.54 ms, *SD* = 20.23) ($p = 8.23^{-4}$). Additionally, the differences between EG2 and CG (*M* = 61.69 ms, *SD* = 30.47) were statistically significant (*p* =

.008). However, /k/ in word-medial position, as was mentioned previously, was not shown to be statistically significant from the results of the analysis of variance (ANOVA), which both negated the need for a post-hoc analysis and simultaneously indicated that said analysis would not yield statistically significant results either. Looking at the means for the three groups for /k/ in word-medial position, this is unsurprising, as EG1 (M = 46.08 ms, SD = 19.39), EG2 (M = 43.91 ms, SD = 18.53), and CG (M = 45.92 ms, SD = 21.44) are all within approximately three milliseconds of each other.

4.2.7 Population Inference from Sample Data (Post-Test Only)

Three multiple linear regressions were then performed on the post-test data in order to determine if there was a relationship between VOT production and group membership and/or the location of the stop in the word pronounced (i.e., word-initial or word-medial). Linear regression provides a unique way of interpreting the data, as it represents a more predictive style of analysis; (multiple) linear regression determines whether the independent variable(s) included in the model can (statistically significantly) predict change in the dependent variable (Uyanık & Güler, 2013). Beta weights were also calculated in order to determine the overall effect size of any independent variable, as they illustrate "how much the criterion variable[92] increases (in standard deviations) when the predictor variable[93] is increased by one standard deviation – assuming other variables in the model are held constant" (Glen, Nov. 2016, par. 4). Specifically, "[b]etas show the influence of all the variables on a +1 to -1 scale [. . .]. Those variables that are closer to +1 or -1 exert a greater amount of influence [than those closer to zero]" (Eddington, 2015, p. 92). Additionally, the sign (i.e., positive or negative) of the Beta weight describes the type of relationship between the dependent and independent variable (Eddington, 2015).

[92] i.e., dependent variable (Glen, 9 December 2014, par. 1)
[93] i.e., independent variable (Glen, 5 December 2014)

However, it is important to note that “[t]he higher the absolute value of the beta coefficient, the stronger the effect” (Glen, Feb. 2016, par. 1).

For these regressions, CG and word-medial position were held as constants, meaning that any differences when looking at the experimental groups or word-initial position are in comparison to CG (for EG1 and EG2) and word-medial position (for word location). For CG, this was done in order to compare the two groups that practiced pronunciation to the group that did not. For word location, word-medial position was chosen as the constant (i.e., intercept) because previous research suggests that L2 learners may show improvement in VOT duration in word-medial position prior to word-initial (González López & Counselman, 2013). Because dummy variables had to be used for all five factors within the two independent variables (i.e., the two experimental groups and one control group within the Group independent variable, and the two positions (initial or medial) within the word location independent variable) the “one unit” of change that is typically associated with linear regressions indicates either a change to another experimental group or a change in word location. Additionally, linear regression interpretation assumes that, although there may be multiple factors (i.e., independent variables) (as is the case in the present study), when predicting one factor, the others are held constant (Glen, Nov. 2016).

Table 45

Coefficients for Multiple Linear Regression for /p/

	Unstandardized Coefficients		Standardized Coefficients		
	Estimate	*SE*	*Beta*	*t*	*Sig.*
(Int.)	40.86	1.35	0.00	30.31	$< 2^{-16}$
EG1	−2.13	1.47	−0.04	−1.45	0.148
EG2	−6.13	1.48	−0.13	−4.13	3.82^{-5}
word.location	14.75	1.06	0.31	13.86	$< 2^{-16}$

For /p/, the multiple linear regression equation was highly statistically significant ($F(3, 1731) = 71.37$, $p < 2.20^{-16}$); it had an R^2 value of 0.1101, and an adjusted R^2 value of 0.1085. As can be seen in the estimate in Table 45, VOT decreased by approximately two milliseconds for participants in EG1 (−2.13), by approximately six milliseconds for participants in EG2 (−6.13), and word-initial positioning of /p/ increased VOT by 14.75 milliseconds. However, only EG2 group membership ($p = 3.82^{-5}$) and word location ($p < 2.00^{-16}$) were statistically significant predictors of change in VOT duration. Looking at the Beta weights (in Table 45) and considering their absolute values, while no variable had a particularly strong effect, word location had a larger effect on VOT duration than group membership generally, and membership in EG2 had a larger effect on VOT duration than membership in EG1.

Table 46

Coefficients for Multiple Linear Regression for /t/

	Unstandardized Coefficients		Standardized Coefficients		
	Estimate	*SE*	*Beta*	*t*	*Sig.*
(Int.)	52.67	1.48	0.00	35.67	$< 2^{-16}$
EG1	−1.03	1.60	−0.02	−0.64	0.52
EG2	−7.99	1.61	−0.15	−4.96	7.50^{-7}
word.location	6.88	1.17	0.13	5.88	4.85^{-9}

The multiple linear regression equation was also highly statistically significant for /t/ ($F(3, 1951) = 24.35$, $p = 1.80^{-15}$), with an R^2 value of 0.036 and an adjusted R^2 value of 0.035. Table 46 shows the coefficients for the multiple linear regression, which indicate that for /t/, the VOT duration decreased by approximately one millisecond in EG1, decreased by approximately eight milliseconds in EG2, and that the stop being in word-initial position increased duration by nearly seven milliseconds. As was seen for /p/, only the treatment for EG2 ($p = 7.50^{-7}$) and the location of the stop in words ($p = 4.85^{-9}$) were statistically significant predictors of VOT duration change.

The Beta weights (in Table 46) indicate that despite the small effect size overall, for /t/, EG2 membership had the largest effect size out of the three variables (considering absolute values), followed closely by word location, while EG1 membership had a very small effect.

Table 47

Coefficients for Multiple Linear Regression for /k/

	Unstandardized Coefficients		Standardized Coefficients		
	Estimate	*SE*	*Beta*	*t*	*Sig.*
(Int.)	46.85	1.30	0.00	36.14	$< 2^{-16}$
EG1	0.04	1.42	0.001	0.03	.977
EG2	−4.33	1.43	−0.09	−3.02	.003
word.location	14.08	1.05	0.31	13.41	$< 2^{-16}$

The multiple linear regression equation was found to be highly statistically significant for /k/ ($F(3, 1722) = 65.37$, $p < 2.20^{-16}$), with an R^2 value of 0.1022 and an adjusted R^2 value of 0.1007. As was the case for the other stops, VOT duration decreased by approximately four milliseconds in EG2, and increased by approximately 14 milliseconds when the stop was in word-initial position (as compared to word-medial), as can be seen in Table 47. However, unlike /p/ and /t/, for /k/, the VOT duration increased very slightly for EG1. Nevertheless, only the treatment for EG2 ($p = .003$) and word location ($p < 2.00^{-16}$)were shown to be statistically significant. The interpretation of the Beta weights for /k/ (in absolute values) indicated that word location had a much larger effect on change in VOT duration within the /k/ dataset than either of the two treatments; however, though quite small, membership in EG2 had a much larger impact on the dataset than the negligible effect of membership in EG1.

Table 48 presents a summary of the output from the multiple linear regression statistical tests performed on the VOT data. Included are the amount and direction of predicted change, the impact of the independent variable on the dataset (i.e., beta, interpreted as absolute values), and a representation of the statistical significance of the calculation.

Table 48

Summary of Multiple Linear Regression Significance with Amount of Predicted Change in VOT and Impact, by Experimental Group and Word Location

	EG1			EG2			word.location (WI)		
	Predicted Change	Impact (Beta)	Sig.	Predicted Change	Impact (Beta)	Sig.	Predicted Change	Impact (Beta)	Sig.
/p/	−2.13 ms	−0.04	nsd	−6.13 ms	−0.13	***	+14.75 ms	0.31	***
/t/	−1.03 ms	−0.02	nsd	−7.99 ms	−0.15	***	+6.88 ms	0.13	***
/k/	+0.04 ms	0.001	nsd	−4.33 ms	−0.09	**	+14.08 ms	0.31	***

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$; nsd: no significant difference; (*): approaches significance

At first glance, Table 48 clearly illustrates that membership in EG2 allows for predictions of greater decreases in VOT duration across all three voiceless stops. Additionally, the voiceless stops being in word-initial position (compared to word-medial) predicts a large increase in VOT duration. In contrast, membership in EG1 predicted very little change in VOT duration.

4.3 Articulatory and Acoustic Phonetics Discussion

4.3.1 Implications for L2 Acquisition (of Pronunciation) and Instruction of Grapheme-Phoneme Mismatches

The present study has implications for the manner in which pronunciation is approached in the field of second language acquisition generally, and more specifically in the L2 classroom. First, analyzing accuracy from the standpoint of odds of a target response allows for the creation of levels of priority for pronunciation elements; this idea has been referred to as a *hierarchy of errors*, though often in reference to prioritization based on native-speaker ratings of non-native speech (e.g., Agostinelli, 2012; Schairer, 1992). Previous research has investigated teachers' attitudes toward pronunciation and the teaching of pronunciation, finding that teachers may avoid pronunciation (or give it a low priority) due to their lack of comfort with teaching pronunciation (Derwing, 2008). An analysis based on likelihood (greater/smaller odds) and the direction of the change in odds can inform where more or less time should be spent, especially if the learners are consistently practicing pronunciation (either in or outside of class). Moreover, the instances in which there was no significant difference in the odds may be just as (or more) informative than where significance was found.

Based on the findings of the present study as a whole, the grapheme-phoneme mismatches could be organized into levels of priority based on the odds of a target response combined with statistical significance, or lack thereof. Looking at each grapheme individually, <ll> appears to be the most resistant to improvement, as the odds of a target response decreased significantly regardless of the method of pronunciation practice, and thus would be ranked as a higher priority. Additionally, <j> and <ch> should be given higher priority. For <j>, the change in the odds of a target response for the two experimental groups went in different directions (increase for EG1 but decrease for EG2), though neither change was statistically significant. For <ch>, both experimental groups showed a decrease in the odds of a target response; however, neither was statistically significant. Organizing the grapheme-phoneme mismatches into a hierarchy of relative importance would provide guidelines to focus pronunciation practice and/or instruction on the areas that require more overt assistance.

Prioritizing grapheme-phoneme mismatches based on L1 interference is also recommended and supported by the findings of this study, showing that three groupings of graphemes would be given a higher priority. The first grouping contains graphemes that, in Spanish, map to phonemes which are not part of the English (L1) phonological inventory, in this case <g{i, e}>, <ll>, <ñ>, and <j>, which map to /x/, /ɟ/, /ɲ/, and /x/, respectively. Since these sounds are not part of the phonological inventory of American English (Odden, 2005, p. 148), one would expect them to be perceived, decoded, and then assigned to the most similar phonetic category (Peperkamp & Dupoux, 2003). However, the addition of English orthography complicates the grapheme-phoneme mapping process; instead of non-native phonetic input simply being received and amended to fit within the L1 phonological inventory (Silverman, 1992), the presence of the L1 orthographic representation misleads the speaker into mapping onto a phoneme that is much less related (or even unrelated) to the original phoneme. This is often the case with the aforementioned grapheme-phoneme pairings. For instance, <g{i, e}> and <j> both map to the voiceless velar fricative, /x/. To adapt a foreign sound to the L1 inventory in the most economical way and with as little loss of phonological information as possible, the transformation requiring the fewest steps is preferred (Paradis & LaCharité, 1997, pp. 386–387). In the case of /x/ and American English as the L1, this would point to /h/, as the next closest fricative (in place of articulation) to the velar /x/ common to both American English and Spanish is /s/ which would require a change from non-sibilant (e.g., /x/, /h/) to a sibilant (e.g., /s/) (Ladefoged, 1975, p. 146), making /h/ the more economical choice. However, in English, the grapheme <g{i, e}> maps to /ʒ/, /d͡ʒ/, or /g/, and <j> maps to /d͡ʒ/, /j/, or /h/ which obscures the selection of the best choice of phoneme or removes it as an option altogether. This potential for interference from the orthography of the L1 makes the graphemes related to sounds not found in the L1 inventory strong candidates for explicit explanation or instruction, ranking them as a higher priority.

Based on a lack of statistically significant change and the direction of the odds of a target response, <c{a, o, u}>, <g{a, o, u}>, and <ll> would be the third grouping to fall into the higher priority category. As was mentioned previously, <ll> would be a very high priority due to the statistically significant decrease in odds found in the present study. For both treatment groups, the likelihood of a target response to <c{a, o, u}> decreased (though not significantly), and the odds of a target response decreased slightly in EG1 but increased slightly in EG2 (again, not significantly). However, unlike the other graphemes designated as higher priority, <c{a, o, u}> and <g{a, o, u}> have a more transparent correspondence between grapheme and phoneme in both languages; the graphemes also correspond to the same phonemes in both languages. It is possible that the decrease in odds was caused by overcompensation, which has been the case in previous L2 acquisition of pronunciation studies (Flege & Hillenbrand, 1984; Flege et al., 2003; Piccinini & Arvaniti, 2015). By requiring the participants to listen in order to focus on pronunciation (either by using the speech recognition functionality of the app or by comparing their recordings to those of a native speaker), the participants may have been so attuned to the fact that they were supposed to find differences in pronunciation between the two languages that they began seeing errors that did not exist. Thus, it could be beneficial to reinforce, briefly, that some consonants are pronounced the same in the L1 and L2. This may reassure learners that not all consonants are problematic, and (at least between English and Spanish) their intuition, influenced by the L1, will assist them much of the time. The time spent on these consonants would be less than the others in the higher priority grouping.

The findings of this study indicate that the selection of graphemes for more explicit instruction should be influenced by theories such as contrastive analysis. This is demonstrated by the discovery that all the graphemes that would be given higher priority individually were subsumed by the list of graphemes created when considering the results from the analyses from the perspectives of the L1 inventory and L1 options. Based on the present findings, higher priority should be given to those whose corresponding phonemes in Spanish are not found in the L1 (English) inventory (<g{i, e}>, <ll>, <ñ>, <j>). From the perspective of number of L1 grapheme-phoneme associations, the present findings indicate that more focus should be placed on those which correspond to one pronunciation option in the L1 (<c{a, o, u}>, <g{a, o, u}>, <ll>), three options in the L1 (<g{i, e}>, <j>), and four options in the L1 (<c{i, e}>, <ch>). This grouping of graphemes is slightly more extensive than the one created simply by considering each grapheme individually, but each analysis producing similar groups of graphemes reinforces the usefulness of approaching L2 acquisition of pronunciation from the standpoint of the language(s) already spoken by the learner, particularly for adult beginners who have the capability to be more meta-linguistically aware. Therefore, comparisons between the L1 and L2 graphemic and phonological inventories ought to be explored by both teachers and students, as well as encouraged during the acquisition process. Moreover, these kinds of comparisons could assist instructors in narrowing potentially problematic pronunciation elements down to those that are most resistant to improving simply with extra practice. Utilizing this type of analysis to inform decisions on inclusion or exclusion of consonants for instruction would combat the reluctance of instructors to teach pronunciation due to deficits in their own understanding (Derwing, 2008) as they could be provided a specific set of consonants to focus on, as well as a ranking within the set by importance.

Approaching language acquisition from the perspective of contrastive analysis also provides the opportunity to explore form in addition to meaning, as proficiency in L2 pronunciation has the capability to impact (positively or negatively) communication (Hurtado & Estrada, 2010; Martinsen et al., 2014; McBride, 2015), where pronunciation focuses more on form and communication focuses on meaning. Decades previously, Brumfit (1984) specifically referred to pronunciation and orthography, especially atypical cases in the L2, as examples of instances in which focusing on form could be beneficial and augment communication.

Moreover, Martinsen et al. (2014) view pronunciation practice and instruction as tools for growth that should extend beyond simply creating comprehensible output as a goal. Extrapolations based on the findings of the present study can be considered evidence that there ought to be a middle ground between intelligibility and comprehensibility. Some of the grapheme-phoneme mismatches investigated showed great improvement with the extra pronunciation practice, while others did not. Therefore, the middle ground should focus on comprehensibility in the classroom supplemented by a focus on form in specific instances such as the grapheme-phoneme mismatches that have been shown to be more resistant to improvement through autonomous practice or training.

Identifying problematic pronunciation elements through comparisons between the L1 and L2 (i.e., contrastive analysis) also appears to fit within current L2 proficiency evaluation criteria, and would allow instructors to maximize the effectiveness of their pronunciation explanation or instruction within the pedagogical framework currently implemented in many second language classrooms. Despite the deprioritization of pronunciation in the majority of the ACTFL Proficiency Guidelines, grapheme-phoneme mismatches arguably fit within the general category of L1 transfer or influence, which the guidelines mention specifically (Lee & VanPatten, 2003; ACTFL, 2012). Additionally, approaching pronunciation from the standpoint of contrastive analysis addresses the fourth *C* in the ACTFL World Readiness Standards, *Comparisons*, which includes metalinguistic awareness (i.e., making connections between the L1 and L2) as a goal for proficiency (ACTFL, 2015), even though there appears to be no mention of pronunciation. Therefore, it could be argued that pronunciation, especially approached from a contrastive analysis standpoint, does fit within the communicative approach.

The differences in performance between the two experimental groups may also reflect different implementations of current pedagogical trends. One possible reason why EG2 outperformed EG1 could have pertained to the debate between focusing on form or meaning. It has been argued that the communicative approach, a popular pedagogy currently, places high value on communicative competency (Wong & VanPatten, 2003; Hurtado & Estrada, 2010; Martinsen et al., 2014), but less on pronunciation (Brumfit, 1984), only briefly mentioning it in terms of evaluation criteria for fluency (ACTFL, 2012; Arteaga, 2000; Lee & VanPatten, 2003), and/or favors suprasegmentals if pronunciation is focused on at all (Lightbown & Spada, 2013, p. 68). However, Trofimovich and Gatbonton (2006) found that the requirement to focus on meaning was detrimental to pronunciation accuracy, especially for participants whose accuracy was already low. As was mentioned previously, the activities provided to the participants in EG2 (i.e., listen-and-repeat style) were included within the set of homework activities that accompanied the textbook for first-semester Spanish classes, *Protagonistas* (Underwood et al., 2012). These activities required participants to record themselves and then compare their own speech to a recording of a native speaker; some of the activities directed participants to focus on particular graphemes and/or sounds, while others had a more communicative goal. In contrast, the app utilized by EG1 provided the opportunity to respond using speech recognition most of the time, but pronunciation was rarely the focus of any activity; these activities focused on concepts (e.g., telling time, *ser* 'to be' versus *estar* 'to be'), and using speech recognition was one of the response options given, which provided an opportunity for pronunciation practice, but it was not the focus or goal of the activity. The claim made by Trofimovich and Gatbonton (2006) could explain the gap in improvement between the two experimental groups: the app's activities focused mainly on meaning, while some of the listen-and-repeat style activities directed the participants' focus to the form of specific pronunciation elements.

Even though there was a partial focus on form, pronunciation practice/training with the listen-and-repeat style activities (included with the course textbook, and utilized by EG2) produced uneven results; thus, a comparison between the grapheme-phoneme mismatch findings and the textbook-provided pronunciation activities could provide insight into the uneven results. While overall, EG2 outperformed EG1 and CG, statistically significant improvement was not seen in every grapheme investigated. For EG2 specifically, there were seven graphemes that did not show statistically significant improvement over time: <c{a, o, u}>, <c{i, e}>, <g{a, o, u}>, <ll>, <j>, <ch>, and <z>; four of these graphemes also demonstrated a decrease in the odds of a target response: <c{a, o, u}>, <ll> (significantly), <j>, <ch>. However, when compared with the topics and contents of the textbook-included pronunciation activities, the graphemes designated in this study as grapheme-phoneme mismatches are in the minority and the attempt at an overview of pronunciation likely negatively impacted pronunciation improvement. The complete breakdown of the pronunciation activities included in the online companion to the *Protagonistas* textbook (Underwood et al., 2012), including identification of activities containing explicit instructions and activities pertaining to the investigated graphemes, can be found in Appendix A, Table A5.

In total, 85 pronunciation activities were included in the guidelines provided to the participants in EG2 (see Appendix I). All of the activities required the participants to pay attention to their pronunciation *generally*, but of those 85, only 37 activities (44%) placed *explicit* focus on the graphemes, meaning that the instructions clearly (i.e., explicitly) stated that students should pay attention to or focus on specific grapheme(s). The remainder of the activities included instructions, but they did not identify and direct participants to focus on specific sounds. Within the 37 activities, graphemes were explicitly mentioned in the directions 77 times in total; 34 (or 44%) of these were the consonants investigated in the present study as grapheme-phoneme mismatches, while 43 (or 56%) were other consonants. Simply from an overview of the activities, it is already clear that the grapheme-phoneme mismatches did not dominate the pronunciation activities.

The breakdown of pronunciation activities that included explicit pronunciation instructions provides insight into the performance of EG2 when compared with the graphemes (<c{a, o, u}>, <c{i, e}>, <g{a, o, u}>, <ll>, <j>, <ch>, <z>) for which there was no statistically significant improvement over time and/or a decrease in the odds of a target response. Out of these graphemes, the most striking is <ll>, for which there was a statistically significant decrease in the odds of a target response as well as no significant improvement over time. This result is less surprising after discovering that out of the 37 pronunciation activities with explicit instructions, only two focused on the grapheme <ll>. Similarly, there was only one explicit mention of <z> (which did not show significant improvement over time) in instructions. For <c{a, o, u}>, the situation is murkier because there was not always a differentiation between <c{a, o, u}> and <c{i, e}>, though by looking at the names of the activities, only one clearly states that it focuses on <c> and vowels. The last grapheme (<j>) in the group that showed neither significant improvement over time nor an increase in odds is more difficult to explain, as it had the most instances of appearing in explicit instructions. However, based on the activity titles, <j> is always included with other graphemes; it is possible that participants would have benefited from practicing <j> by itself as well as with other graphemes, indicating that more repetitions may not be as helpful without clear explanation to begin with.

The investigation into the pronunciation activities that accompanied the textbook also shed light on the type of focus placed on pronunciation, which appeared to be simultaneously general and nuanced. Evidence for a more general focus would include the observation that the majority of the Spanish consonants were included in at least one activity containing explicit instructions. Additionally, topics such as intonation, syllable/word stress, diphthongs, and linking (i.e., synalepha) were included within the 85 pronunciation activities, adding to the overview of Spanish pronunciation (i.e., many topics with few repetitions of each). In contrast, some of the pronunciation activities focused on very nuanced elements, which, one could argue, would cause fewer issues with comprehensibility and/or intelligibility than ones that could change the entire meaning of a word (e.g., *llamo* 'I call' (1SG-PRS) [ˈɟa.mo] and *lamo* 'I lick' (1SG-PRS) [ˈla.mo]). For example, included in the activities for 'm, n' was information regarding the assimilation of nasals to the place of articulation of the subsequent sound (Underwood et al., 2012), which could be considered quite advanced for first-semester learners. Also considered very advanced would be activities covering linking (or synalepha), which were included within activities about vowels, in which the focus was on producing a longer vowel (e.g., *mi hijo* 'my son' as [ˈmi:.xo][94]) instead of as two separate instances of the same vowel. Additionally, while it is true that [v] can occur in Spanish due to voicing assimilation of /f/ (Hualde, 2014; Hualde et al., 2010), the lack of a phonemic contrast could support the argument that sounds that do contrast phonemically should be prioritized. While providing an overview of Spanish pronunciation could be useful, the findings of the present study support a more targeted approach, particularly favoring phonemic contrasts and possible orthographic interference over more advanced and nuanced topics that would not cause deterioration of intelligibility and

[94] Example from "Diphthongs and Linking" activity from *Protagonistas* online activities (Underwood et al., 2012)

comprehensibility. The analysis of the textbook-provided pronunciation activities could also provide evidence for the argument that while EG2 did show the most improvement, a likely culprit for the uneven improvement would be the lack of a clear focus and attempt to introduce too many pronunciation elements without enough practice of each.

4.3.2 Implications for Second Language Acquisition (of Pronunciation) and Instruction of Voice Onset Time

In contrast to the grapheme-phoneme mismatches, contrastive analysis cannot explain the findings pertaining to VOT. Even though there has been research investigating VOT in the foreign language classroom generally (Counselman, 2015; Kissling, 2013, 2015; González López & Counselman, 2013; Offerman & Olson, 2016; Olson, 2014) as well as with explicit instruction as part of the methodology (González López & Counselman, 2013; Kissling, 2015; Counselman, 2015; Offerman & Olson, 2016), direct comparison between the L1 (English) and the L2 (Spanish) is much more subtle for VOT, especially because the presence or absence of aspiration does not change the meaning of words in either language. As has been mentioned previously, the accepted ranges of VOT values for English and Spanish are mutually exclusive (Keating, 1984; Lisker & Abramson, 1964; Zampini, 2008), and while previous studies have noted improvement (i.e., reduction) in VOT length after instruction (González López & Counselman, 2013; Kissling, 2015), it is important to note that the resulting VOT values did not fall within the accepted range for the L2 (Flege & Hillenbrand, 1984). The present study has similar findings to those of Flege and Hillenbrand (1984); while VOT values decreased overall, they did not consistently decrease enough to fall within the acceptable range for Spanish.

While an overall decrease in VOT was found in all three participant groups over time, EG1 demonstrated an interesting pattern depending on the location of the voiceless stop within the target word. For this particular group, no statistically significant difference was found between the duration of VOT between the pre-test and post-test for any of the three voiceless stops when they occurred in word-initial position. Conversely, the differences for word-medial VOT were quite statistically significant. This pattern was also found by González López and Counselman (2013), who observed that "[t]he significant effect of the factor 'site' [. . .] may indicate that learners improve their VOT production of voiceless stops in medial-position first" (p. 124). As EG2 consistently outperformed EG1, it is possible that the participants in EG2 improved more rapidly, and given enough time, those in EG1 would catch up. Although CG also demonstrated statistically significant improvement over time, their overall duration values did not decrease as much as the other groups, though compared to themselves the improvement was significant.

Pertaining to only the post-test data, as was found for the grapheme-phoneme mismatches, membership in EG2 played a significant and consistent role in VOT improvement, not only between EG2 and CG, but also between both experimental groups. The post-hoc analysis (Tukey's HSD) of the ANOVA conducted on the VOT durations from all three groups indicated that the most significant difference (i.e., gap) in VOT duration existed between the two experimental groups, with all three voiceless stops showing highly statistically significant differences in word-initial position, and significant differences for two of the three stops (/p/ and /t/) in word-medial position. Similar differences were found between EG2 and CG. Conversely, the only statistically significant difference found between EG1 and CG was found between the VOT durations for /t/ in word-medial position. These differences place EG1 slightly closer to EG2 in terms of improvement, but much closer to CG. The combination of these findings and those over time indicate that while the use of each type of technology (i.e., app with speech recognition, listen-and-repeat style activities, or grammar activities with no pronunciation element) produced overall improvement in VOT duration when compared to themselves, comparison across groups revealed that the listen-and-repeat style of activities produced more significant change in VOT duration than the other two technologies.

4.3.3 Inferred Impact of Group Membership and Word Location on L2 Learner Population

The linear regression conducted on the data allows a prediction to be made for the population as a whole based on the sample in the present study. Similar to the logistic regression, which focused on the odds of obtaining a particular response, the linear regression produced the amount of change (in milliseconds) predicted for each experimental group (as compared individually to the control group) and for word-initial VOT (as compared to word-medial). The output from the linear regressions produced two results: first, how much technology impacted the participants' VOT (and in which direction), and second, how much the position within the words affected VOT (and in which direction). As was the case for the logistic regressions performed on the grapheme-phoneme mismatches, this data can be used to inform decisions made regarding the need for explicit instruction.

For VOT, explicit pronunciation instruction would be recommended for the instances in which the participants' predicted response either was not statistically significant or demonstrated a trend toward larger VOT values (as a decrease in milliseconds demonstrates improvement). Based on the results of the present study, the most striking, and most consistent, area that would benefit from explicit instruction on VOT is word location. For all three voiceless stops, the linear regressions indicated large increases in VOT duration (between approximately seven and 15 milliseconds) when word-initial VOT was compared to the "baseline" word-medial VOT, all of which were highly statistically significant. This predicted difference could indicate that the participants were not attending to the pronunciation of the voiceless stops as much when the stops began a word; instead, the participants may have been focusing on the upcoming sounds within the word, especially since <p> (/p/), <t> (/t/), and <c{a, o, u}> (/k/) are already familiar to L1 English speakers, and the difference between the sounds in the L1 versus the L2 is much more subtle than the differences between some of the other elements they might encounter within a word. The prediction that VOT duration would be significantly longer for voiceless stops in word-initial position also appears to support previous claims that word-medial VOT duration may improve before word-initial VOT duration (González López & Counselman, 2013). Moreover, a comparison of the Beta weights for word location to those for either group type demonstrates that word location has a larger overall impact on the increase or decrease in VOT duration than group membership.

As was seen for the grapheme-phoneme mismatches, the type of technology used by the participants also affected the change in VOT. For all three voiceless stops, EG2 demonstrated statistically significant reductions in predicted VOT duration in comparison to CG. Conversely, the predicted VOT values for EG1 decreased for /p/ and /t/, though not significantly; however, for /k/, the predicted VOT value increased slightly. A comparison of the Beta weights for both experimental groups revealed that, despite the effect sizes being small, the listen-and-repeat style of activity utilized by EG2 consistently had a larger impact on VOT duration. For /p/, the effect size of membership in EG2 was approximately three times larger than that of EG1. For /t/, the effect size of EG2 membership was approximately seven times larger, and for /k/, it was approximately ten times larger.

Pronunciation training, or at the very least sustained focus on pronunciation, had an impact on the VOT production of the participants, despite the lack of explicit instruction. Pre-test to post-test, there was a statistically significant difference in VOT duration for all three voiceless stops in word-medial position for EG1, and in both positions (initial and medial) for EG2. Despite having taken part in no extra pronunciation practice there was also a statistically significant difference over time for CG. These findings fit with what has been found previously, but they also stand out due to one specific difference: instruction. Many studies have found an improvement in VOT duration (i.e., in the case of English and Spanish, a decrease in duration) after participants receive instruction; the majority of this instruction focused on articulatory and/or acoustic phonetics (or elements therein) and is typically explicit (Aliaga-García, 2007, 2017; Alves & Luchini, 2017, 2020; Alves & Magro, 2011; Camus, 2020; Counselman, 2015; González-Bueno, 1997a; González López & Counselman, 2013; Kissling, 2013; Yang, 2017). In contrast, a small amount of studies investigate VOT in immersion settings (Hanzawa, 2018; Harada, 2007; Netelenbos et al., 2016; Rao et al., 2020) including ones utilizing content-based L2 instruction (Hanzawa, 2018), or tracking improvement in VOT simply from studying Spanish over time (Nagle, 2019); in these studies, there is little if any focus on pronunciation itself. The present study, however, cannot be classified as either explicit instruction or immersion. While the participants were instructed to either use the speech recognition function whenever it was available (for the app group) or complete pronunciation activities that would require the use of a microphone and perhaps headphones (for the listen-and-repeat style group), they were never given specific information about VOT (or grapheme-phoneme mismatches), nor were they in an immersion setting, as the typical exposure to the L2 (Spanish) for the participants was three 50-minute classes per week. Thus, the findings from the present study appear to indicate that improvement in VOT duration can be achieved at beginning levels of acquisition without VOT being the focus of pronunciation practice or instruction. While pronunciation improvement might stagnate at some later point in the acquisition process, at which point more explicit instruction would be beneficial (Martinsen et al., 2014), consistent practice and/or training in L2 pronunciation was sufficient to affect a change in the participants' VOT duration.

4.3.4 Impact on Current Technological Implementation

The findings of the present study indicate that the listen-and-repeat style of activities included with the textbook for first-semester Spanish courses had a greater impact on grapheme-phoneme mismatch accuracy than the mobile application which had speech recognition capability and provided a minimal amount of feedback on pronunciation accuracy. It is important to note, however, that though there was less improvement overall by participants interacting with the app (EG1), it is not true that there was no improvement. The overall improvement found in both experimental groups, combined with the evidence that at least one of the experimental groups consistently outperformed CG (though not always statistically significantly), indicates that pronunciation practice (or training) generally improves accuracy in navigating grapheme-phoneme mismatches. However, since the type of technology used produced different results, the type of technology (and the type of activities therein) should be taken into account and inform decisions regarding explicit explanation and/or instruction.

The style and/or type of technology should be considered when selecting practice (i.e., training) activities. If pronunciation practice is assigned to L2 learners, based on the results of the present study, the style of the practice activities should also be taken into account when deciding how the grapheme-phoneme mismatches are approached in a classroom setting in order to maximize the efficacy of pronunciation training/practice. As has been mentioned previously, the app that EG1 interacted with used speech recognition to provide binary (i.e., correct or incorrect) feedback to the participants; the participants were instructed to utilize the speech recognition feature when it was available (which was in the majority of activities). However, when interacting with the app, participants' focus was divided between the pronunciation and the concepts being presented. Conversely, the participants in EG2 completed activities that were more closely focused on pronunciation; the activities varied, sometimes asking participants to listen to and repeat single words, and other times whole sentences, but they were mostly focused on simply practicing pronunciation. This difference likely impacted the improvement of the participants.

Using the present study's findings as a template, the majority of the grapheme-phoneme mismatches should be given additional attention (i.e., explicit instruction) if the practice activities were styled like the app. Participants using the app only demonstrated a significant increase in the odds of a target response to <c{i, e}>; the other nine graphemes were divided nearly equally between an increase or decrease in the odds of a target response (five decreased and four increased). Thus, the recommendation based on the present findings would be for more explicit instruction of grapheme-phoneme mismatches in general prior to interaction with the app. In terms of a hierarchy of priority, the graphemes that resulted in a decrease in odds g(<c{a, o, u}>, <g{a, o, u}>, <g{i, e}>, <ll>, and <ch>) would be prioritized over those that resulted in an increase in odds (<h>, <ñ>, <j>, and <z>), though explicit instruction could be beneficial for all nine graphemes. As has been mentioned previously, within the higher priority grouping, <ll> would be given the highest priority due to a statistically significant decrease in the odds of a target response.

Conversely, if the practice activities focused more specifically on pronunciation (i.e., the listen-and-repeat style of activities), a more targeted approach to grapheme-phoneme mismatches could be taken. In the present study, participants using this type of activity showed statistically significant increases in the odds of a target response to five graphemes (<c{i, e}>, <g{i, e}>,

<h>, <ñ>, and <z>); the odds of a target response also increased for <g{a, o, u}>, though not statistically significantly. Thus, prior to practicing on their own, students could be introduced to the concept of grapheme-phoneme mismatches (though possibly in less technical terms), but explicit instruction could be limited to the graphemes for which there was a decrease in the odds of a target response in the present study (<c{a, o, u}>, <ll>, <j>, and <ch>), with priority being given to <ll> within the grouping, as there was a statistically significant decrease in odds in response to that particular grapheme.

The trends seen in VOT duration underscore what was found for the mismatches: the need for explicit instruction may vary based on what type of technology is being used for pronunciation practice. Drawing on the findings from the present study, it would be beneficial to call students' attention to the voiceless stops and the differences between the L1 and the L2 (and to /k/ in particular) if they were going to interact with technology styled more like the app used in this study (which had speech recognition capability but also focused on more than just pronunciation in the activities). If the activities were more focused on pronunciation alone, like the listen-and-repeat style activities given to EG2, there could be less of a need for explicit instruction. However, some type of intervention would likely be needed at a certain point in order for the learners to continue progressing, especially as their fluency increased (Martinsen et al., 2014).

4.3.5 Impact on Future Technological Development and Implementation

As fields such as artificial intelligence and natural language processing develop, these technological advances should continue to be utilized to aid in language acquisition. Apps such as Duolingo, Babbel, and Rosetta Stone incorporate technological innovations (as did the Spanish SOLO app), and textbooks published for use in foreign language classrooms have also begun incorporating these innovative elements, such as speech recognition (Blanco, 2016). However, even mainstream speech recognition systems such as Alexa or Siri have issues with inaccuracies and glitches, and the technology included with some L2 textbooks and/or developed for L2 acquisition is typically less advanced and/or accurate (McCrocklin, 2016).

Beyond implications for language learning within and/or outside of the classroom, the findings of the present study should also be taken into account when making improvements on language learning ASR systems. McCrocklin (2016) acknowledged that the ASR software she utilized in her study needed improvement, particularly in word recognition, and Kolesnikova (2017) proposed improving identification accuracy by adding phonetic features to the information contained in ASR databases, which could then be utilized to provide better feedback. In fact, in the survey of foreign language learning technology conducted by Golonka et al. (2014), one of the few elements that had a positive impact on language learning was speech recognition functionality and its capacity for providing feedback on pronunciation, though, as Kolesnikova (2017) later reiterated, the feedback was only as accurate as the knowledge base provided to the ASR program. To remedy this, Kolesnikova (2017) recommended providing information regarding common pronunciation errors to the database(s) used by ASR systems. The findings of the present study could be incorporated into the knowledge database(s) of existing language learning ASR programs in a similar manner to the one proposed in Kolesnikova (2017).

The results of the present study would positively impact the amendment of ASR databases (as proposed by Kolesnikova (2017)) from a language acquisition standpoint as well as a technological development standpoint. First, databases of knowledge for ASR systems provide a unique opportunity to utilize contrastive analysis. For instance, if users (of a given ASR program) selected English as their L1 and Spanish as the L2, then the program would be able to pull information from both databases. Then, minimally, commonalities between the two languages could be pinpointed and potentially minimized in favor of more divergent elements. Additionally, if the program were provided this knowledge, information regarding common errors and their causes could be readily available to users, as well as extra practice activities (Kolesnikova, 2017) which could be created based on linguistic elements found to be problematic both generally and between the selected L1 and L2.

Making improvements to consonant recognition based on information such as grapheme-phoneme mismatches between two languages (i.e., contrastive analysis) would also be beneficial because it would help to move ASR away from making decisions based mostly (if not solely) on statistical probability, algorithms, or other mathematical functions. Presently, one of the most common metrics used in ASR is Goodness of Pronunciation (GOP), an "algorithm [that] calculates the likelihood ratio that the realized phone corresponds to the phoneme that should have been spoken according to the canonical pronunciation" (Kanters et al., 2009, p. 49). A GOP score is calculated for each sound in a word and compared to a predetermined threshold to determine if the pronunciation was correct (Kanters et al., 2009; Odriozola et al., 2012). The issue with utilizing this kind of prediction model is that the overall evaluation of a word could still be "correct" despite there being glaring issues in pronunciation. For example, the Spanish word *llamo* 'I call' (1SG-PRS) [ˈʝa.mo] contains four sounds, two of which are very similar to English, with the third being identical. If it were assumed that the probability of "correct" pronunciation would be high for sounds that overlap between the two phonological inventories (even with small differences between the English and Spanish vowels), that would only leave one sound that could negatively impact the overall evaluation of the word; for an L1 English speaker, a common mispronunciation in this case would be to pronounce <ll> as [l]. Although *lamo* 'I lick' (1SG-PRS) [ˈla.mo] is a real word, and exemplifies that /l/ and /ʝ/ are separate phonemes through this minimal pair example, an evaluation of *llamo* as *[ˈla.mo] would likely still be evaluated as "correct" overall due to the other sounds having high probabilities of "correct" pronunciations. In other words, the mispronunciation of one sound out of the four might not detract from the overall evaluation enough for the word as a whole to be considered to be pronounced "incorrectly". However, if information such as common mispronunciations (Kolesnikova, 2017) and/or data from contrastive analyses (such as the grapheme-phoneme mismatches investigated in the present study) could be added to databases that are used to train ASR systems, more weight could be given to errors like the one in the aforementioned example so that pronouncing *me llamo* 'I call myself' as *me lamo* 'I lick myself' would be marked as "incorrect" by the ASR system, in spite of the pronunciation of the other sounds in the word.

The proposed additions to ASR databases create an opportunity to improve and/or further develop feedback given in response to L2 pronunciation utterances. In their review of more than 350 studies, Golonka et al. (2014) found that of all the technology and activities reviewed, one of the few areas in which the technology had positively impacted language acquisition was in the area of pronunciation, specifically due to speech recognition functionality and the feedback provided therein. The availability of data for contrastive analysis combined with the modifications suggested by Kolesnikova (2017) would also be beneficial in improving the quality of feedback provided in response to L2 pronunciation. The feedback provided could be based in contrastive analysis between a given L1 and L2, provided the ASR database contained information on both languages. This feedback could take various forms, such as aural, visual, or meta-linguistic. Visual feedback, for example, has already been proven to have a positive impact on pronunciation. Offerman and Olson (2016) found that VOT duration improved (i.e., the values trended toward native speaker VOT values) after participants were taught about waveforms and spectrogram representations of VOT and were able to compare their own recordings to those recorded by native speakers. If the ASR software had audio signal analysis capability, and the databases had spectrogram representations of preset words or phrases, the L1 and L2 spectrograms could be superimposed on each other for ease of comparison. This feedback could be nearly immediate, and could even be accompanied by playing back the user's own recording. Though VOT duration is more subtle than other types of errors, receiving feedback on it (and/or other more subtle adjustments toward native-like speech) would help learners keep focusing on and noticing their own pronunciation, especially at more advanced levels of L2 acquisition, which would avoid a plateau in the development of their L2 pronunciation (Martinsen et al., 2014).

LANGUAGE ANALYSIS OF LEARNERS' EXPERIENCES

This chapter presents and discusses the findings from the analysis of the Technology Questionnaire responses. This questionnaire was filled out by every participant who completed the final recording (i.e., post-test) in the Fall 2018 semester. It contained, in addition to background and experimental group-related questions, seven open-ended questions. The participants were instructed to write as much as they wanted in response to each one. The purpose of these questions was to probe the participants' thoughts about and attitudes towards the technology they interacted with throughout the Fall 2018 semester. The full text of the questionnaire (which was deployed using Qualtrics) can be found in Appendix G.

The responses to the seven open-ended questions were processed through Linguistic Inquiry and Word Count (LIWC), a psychometric text analysis software that parses text, comparing the words within each line to entries in its dictionary that have been tagged with psychologically-impactful categories. The output is then presented line by line, with each category denoting the percentage of words that fell into that category within the specific line of text (Masip et al., 2012; Pennebaker, Booth, et al., 2015; Tov et al., 2013). For the purposes of this study, the LIWC categories under consideration were limited to *positive emotion, negative emotion, tense, pronouns,* and *tentative language*. These five designations were selected beforehand as the most likely to illustrate the participants' attitudes towards their assigned technology and the experience of practicing their pronunciation throughout the semester (see Section 2.6 for a full description of LIWC and its categories).

This chapter first presents the output from processing all of the responses through LIWC. Then, an analysis based on technology-related words is presented. These analyses provide a quantitative interpretation of the results. The quantitative analysis is followed by a more qualitative one, in which examples of the participants' responses are presented and compared; these responses are grouped by topic, and aim to illustrate the participants' opinions regarding the technology they used in the present study. Finally, the discussion presents an interpretation of the findings in terms of answering Research Questions (2) and (3) (see Chapter 1).

5.1 Quantitative Analysis by Group using LIWC

The output from LIWC, which bases its calculations on the percentage of a given category found in each individual entry within the responses (see Figure 17) (Tov et al., 2013), was analyzed using the Kruskal-Wallis statistical test. This test was selected because it allows for the comparison of more than two groups and is non-parametric, which was necessary based on the small sample size. The Kruskal-Wallis test was performed on each individual LIWC category and/or subcategory separately; there were five categories investigated in the present study. The categories *positive emotion* and *negative emotion* (from which *anxiety*, *anger*, and *sadness* were extracted) were also subsumed under *affective processes*. Another category chosen was *tentative language*, which was selected due to its relationship to cognitive processes (Pennebaker, 2013; Pennebaker, Boyd, et al., 2015). The present study also included *tense* (which was divided into *focus past*, *focus present*, and *focus future*). Finally, the study also investigated the category *pronoun*, which was analyzed as a whole but also divided into the subcategories *personal pronoun* and *impersonal pronoun*. Additionally, data in the *pronoun* category were also separated by person (i.e., the subcategories *I*, *we*, *you*, and *they*[95]).

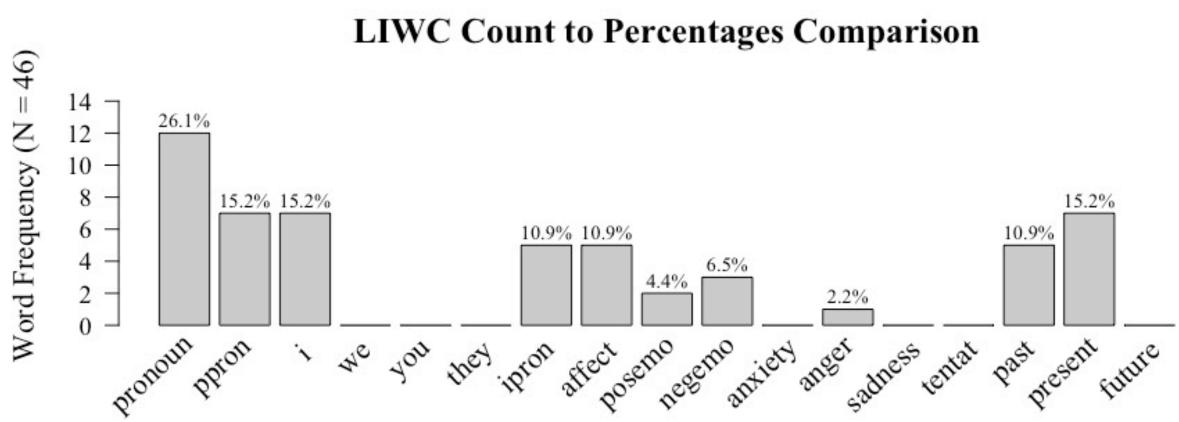
[95] Third person singular (i.e., she/he as designated by LIWC) is a subcategory of *pronoun* but was excluded due to there being no instance of it in the survey responses.

After the Kruskal-Wallis test was performed, the post-hoc Dunn pairwise test with a Bonferroni adjustment was used in order to determine between which group pairings statistical significance was found. To further explore differences in language use between the participant group pairings, the Dunn test was performed on the data from any category in which the Kruskal-Wallis test produced a p-value of $p = .125$ or smaller. While unusual, this helped determine whether the result would have been statistically significant or approached significance, had the comparison been made only between two of the groups.

As has been mentioned previously, LIWC data are presented as percentages out of the total number of words in the text sample. However, as the percentage depends on the number of words in the sample, which is highly variable, the true number of tokens being counted in a given category can be unclear. Figure 17 illustrates the difference between the percentages used in the LIWC analysis in comparison to the raw frequency of tokens.

Figure 17

Comparison of Raw Frequency to the Percentages used by LIWC for Analysis



Note. Percentages and frequency in Figure 17 are based on the following EG1 participant comment: It was frustrating that it told me i [*sic*] was saying the word wrong, but it could not provide a correction on how i [*sic*] can fix it. I liked to be told I was wrong, but i [*sic*] would have preferred to know how i [*sic*] can do better. (46 words)

In Figure 17, the raw frequency is presented in parentheses for each category that contained at least one token. For comparison, the y-axis represents the percentage out of the total words for that specific response ($N = 46$). Table 49 provides descriptive statistics for each LIWC category investigated in the present study for the three participant groups.

Table 49

Descriptive Statistics for Percentage of Responses in LIWC Categories by Participant Group

	EG1 ($N = 115$)		EG2 ($N = 98$)		CG ($N = 49$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
pronoun	17.11	8.01	14.99	7.92	16.91	8.49
ppron	9.54	5.69	8.60	5.69	8.95	5.36
I	7.64	5.80	7.22	5.68	7.63	5.70
we	0.23	0.89	0.22	1.14	0.39	1.40
you	1.10	2.72	0.74	2.44	0.72	2.61
they	0.56	1.60	0.41	1.25	0.21	0.91
ipron	7.57	5.59	6.40	4.88	7.92	5.75
affect	4.77	4.18	5.00	5.00	6.21	4.58
posemo	3.67	3.79	4.32	4.91	4.29	3.75
negemo	1.08	2.61	0.64	1.52	1.92	3.93
anxiety	0.09	0.64	0.12	0.60	0.07	0.49
anger	0.12	0.58	0.02	0.17	0	0
sad	0.06	0.67	0	0	0.16	0.63
tentat	2.95	4.19	3.26	4.26	3.07	4.19
focus past	6.98	5.37	6.78	5.57	7.62	6.50
focus present	6.46	5.35	7.19	5.66	8.99	7.06
focus future	0.31	1.09	0.68	1.62	0.30	0.83

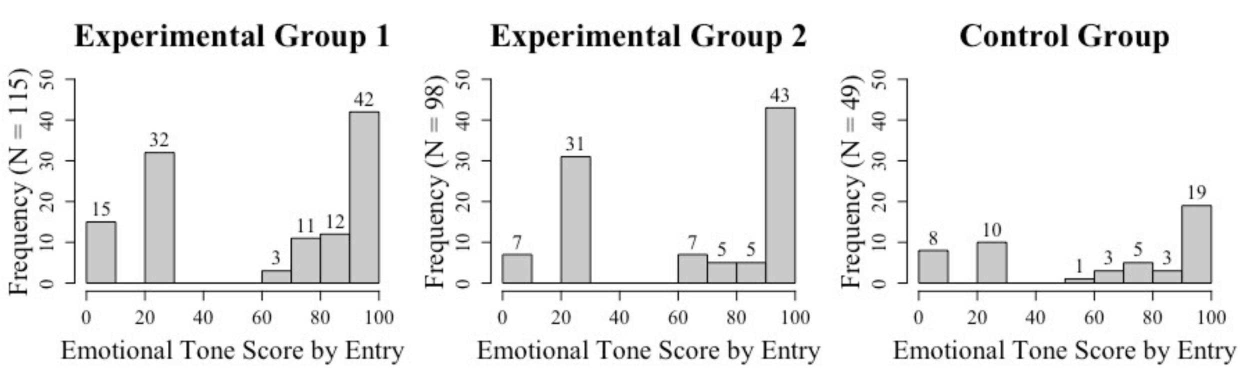
Due to the small number of responses overall (i.e., 115 individual entries for EG1, 98 for EG2, and 49 for CG), the present study analyzes both the percentages populated by LIWC (using the Kruskal-Wallis and Dunn pairwise statistical tests) and the total frequency of words from the responses to the Technology Questionnaire. The following sections examine the findings pertaining to the aforementioned LIWC categories .

5.1.1 Emotional Tone, Affective and Cognitive Processes

LIWC calculates a summary variable referred to as *emotional tone*, which, "provides information about both positive and negative emotions through an algorithm that accounts for them in a single summary variable. The higher the number reported, the more positive the emotions expressed" (Zapata & Ribota, 2021, p. 11). Included in the calculation are words tagged as *positive* and *negative emotion*, but unlike the other individual categories, the use of words in these categories is compared to the dictionary entries programmed into LIWC to find the difference (Rocklage et al., 2018). A score higher than 50 is generally considered to represent positive tone, while below 50 is considered negative (Jordan & Pennebaker, 2016; Nicholes, 2016). To ascertain the emotional tone of the responses, the score assigned to each individual entry from the three participant groups was compiled. Figure 18 utilizes histogram bins to demonstrate the number of times an entry received an emotional tone score within each range of ten. The actual number of responses in each bin is also provided as the number above each bar in the graph (e.g., in Figure 18, in the leftmost histogram, the bin representing zero to 10 contains 15 entries, and the bin representing 90 to 100 contains 42 entries).

Figure 18

Distribution of emotional tone *Scores per Questionnaire Entry, by Participant Group*



As can be seen in Figure 18, based on emotional tone, the overall impression of the technology utilized in the present study is positive, regardless of participant group, as more of the entries received emotional tone scores above 50 in each case. Mathematically, CG appears to have the highest emotional tone, with 63.3% of responses having scores of 50 or higher (compared to 59.1% for EG1 and 61.2% for EG2). However, it is important to note that CG had approximately half the participants and responses than the other groups. Cross-multiplication for normalization (everything compared using $N = 115$) allowed for more direct comparisons of the responses, the results of which can be seen in Table 50.

Table 50

Normalization of emotional tone *Distribution (N = 115) per Participant Group*

	EG1	EG2	CG
Above 50	68	70	73
Below 50	47	45	42
90+	42	50	45
80-89	12	6	7
70-79	11	6	12
60-69	3	8	7
50-59	0	0	2
20-29	32	36	23
0-9	15	8	19

The normalization, shown in Table 50, indicates that when looking at the data from the perspective of above or below the emotional tone score of 50, it appears that CG has higher emotional tone overall than the other two groups. When expanded out to look at every range (i.e., histogram bin) in which there was at least one entry assigned that score, however, it is evident that the despite the higher emotional tone scores overall, CG's scores are more widespread than the other two groups. Nevertheless, Figure 18 and Table 50 indicate that emotional tone was high for all three groups, which has previously been associated with positive reactions to activities (Oliver et al., 2021).

Within LIWC's summary variable *emotional tone* is *affective processes*, or *affect*. According to Walla (2018), "affective information is evaluative leading to a decision on how something is" (p. 144). As the category *affect* subsumes the other emotionality categories (both positive and negative), the percentages calculated by LIWC for this category give more insight into emotional language use than what type of emotion it is; thus, the presentation of *affect* data is followed (and supplemented) by presentations of *positive emotion,* and then *negative emotion*, which includes *anxiety, anger,* and *sadness*. Table 51 presents the most frequent *affect* words used by the participants.

Table 51

Highest Frequency Words Coded as affect *by LIWC, by Participant Group*

EG1 (N = 187)		EG2 (N = 155)		CG (N = 80)	
Word	Count	Word	Count	Word	Count
I liked	19	better	14	I liked	4
definitely	11	best	10	sure	4
better	9	I liked	9	better	4
helpful	9	helpful	9	easy	3
easy	9	improve(d)	8 (+6)	helpful	3
wrong	7	good	6	wrong	3

As can be seen in Table 51, the *affect* category encompasses both *positive* and *negative emotion* words; it is interesting to note that the word 'wrong,' tagged as *negative emotion* was among the top six most frequently used words in this category for two of the three groups. The Kruskal-Wallis statistical test revealed that there was no statistically significant difference between the groups ($H(2) = 4.18$, $p = .124$). A Dunn pairwise test was performed, which revealed that the differences in the use of words in the *affect* category approached significance when comparing EG1 to CG ($p = .078$).

Delving into the categories that are subsumed by *affect*, Table 52 presents the frequency of word use in the category *positive emotion* (abbreviated *posemo*). As was mentioned previously, higher (i.e., positive) *emotional tone* scores, which are generated due to higher use of *positive emotion* words, have been found to correspond positively to engagement as well as positive reactions to activities (Oliver et al., 2021). Additionally, "positive emotion may denote mood rather than a stable psychological trait such as happiness" (Jaidka et al., 2020, p. 294), which is reflected in the most frequent words shown in Table 52.

Table 52

Highest Frequency Words Coded as posemo *by LIWC, by Participant Group*

EG1 (N = 139)		EG2 (N = 131)		CG (N = 56)	
Word	Count	Word	Count	Word	Count
I liked	19	better	14	I liked	4
definitely	11	best	10	sure	4
better	9	I liked	9	better	4
helpful	9	helpful	9	easy	3
easy	9	improve(d)	8 (+6)	helpful	3
best	7	good	6	easily	2

From Table 52, it is clear that many of the same words were used with higher frequency across the groups. The use of 'definitely' stands out from the rest, as it was used with high frequency in EG1, but was not in the six most frequently used words for the other two groups. This difference, however, had little statistical impact, as the Kruskal-Wallis test revealed no significant difference in the use of *positive emotion* words between the groups ($H(2) = 1.50$, p = .471).

Negative emotion, in contrast, includes words that connote negative feelings or negative tone. While this category includes words such as 'wrong,' 'disliked,' 'frustrating,' and 'annoying,' and are likely more straightforward to connect with *negative emotion*, this category also includes words such as 'difficult,' 'confused,' and 'problem' which relate to situations in which negative emotions might be felt. Despite including the word 'frustrating' in the category, previous research has found that higher use of general negative emotion words is correlated with feelings of frustration (Valenti et al., 2019). Within the category *negative emotion*, certain words are tagged with additional specifications for *anxiety, anger,* or *sadness*. Table 53 presents the highest frequency words for the *negative emotion* category in general, as well as for *anxiety, anger,* and *sadness*.

Table 53

Highest Frequency Words Coded as negemo *by LIWC (with* anxiety*,* anger*,* sadness*), by Participant Group*

Category	EG1 Word	EG1 Count	EG2 Word	EG2 Count	CG Word	CG Count
	(N = 28)		(N = 19)		(N = 24)	
negemo	wrong	7	difficult(y)	5 (+1)	wrong	3
negemo	frustrating	4	dislike(d)	1 (+3)	problem	2
negemo	difficult(ies)	2 (+2)	wrong	2	tedious	2
negemo	struggling	2	confused	1	difficult(y)	2 (+1)
negemo	disliked	2	struggled	1	dislike(d)	1 (+1)
	(N = 3)		(N = 4)		(N = 1)	
anxiety	struggling	2	confused	1	insecurity	1
anxiety	confused	1	struggled	1		
anxiety			nervous	1		
	(N = 5)		(N = 1)		(N = 0)	
anger	frustrating	4	annoying	1		
anger	annoying	1				
	(N = 1)		(N = 0)		(N = 3)	
sadness	low	1			hurt	1
sadness					missed	1
sadness					lost	1

As can be seen in Table 53, while there is some overlap in the most frequent words (tagged as *negative emotion*) used by the two experimental groups, there are some notable differences, such as higher frequency of ‘wrong’ and ‘frustrating’ by EG1. The Kruskal-Wallis test revealed that the differences in the usage of *negative emotion* words approached significance ($H(2) = 5.61$, $p = .061$). The subsequent Dunn pairwise test revealed a statistically significant difference between EG2 and CG ($p = .029$), and a difference that approached significance between EG1 and CG ($p = .088$). Conversely, the two experimental groups used words tagged as *negative emotion* similarly

(p = .797), though the higher frequency of words such as ‘wrong’ and ‘frustrating’ likely contributed to the lack of statistical significance between EG1 and CG.

The *negative emotion* category has three subcategories: *anxiety*, *anger*, and *sadness*. That is not to say that all words found within the negative emotion category are sorted into one of these three subcategories, but rather that some words have been tagged as not only belonging to negative emotion, but also representing *anxiety*, *anger*, or *sadness*. Of the three categories illustrated in Table 53, no statistically significant difference was found for *anxiety* ($H(2) = 0.59$, $p = .746$) or *anger* ($H(2) = 4.05$, $p = .132$). However, the use of words coded for *sadness* was found to be highly statistically significantly different between the three participant groups ($H(2) = 8.62$, $p = .013$). The post-hoc Dunn pairwise test revealed that there were statistically significant differences between two of the three possible group pairings: EG1 and CG ($p = .019$) as well as EG2 and CG ($p = .007$). While the results of the Dunn pairwise test indicate that both of the experimental groups performed significantly differently than CG, it is important to note that the *sadness* subcategory contained very few tokens.

In addition to *affective processes,* LIWC also designates *cognitive processes* as another summary variable. In contrast to *affect*, however, “[c]ognitive information focusses on semantic features that lead to an understanding [of] what something is” (Walla, 2018, p. 144). Tentative language is one of the categories classified by LIWC under cognitive processes (Pennebaker, Boyd, et al., 2015). Tentative language has been associated with uncertainty or insecurity as well as with certain aspects of storytelling; studies have found that previously undisclosed stories often have higher numbers of tentative language words (Tausczik & Pennebaker, 2010, p. 36). In addition, tentative language has been found to be negatively indicated with authoritativeness (Sell & Farreras, 2017), meaning that use of more tentative words or phrases lowers the impression of authority, as well as to soften potentially controversial or absolute statements (Medimorec & Pennycook, 2015, pp. 600–601). Table 54 illustrates the most frequent words labeled as *tentative language* by LIWC in each group.

Table 54

Highest Frequency Words Coded as tentat *by LIWC, by Participant Group*

EG1 (*N* = 98)		EG2 (*N* = 104)		CG (*N* = 46)	
Word	Count	Word	Count	Word	Count
or	13	or	17	or	4
if	12	if	14	if	4
some	12	lot	11	question(s)	4 (+3)
sometimes	12	some	10	anything	3
most	8	any	6	probably	3
someone	3	sometimes	5	lot	3

As can be seen in Table 54, the most frequent tentative words used by the two experimental groups are very similar, in contrast to those used by CG. Despite the difference in lexical items, however, the Kruskal-Wallis test determined that there was no statistically significant difference between the three groups ($H(2) = 0.18$, $p = .915$).

5.1.2 Time Orientation (i.e., Temporal Focus, Tense)

LIWC also designates three classifications words based on *time orientation*: *focus past, focus present,* and *focus future* (Pennebaker, Boyd, et al., 2015). Use of different tenses, particularly present and future in contrast to past tense, has been linked with more positive impressions (Rúas-Araújo et al., 2016; Tausczik & Pennebaker, 2010). In terms of academic success, more frequent usage of present tense has been found to predict poorer performance (Robinson et al., 2013). For the purposes of the present analysis, it is important to note that, as LIWC is a psychometric tool, *tense* is more aligned with *temporal focus* than strict syntactic verbal tense. For example, one of the words that was tagged as *focus past* from the responses in the present study is ‘remember’. Though syntactically the verb is in present tense, the meaning of the verb indicates something that has to have already happened, thus leading to its classification as *focus past*. While the majority of the words that were classified as *focus past* in the present study are verbs, this category also includes words such as 'already,' 'previously,' and 'earlier' (all of which can be found in the responses to the Technology Questionnaire in the present study) which reference something that already occurred; these words are included because LIWC classifies words based on psychological properties associated with them (McHaney et al., 2018) rather than solely considering syntax and/or morphology. Table 55 includes the highest frequency words classified as *focus past,* Table 56 presents *focus present*, and Table 57 *focus future*.

Table 55

Highest Frequency Words Coded as focus past *by LIWC, by Participant Group*

EG1 ($N = 276$)		EG2 ($N = 229$)		CG ($N = 114$)	
Word	Count	Word	Count	Word	Count
was	84	was	70	was	34
helped	21	helped	17	helped	10
liked	21	had	15	had	10
were	17	were	13	were	8
did	17	liked	13	did	7
had	16	did	9	used	6

As can be seen in Table 55, the most frequent words are nearly identical across all the three participant groups (the frequency of 'liked' in CG's responses was five (5)). Additionally, the Kruskal-Wallis test indicated that there was no statistically significant difference between the use of words coded as *focus past* by the three groups ($H(2) = 0.47$, $p = .791$). As for 'did,' which occurs frequently across groups, it appears to be used for emphasis (e.g., 'I did like' instead of 'I liked') more than as the past tense of the verb 'do'.

As was the case for *focus past*, *focus present* relates to temporal focus. Thus, words like *today's* are included in the category in addition to verbs. Also, as was the case for *focus past*, the use of ‘do’ or ‘does’ (seen in Table 56), is often more emphatic than actual use of the verb ‘do’.

Table 56

Highest Frequency Words Coded as focus present *by LIWC, by Participant Group*

EG1 (N = 240)		EG2 (N = 246)		CG (N = 137)	
Word	Count	Word	Count	Word	Count
have	28	think	22	practice	13
is	18	is	22	have	13
use	18	have	18	is	7
learn	15	be	17	be	7
practice	15	do	15	do(es)	6 (+1)
say	12	learn	10	get	6

Table 56 illustrates that there is some variation in the most frequent words coded as *focus present* by LIWC between the three participant groups. Although the Kruskal-Wallis statistical test revealed no statistically significant difference between the groups’ usage of *focus present* tagged words ($H(2) = 4.57$, $p = .102$), the Dunn pairwise test revealed a statistically significant difference in usage between EG1 and CG ($p = .049$), with CG using more *focus present* words ($M = 8.99$) than EG1 ($M = 6.46$).

Focus future is different from the other two time orientation categories in that, due to the syntax of English (i.e., the lack of verbal conjugation in future tense), the number of verbs included is lower than what was seen in *focus past* and *focus present*. Instead, this category contains words that can be interpreted as having psychological properties that would indicate *future*. Table 57 includes all of the words categorized as *focus future* by LIWC from the responses to the Technology Questionnaire.

Table 57

Highest Frequency Words Coded as focus future *by LIWC, by Participant Group*

EG1 ($N = 12$)		EG2 ($N = 24$)		CG ($N = 8$)	
Word	Count	Word	Count	Word	Count
going	4	then	11	might	2
then	2	might	3	then	1
wish	2	wanna	2	prepare	1
will	1	(up)coming	(+1) 1	future	1
might	1	gonna, going	1, 1	going	1
prepared	1	potential	1	sometime	1
future	1	expected	1	hope	1
		may	1		
		will	1		

The Kruskal-Wallis test revealed no statistically significant differences between the three groups in terms of their use of words coded as *focus future* ($H(2) = 3.39$, $p = .184$). Despite the lack of statistical significance, of interest is the higher frequency of words tagged as *focus future* in EG2 in comparison to the other two groups (see Table 57). As *present* and *future focus* can indicate a more positive viewpoint, this might show a difference between the groups that did not appear statistically.

5.1.3 Pronouns

The category *pronoun* encompasses the categories *ppron, ipron, I, we, you,* and *they*. Table 58 includes the seven most frequent words from each of the three participant groups. These values can also be found in subsequent tables which are divided into specific pronoun categories.

Table 58

Highest Frequency Words Coded as pronoun *by LIWC, by Participant Group*

EG1 (N = 605)		EG2 (N = 513)		CG (N = 264)	
Word	Count	Word	Count	Word	Count
I	161	I	168	I	79
it	137	it	101	it	54
me	66	that	44	that	24
that	65	my	41	my	20
my	38	me	32	you	15
you	37	this	25	me	12
this	18	you	21	what	9

When analyzed as a whole, the Kruskal-Wallis statistical test indicated that there was no significant difference between the overall use of pronouns between the three participant groups ($H(2) = 4.17$, $p = .124$). The Dunn pairwise test, however, revealed that the comparison of the two experimental groups approached significance ($p = .098$). Pronouns have been shown to relate to status and attentional focus (Tausczik & Pennebaker, 2010); as the participants were all students, they would have the same status, so it is possible the difference between the two experimental groups (shown in Table 58) stems from differences in attentional focus.

Within the category *pronoun*, LIWC divides the pronouns by person but also creates a category that encompasses all personal pronouns (*ppron*); it also tags impersonal pronouns (*ipron*). Table 59 includes the most frequent *personal* and *impersonal pronouns* found in the responses to the Technology Questionnaire. As was the case with *pronoun*, the values for *ppron* are repeated in their respective *person* categories (e.g., *I, we, you, they*).

Table 59

Highest Frequency Words Coded as ppron *or* ipron *by LIWC, by Participant Group*

	EG1		EG2		CG	
	Word	Count	Word	Count	Word	Count
	(N = 336)		(N = 296)		(N = 144)	
ppron	I	161	I	168	I	79
ppron	me	66	my	41	my	20
ppron	my	38	me	32	you	15
ppron	you	37	you	21	me	12
ppron	them	11	them	9	I'm	5
ppron	they	9	they	5	we	4
	(N = 217)		(N = 217)		(N = 119)	
ipron	it	137	it	101	it	54
ipron	that	65	that	44	that	24
ipron	this	18	this	25	this	9
ipron	what	8	what	14	what	9
ipron	which	7	other(s)	4 (+2)	anything	3
ipron	other(s)	6 (+2)	someone	4	things	3

As shown in Table 59, the highest frequency words are very similar across all three participant groups; additionally, the frequency of the words is similar between the two experimental groups. It is not surprising, then, that the Kruskal-Wallis test determined that there was no statistically significant difference in the use of *personal pronouns* ($H(2) = 1.39$, $p = .498$), or *impersonal pronouns* ($H(2) = 3.04$, $p = .219$) between the three groups.

Within the category of *pronoun*, the pronouns themselves are grouped individually into *I, we, you,* and *they*. As can be seen in Table 60, the use of *I* overwhelmed the other pronoun groupings, which is unsurprising given that the participants were asked to evaluate their own experiences.

Table 60

Highest Frequency Words Coded as I, we*,* you*, or* they *by LIWC, by Participant Group*

	EG1		EG2		CG	
Category	Word	Count	Word	Count	Word	Count
	(N = 268)		(N = 248)		(N = 121)	
I	I	161	I	168	I	79
I	me	66	my	41	my	20
I	my	38	me	32	me	12
I	I'm	1	myself	2	I'm	5
I	myself	1	I'm	3	I'd	3
I	I've	1	I'd	1	myself	2
	(N = 8)		(N = 7)		(N = 5)	
we	we	7	we	4	we	4
we	our	1	our	3	our	1
	(N = 39)		(N = 24)		(N = 15)	
you	you	37	you	21	you	15
you	you're	1	yourself	2		
you	your	1	your	1		
	(N = 21)		(N = 17)		(N = 3)	
they	them	11	them	9	they	2
they	they	9	they	5	their	1
they	their	1	their	3		

All three participant groups pattern very similarly in terms of which person pronouns were more frequent than others (enumerated in Table 60), which provides insight into the lack of statistically significant differences between them. The Kruskal-Wallis statistical test found no significant differences between the groups for *I* ($H(2) = 0.34$, $p = .843$), *we* ($H(2) = 1.31$, $p = .520$), *you* ($H(2) = 3.80$, $p = .149$), or *they* ($H(2) = 3.09$, $p = .213$). As has been mentioned previously, third person singular is also a subcategory that would be grouped with the ones listed in Table 60; however, there was no use of third person singular pronouns in the questionnaire responses. The only third person references in the responses were regarding 'John,' the student whose study abroad story was illustrated in the app, and references such as 'teacher,' which were not pronominalized and thus not included in this analysis.

5.2 Analysis of Technology-Related Words

Referring back to the analysis and discussion in Chapter 4, one potential explanation for the differences in the performance of the two experimental groups was that the type of technology used affected each group differently. As has been explored in the present chapter, few statistically significant differences have been found between the three groups when analyzing their language use based on LIWC categories. However, trends begin to appear when the analysis focuses on particular terms, such as technology. When reviewing the participant's responses to the Technology Questionnaire, words such as 'glitch,' 'crash,' and 'phone' stood out, particularly in the responses from EG1. Therefore, technology-related words and their frequencies were separated out from the whole dataset (see Chapter 3 for an explanation of method) in order to determine whether the use of these words revealed anything that could demonstrate differences between the three participant groups. Chi-Square Goodness of Fit tests were run on the total number of instances of the words designated as *technology-related*.

Supporting the claim that the participants behaved differently from each other specifically in their use of technology-related words, the raw frequency totals for these words differed based on technology. Table 61 includes the words and count of the most frequent technology-related words found in the questionnaire responses. Unsurprisingly, the words 'technology' and 'feature' occurred frequently, as they were included in the prompt questions. Table 61 also includes information regarding plurals or other derivations; when included, their information can be found in parentheses.

Table 61

Highest Frequency Technology-Related Words, by Participant Group

EG1 (*N* = 145)		EG2 (*N* = 94)		CG (*N* = 67)	
Word	Count	Word	Count	Word	Count
app(s)	37 (+1)	technology	25	technology	11
technology	26	VHL	10	Blackboard	11
feature(s)	10 (+2)	recording(ing)(s)	7 (+7) (+1)	VHL	6
phone(s)	7 (+1)	feature(s)	6 (+2)	feature	3
glitch(es)	4 (+5)	session	6	studio	3
crash	4	accessible	4	recordings	2
voice	4	voice	2	headset	2
access(ible)	3 (+2)	recorded	2	access(ible)	2 (+1)
recognition	2	computer	2	button	2

As can be seen in Table 61, the highest frequency words in both experimental groups are either 'technology' or a reference to the technology the participants used in the study. Closely following those is 'feature(s)' which, as was mentioned previously, was included in many of the question prompts. After those, however, the higher frequency words demonstrate a difference in focus between the two groups. For EG1 (who used the app), the next most frequent words indicate what was likely most striking to them: 'phone,' as the requirement for interacting with the app was having an iPhone, and thus this was the only technology available to them to interact with the app; 'glitch(es),' followed by 'crash,' which indicates that one of the more memorable elements of the app was that it had technological issues. Within the nine most frequent words for EG1, however, were 'voice' and 'recognition'; their frequencies might be on the smaller side, but these ideas appeared in other ways as well, often as descriptions of how the speech recognition function worked.

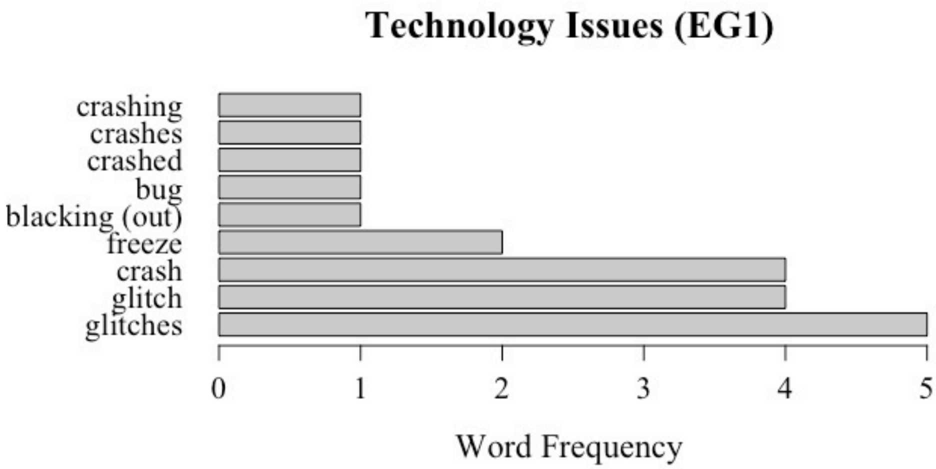
For EG2, however, the focus is somewhat different. Higher frequencies of 'recording(s)' and 'session' indicate that these participants' focus was directed more toward the times when their voices were recorded; as every interaction with their pronunciation activities resulted in their voices being recorded (and then compared with a native speaker), the high frequency of 'recording(s)' is less surprising. 'Recorded' also was among the nine most frequently used words, as was 'voice,' which again refers mostly to their assigned activities. What is missing from this group of high-frequency technology-related words for EG2 are words that negatively impact technology or indicate problems (unlike EG1). This difference in high-frequency words adds more support to the argument that technology type affected the participants' pronunciation improvement (see Chapter 4) due to the distraction caused by the technological issues that the participants using the app had to navigate (but the participants using the textbook provided activities did not).

In contrast, the higher frequency words used by CG indicate a different focus. First, many made connections between their homework activities (completed on a website referred to by students and instructors as VHL), possibly comparing the technology, but based on participant comments, differentiating between their homework site and its restrictions and the grammar-related activities on the Blackboard site used in this study. Additionally, the participants in CG used words like 'headset,' 'studio,' and 'recordings' with higher frequency than other technology-related words, indicating that the four recording sessions throughout the semester were memorable to them and worth mentioning.

In addition to the differences in frequency overall, there was a clear difference in the use of words relating to problems with the technology. Figure 19 highlights the frequency of words related to issues or problems with technology found within the technology-related words. It is important to note that Figure 19 only contains data from EG1 because only one of these words ('glitch') occurred outside of that group and only once (in EG2).

Figure 19

Frequency of Words Related to Technological Issues from Experimental Group 1 Responses



As can be seen in Figure 19, words related to *glitch* and *crash* make up the majority of the words pertaining to technological issues. The disparity between EG1 and the other two groups in terms of words relating to technological issues or problems reveals that the participants from EG1 appear to have spent a lot of mental energy thinking about or focusing on the functionality of the app that should have been allotted to learning Spanish and practicing pronunciation.

As was mentioned previously, a Chi-Square Goodness of Fit test was performed on the total number of technology-related words used by each of the participant groups in order to determine whether there was a difference between the groups in terms of how often they mentioned technology in the questionnaire. The results of the Chi-Square test indicated that there was a significant difference in the frequency of use of technology-related words between the participant groups. There was a highly statistically significant difference between all three groups ($\chi^2(2, N = 306) = 30.77, p < 1.00^{-5}$). However, the comparison between the three groups did not reveal precisely where the differences in usage could be found. Thus, three subsequent chi-square tests were performed on each distinct pair of groups, and statistically significant differences were found between all three groups. The use of technology-related words was highly statistically significantly different between the two experimental groups ($\chi^2(1, N = 239) = 10.88, p = 9.70^{-4}$) as well as between EG1 (the app-using group) and CG ($\chi^2(1, N = 212) = 28.70, p < 1.00^{-5}$). Additionally, the difference in frequency was statistically significant when comparing EG2 (who used the listen-and-repeat style activities that accompanied the course textbook) and CG ($\chi^2(1, N = 161) = 4.53, p = .033$). The clear differences between EG1 and the other two groups could be interpreted as a larger focus on the technology itself, which, when compared to the findings discussed in Chapter 4, supports the claim that technology type impacted pronunciation performance. In this case, it is likely that the high frequency of technology-related words indicates that focus was taken away from the pronunciation practice/training because the technology became a distraction.

5.3 Insight from Trends in Participant Commentary

This section aims to explore (more qualitatively) the experiences of the participants and their opinions regarding either the Spanish SOLO app, the pronunciation activities found in the online companion to the participants' first-semester Spanish course textbook, or the grammar activities completed via Blackboard. While the previous findings were presented by emotional category, this sections presents the participants' responses to the Technology Questionnaire by topic. While there were many topics covered in the questionnaire responses, the present study presents those that most closely relate to the study's research questions. The five topics discussed in this section are: *positive impressions of technology*; *impressions of feedback*; *technological issues*[96] *and/or suggestions for improvement*; *impact on performance*; and *rationales behind recommendations*. The specific participant comments included in this section represent frequently expressed ideas found throughout the questionnaire responses (referenced

[96] To avoid redundancy, the section on technological issues subsumes negative impressions of technology.

by comment number in parentheses) though in most cases, they are not an exhaustive list of all commentary pertaining to a given topic. Appendix A contains additional participant commentary, shown in Tables A6 through A11.

5.3.1 Positive Impressions of Technology

Analyses of the first topic, *positive impressions of technology*, shows that on the whole, the participants reacted positively to the technology, regardless their group membership. The general impressions of the Spanish SOLO app utilized by EG1 were mostly positive and varied, with participants focusing on different aspects of the app. The positivity of the emotional content of the responses is supported by the presence of words tagged as *positive emotion* by LIWC, such as 'enjoyable' (1), 'nice' (2), 'I like(d)' (2, 3), and 'enjoyed' (3):

(1) I think the activities were more like games than actually worksheets or homework. So it made it more enjoyable.

(2) I liked that I [*sic*] could hear the words being spoken to me so i would know how to pronounce them properly. I also like how easy it was to use the technology. The length of the lessons was also very nice because you could do several lessons in the amount of time you were supposed to interact with the technology.

(3) I liked that the technology was accessible and easy to come back to if needed. I liked that I needed to record my voice for certain words to see if I was pronouncing them correctly. Overall I definitely enjoyed using the app more than others I have seen in the past.

(4) I feel like I got more out of my experience using the app than I would have received if I had been working on VHL homework or blackboard assignments.

Some of the commentary focused on elements pertaining to pronunciation[97], such as in (2) and (3), but the variety of comments pertaining to other features indicates that the participants' responses to the app are likely based off of more than one specific feature (i.e., speech recognition). Availability (or accessibility) also appears to have been favored by the participants

[97] Much of the commentary regarding speech recognition also focused on feedback (see Section 5.3.2)

in EG1 (2, 3). Additionally, 'grammar' and 'vocabulary' were often mentioned in these responses, two themes that persist throughout the responses. Comments (1), (3), and (4) also differ from the others slightly due to the comparisons they make. Comment (1) equates the app to a game, which has been shown to correlate with engagement (Reinhardt & Sykes, 2014, p. 4); engagement could also apply to comment (4) which clearly states a preference for the app over the hypothetical outcomes of interacting with the other technologies[98]. These comments demonstrate that the participants in EG1 appreciated the app for more than just the speech recognition functionality it offered.

The comments from EG2 mirror those from EG1 in terms of general impressions of the online pronunciation activities. The language use is also similar, the responses featuring *positive emotion* words such as 'I liked' (5), 'easy/easily' (5), 'fun' (6), 'great' (7), and 'best' (8):

> (5) I liked that the technology was easily accessible and in the same place as my spanish [*sic*] homework. It made it easy for me to be able to switch from one activity to the other when I was able to. It also was one less log in information that I had to remember.
>
> (6) I had a lot of fun working on these activities, especially the grammar. I've always wanted to learn Spanish and this has helped me in a lot of ways to become a better speaker of the language.
>
> (7) I think being able to hear the word or phrases spoken by a native speaker was great. Being able to listen to how it should actually sound right before I recorded really helped me pronounce it well.
>
> (8) The best feature of the technology was the sections where I drilled vocabulary because it helped me greatly in working on my pronunciation of different kinds of words.

As is the case for EG1, 'grammar' (6) and 'vocabulary' (8) are mentioned consistently by the participants in EG2. EG2, however, appears to have focused on the functionality of the

[98] Participants in EG1 had experience with the other two technologies as the activities for EG2 came from the online companion to the course textbook, and Blackboard is the platform adopted by the university.

technology in terms of pronunciation generally more than EG1, describing features such as *listening to a native speaker* (7), but did not typically describe detailed experiences with the technology in general commentary. Accessibility (or availability) appears to have ranked high in importance for EG2 as well (e.g., comment (5)). Overall, it seems that EG2 evaluated their general experiences similarly to EG1.

Despite using a dramatically different technology, the comments from CG seem to be positive as well regarding general impressions of their technology. As was mentioned in the previous section, there was no statistically significant difference between the three groups' use of *positive emotion*. Similar words are evident in these responses as well, such as 'best' (9), 'favorite' (10), 'easy' (11), 'fun' (11), and 'enjoyed' (12):

(9) The best feature of this technology was the convenience of being able to work with it on my own time and being able to work around a schedule that could benefit me and the study in itself. The layout of the work made for quick studying quizzes that helped me to practice things that I needed to learn a little more.

(10) With Blackboard there were only grammar activities so I guess those were my favorite. They were really helpful.

(11) The technology was easy to understand and was very straightforward when using. The recordings were difficult but were fun to read as it tested my ability of the pronunciation of words that I don't usually have the opportunity to use.

(12) I thoroughly enjoyed the set up. The recording studio has a very legitimate feel to it. Also, I probably wouldn't have visited the Media center as early as I did if it were not for this studio.

The themes in the comments are similar as well, referencing ideas such as *accessibility* and/or *availability* (9) and specifically mentioning 'grammar' (10). Unsurprisingly, the comments were more limited due to the less sophisticated type of technology used by this group. However, some of the participants from CG interpreted 'technology' to be more than the grammar activities on Blackboard; comments (11) and (12) refer to the periodic recordings completed in a recording studio in the media center over the course of the semester. The unusual interpretation of the question(s) provides insight into how CG viewed the recording sessions, as they were the only exposure to anything pronunciation-related (other than class) that they experienced.

As was mentioned previously, the comments included in this section reference technology very generally. However, the majority of the comments from all the groups were more expressive regarding the technology they used. Many of the comments also focused on a particular feature, *feedback*, which is explored in the next section.

5.3.2 Impressions of Feedback from Technology

Looking at the second topic, *impressions of feedback from technology*, within the comments pertaining to positive and negative learning experiences stemming from the present study, an idea that was consistently expressed was *feedback*. The specific type of feedback being described focuses on the information provided by the technology to the student regarding their performance. The consistency of these comments regarding *feedback* over all three groups (though in varying frequencies) was surprising in that most of the comments centered around the same ideas (within each participant group). Nevertheless, the same theme, *feedback*, was repeated throughout most of the responses, particularly in the two experimental groups.

As the feedback provided by the app was mainly provided by the speech recognition functionality (other than the standard grading of answers), it is unsurprising that speech recognition is the main focus of the responses from EG1, even if sometimes it is not specifically named:

> (13) I liked that it would not allow you to "pass" saying a word until it was said perfectly. It forced me to continuously practice words that I was struggling with. Otherwise, I would have assumed i [*sic*] was correct and moved on.

(14) The speech recognition really makes sure that you're saying your words right or it will not let you pass or you get the answer wrong. This made me learn the words more carefully and it helped me out the most.

(15) The best feature of the technology was its ability to listen to my voice to see whether the pronunciation was correct or not. The only other time I have experienced vocal interaction was in a Spanish classroom, which is only 3 hours per week. This app allowed me to correct myself by listening to the recordings and comparing them to the spanish [*sic*] speaker programmed in the app to see how I could improve my pronunciation.

Comments (13) and (14) use the term 'pass' in reference to a mechanism (the speech recognition) that prompted them to correct their response. The remainder of comment (13) is a striking insight into the mindset of a first semester language learner: *I would have assumed I was correct*. The implied rationale likely is: *because I was not told otherwise (by someone or something)*. Comment (14) gives the speech recognition function similar agency and responsibility: *[it] really makes sure that you're saying your words right*. Comment (15), however, made the action of correction reflexive (*correct myself*), but also still gave agency to the app: *this app allowed me*; *its ability to listen to my voice*, while reiterating the desire for knowledge of the 'correct' answer provided by the technology.

Although EG2 had fewer responses that could be classified as either directly mentioning or implying the idea of *feedback*, their responses are similar to those of EG1. Despite the fact that feedback in their activities was not automatic and required them to take the initiative to compare their recording to the native speaker's, the perception from participants in EG2 appears to view it similarly to automatic feedback:

(16) I appreciated the immediate feedback I received when completing some of the parts of the activities. I could record my voice then hear the phrase said correctly, then hear my voice again. This helped me to see the exact times that I pronounced words or phrases incorrectly.

(17) I liked having the ability to practice speaking Spanish after hearing the computer say it. I could practice saying it as many times as I liked before

submitting it. I also liked listening to my own personal recording before submitting and comparing it to the computer.

(18) Many of the simple pronunciation activities were one word or short phrases and I could try to emulate the exact pronunciation of the native speaker. It was also nice to be able to re-record anything that I knew I pronounced very wrong.

(19) I personally believe the best feature of the technology was being able to listen back to my personal recording before submitting. Many times I caught myself so focused on reading my response that I did not even notice I was pronouncing some phrases and words incorrectly.

In contrast to the responses from EG1, however, those from EG2 appear to portray more agency on the part of the participant. While some responses did give agency to the website, more of the EG2 responses show the participants taking the initiative, such as (17) *I liked listening [. . .] and comparing*, and (18) *I could try to emulate*. Additionally, (16) described the experience of completing the activity (*I could record my voice then hear the phrase [. . .] then hear myself*). Comment (19) introduces a different issue: *I was so focused [. . .] I did not even notice*. These responses from EG2 could indicate more involvement in the process of figuring out the differences in pronunciation, which may have had an effect on their performance.

In terms of impressions of feedback, CG also had a few (*n* = 3) responses in this category, but unlike the other two groups, the word 'feedback' was never explicitly mentioned within the responses. However, the content and descriptions conveyed the idea of *feedback*. Due to the difference in activity type, the type of feedback described also differs from the other groups.

(20) The best feature would probably be the instant grading. You can see exactly what you missed right after you finish an assignment. It shows you the right answer so you know exactly what you need to do better next time.

(21) I was already familiar with BlackBoard [*sic*], so it was not hard to access. I also like how it told you the correct answers whenever you answered incorrectly. I liked having to speak aloud, because it made me practice my pronunciation.

(22) I also liked how I had unlimited tries so I could really learn form [*sic*] my mistakes and understand the material.

The ideas *instant grading* (20), *it told you the correct answers* (21), and *learn from mistakes* (22) are phrases that could be used to describe what feedback can accomplish, indicating that the participants attempted to glean as much information as possible from the technology they were given access to, specifically confirmation that they were understanding the material. In terms of agency, two CG responses, (20) and (21), contained statements that put the technology in the agentive role (*it shows you*; *it told you*); Comment (22), however, was more agentive (*I had*; *I could learn*). Comment (21) also mentioned something unusual for this group, as it mentions speaking and pronunciation; since their activities were designed to not feature pronunciation, it is possible this comment pertains to the recording sessions throughout the semester. The themes within CG's comments are similar to those from the other two groups, possibly indicating common desires from language learners to know whether they are successfully grasping the concepts they are learning. However, they may not have been as active of participants as those in EG1 and EG2.

While the commentary on the technology was typically positive across the three participant groups, two groups of participants also expressed negative experiences receiving feedback from their assigned technology. As CG utilized the least-interactive technology (i.e., grammar activities on a Blackboard course site), it was somewhat unsurprising that there were no comments from CG that could be categorized as expressing a negative impression pertaining specifically to feedback provided by the technology itself. However, the difference in comments between the two experimental groups begins to shed some light on the differences in their pronunciation performance.

Most of the comments that portrayed negative aspects of feedback focused on a malfunction with the speech recognition functionality (used by EG1):

(23) In the questions that were multiple choice and I would have to say the correct answer, sometimes I would say one answer (that was correct) but it would hear a different one and mark me wrong. That was kind of annoying.

(24) There were some words that i [*sic*] repeated a dozen times and it just wouldn't pick up on it, but I couldn't figure out exactly where in my inflection that I was going wrong. In addition, sometimes, the entire sentence that it wanted you to repeat wasn't visible. Also a glitch during the fill in the blank section of the program.

(25) It was frustrating that it told me i [*sic*] was saying the word wrong, but it could not provide a correction on how i [*sic*] can fix it. I liked to be told I was wrong, but i [*sic*] would have preferred to know how i [*sic*] can do better.

Comments (23) and (24) describe situations in which their speech was not recognized by the speech recognition mechanism. They describe situations in which the participants knew they were accurate, as in (24); comment (24) also mentions the outcome of one of these errors: *it would hear a different one*. Additionally, from a lexical and emotional standpoint, words were used to describe these situations that indicate a higher level of *negative emotion*, such as 'annoying' (23), 'frustrating' (25), and 'wrong' (23–25). Comments (24) and (25) also appear to request feedback as they critique it: (24) *I couldn't figure out where [. . .] I was going wrong*; (25) *it could not provide a correction [. . .] I would have preferred to know how I can do better*. The emotions expressed in these comments, particularly frustration, likely negatively impacted the participants' performance; they took the focus away from language learning and placed it on the technology itself.

In contrast to EG1, there was only one response from EG2 that could be classified as a negative evaluation of feedback provided or the feedback process:

(26) I disliked that, as far as my knowledge, you could only complete the assignments one time. After it was submitted once, you could only review your

response instead of try again. If you could do it multiple times, since it is not graded, it might better benefit students in the learning aspect.

Comment (26) illustrates a technology issue that may not have been experienced by the majority of the participants, as there was no consultation with the investigator regarding this issue, though the participants could have taken the initiative to ask their individual instructors to reset the activities without ever mentioning it to the investigator. However, the comment at least shows that though this participant received feedback, they could not use it as it was intended once they closed the activity. Aside from the aforementioned situation, EG2 clearly had fewer frustrations with getting feedback from the technology; the differences in expressions of *negative emotion* and the technology becoming distracting is a trend that continues when considering all aspects of the technology as well.

5.3.3 Technological Issues, Suggestions for Improvement

Among the questions the participants answered related to the third topic, technological issues and suggestions for improvement, were *what did you dislike?* and *what would you improve?* in reference to their assigned technology. As the answers to these questions have high potential for overlap, participant commentary is presented here by participant group, addressing both technological issues (including negative impressions of the design of the technology) and suggestions for improvement together. In keeping with the sentiment analysis results, variation in emotion is observable in these responses, particularly for EG1.

As was discussed in Section 5.2 and illustrated in Figure 19 , the responses from EG1 contained a high frequency of words and/or descriptions relating to problems with technology:

(27) It was prone to glitch out so I couldn't even complete certain activities because they would crash on my phone. For the most part it was okay but sometimes not being able to complete the beginning of some of the activities was bad because I wasn't sure what I was doing.

(28) It did have a habit of making my phone freeze or blacking out while I was using the app.

(29) The only error that I noticed with the technology is sometimes the app would exit me out of a lesson I was working on at the time. This became very frustrating sometimes, but I was able to go back and skip that section of the less and continue with another.

(30) I did have some technical issues such as not being able to complete some lessons fully. It became frustrating after a while since it was such a great help to me.

The frequency of these comments, the words used to describe the issues, and the fact that these descriptions came from many, if not most, of the participants in EG1 exacerbates the problem described in the previous section. Not only was there an obvious focus on the technology problems, but the detail in some of the descriptions could lead to the conclusion that these problems were also memorable. Additionally, there is a high level of *negative emotion* expressed, some of it overtly, such as ‘bad’ (27) and ‘frustrating’ (29, 30), but also in the descriptions of the problems themselves, using specific words including ‘glitch’ (27), ‘crash’ (27), ‘freeze’ (28), and ‘blacking out’ (28). Participants also described the problems in more vague terms as well, such as ‘error’ (29) and ‘technical issues’ (30). This commentary further supports the claim that pervasive technology issues were distracting and negatively impacted performance.

Unsurprisingly, many of the suggestions for improvement coming from EG1 mentioned the aforementioned issues (e.g., ‘glitches’), but otherwise, the commentary either pertained to app design or specific features to add, often mentioning cost in some way:

(31) I would just fix some of the glitches like the one mentioned above as well as maybe make the lessons worth less solos since you can not [*sic*] earn that much in one lesson.

(32) Although difficult, it would be so helpful if the app could produce a free version with adds [*sic*] so that more people could benefit from it!

Comments (31) and (32) pertain to a choice made by the investigator to pay for enough *solos* to last through the whole semester for each participant in EG1; otherwise, as comment (31) points out, *you cannot earn that much*, which would likely have increased the participants frustration and potentially caused dissatisfaction with the app. Comment (32) suggests a feature that is common to most apps currently, ads, as a potential solution. Other comments from EG1 participants suggested specific elements they would like to see added:

> (33) I would add a feature that could provide specific feedback for the words i [*sic*] was struggling with.

> (34) I wish that it had activities to carry on conversations with others or to write complete conversations or phrases without having a word bank.

Comments (33) centers around improving the feedback by stating what specifically was incorrect, expanding on the binary "correct/incorrect" style of feedback already included. Sadly, comment (34) would have been partially addressed had there not been glitches with the fill in the blank activities, as those were often quasi-dialogue in nature (i.e., the app would play a question, then the user would provide the answer and could use the speech recognition feature to "converse").

In contrast to EG1, the comments from EG2 focused much less attention on technology and its problems. Instead, these participants focused more on the activities themselves and specific features, often describing how or why the features negatively impacted their learning or performance:

> (35) It was ok, there weren't a lot of speaking activities. The ones that asked a question and you had to respond were not helpful for pronunciation.

> (36) There wasn't a lot of activities for pronunciantion [*sic*] on VHL there was 3 maybe 4 activities per part of chapter and it wasnt [*sic*] really on the vocab but more on how to pronounce certain letters from the alphabet which wasnt [*sic*] really much help in understanding the words.

(37) Sometimes the ones that included sentence building and grammar flow were difficult to follow because the speaker spoke too fast and the sentences were so long they were hard to remember. This hindered my ability to work solely on pronunciation.

(38) Some of the activities where you had to answer a question after looking at a graphic or just by looking at a few words were difficult because I didn't know if I was correct or not. There were a few that I was pretty confused on and since they are graded together, I couldn't tell if I was on the right track or not.

These examples illustrate that EG2 appears to have had a more negative viewpoint of some of the features of the activities. Among the more common criticisms from EG2 were the rate of speech of the speakers in the recordings (37), and the length of the activities before receiving feedback (37, 38). The smaller number of activities overall was mentioned (36), but a specific activity type was mentioned with higher frequency and often criticized: receiving input, then recording the response (35, 38).

As was the case with EG1, some of the suggestions for improvement from EG2 mirror the criticisms mentioned previously. These suggestions included improvements based on criticism of the site itself (39), better timing of feedback (40), and, as in the comments critiquing the technology, a desire for the native speakers in the recordings to speak slower (41).

(39) My main dislike was honestly the appearance of the website, VHL. It did not make me want to get on and study Spanish. I have used IPhone [*sic*] apps to learn languages and it's much more physically enticing. VHL feels more like work and less like a benefit. I just personally believe more individuals especially in today's technological culture that if VHL was more attractive and easier to navigate.

(40) The technology would have been better if all of the activities included the correct answers right after I said a word or phrase.

(41) I would have the speakers have a normal speaking part and then a slower speaking part.

(42) It was good, but I think after the session there should be a meet up about the session. The meet-up is so the individual can learn what to fix in the next time

there is a recording. So a check-up after the session to give the student some advice.

Comment (42) is particularly revealing and relates to the set-up of the study; this participant suggested a meeting between the investigator and participant after the recordings to get feedback on their performance in the recordings. Though improvement with explicit (i.e., more than binary) feedback was not the purpose of the present study, the comment shows the participant's desire to improve and make use of resources; it also presents a related avenue of study for future research (see Section 6.6.2).

As the form-focused grammar activities within the Blackboard learning management system were the least interactive and sophisticated out of the three technologies, exemplified by comment (43), it is less surprising that CG's comments would focus on the nature of the technology itself and what it may be lacking:

(43) It was BlackBoard [*sic*], so it has the same disadvantages. The GUI was weird at points. Sometime the test would take an excessive amount of time to load. Sometimes BlackBoard [*sic*] would be down as a whole also.

(44) It was not the easiest to navigate. You could not just hit enter after you entered a response, you had to go back and hit the submit button and then got back and put the mouse in the text box in order to do the next problem. It made the assignments really tedious especially when there was over ten fill in the blank questions. It was also pretty much the same if not easier than our VHL homework, so sometimes it felt like a waste of time.

(45) It was very hard to input accents into Blackboard - I ended up just giving up and taking the lost points because I would have had to go through a whole rigamarole [*sic*] to get the accents inserted; on VHL they are accessible at the bottom of the screen.

There were only four comments that could be interpreted as suggestions for improvement; three are shown here, and the fourth reiterated the comment regarding accent marks found in (45). Two of the three comments compared their experience to the homework they completed for class (44, 45), with accent marks being a prominent source of frustration.

Similar themes were found among the suggestions for improvement. Commentary discussed the functionality of the Blackboard site, with descriptions including *negative emotion* words such as ‘tedious’ and ‘boring’ (46), as well as a general lack of motivation (46). Comment (47), specifically, represents a suggestion of how to fix an oft-repeated complaint regarding the availability and/or utility of accent marks; the reference to ‘copied and pasted’ implies what other comments stated outright: many participants found it difficult to manually input accent marks.

(46) I would make it easier to go from question to question. Blackboard just probably isn’t the best program to use for homework, since it is so tedious. And it isn’t pretty to look at, it looks quite boring. Which doesn’t make me want to do the homework anymore [*sic*] than I already do.

(47) As stated above, I would make common accent marks readily available in the description of the activity that can be easily copied and pasted for use in answering the questions.

(48) When I get a question wrong, I’d like it to highlight the parts I got right, so I know what part of the answer I’m having trouble with.

(49) I understand the importance of reading the same passage for the recordings but I think it would have been a little interesting to see some changes in the readings so that we could practice pronunciation of other phrases as well.

Comments (48) and (49), however, are a divergence from the aforementioned themes. Comment (48) reiterates a theme seen throughout the commentary in general: the desire to know the answer. However, comment (49) mentions the recordings spaced throughout the semester. As the recording sessions were the only times throughout the semester that this participant group completed any activity for the study that was related to pronunciation, it is understandable to desire new material to read aloud. Similar to EG1, the more negative tone in response to the technology likely played a role in the participants’ performance, in addition to their group being assigned the technology with no pronunciation element.

5.3.4 Impact on Performance

The fourth topic that was identified in the participant's responses was *impact on performance*. In the Technology Questionnaire (see Appendix G), the participants were asked if they thought interacting with the technology they were assigned affected their performance in their Spanish course. The commentary included in this section exemplifies the trends found in the questionnaire responses. In relation to language use, the participants appear to be self-focused (i.e., a high frequency of *I*-words[99]) and *positive emotion* was expressed more frequently than *negative emotion*, particularly in the two experimental groups.

The commentary pertaining to performance from EG1 appears quite positive overall. The topics covered in these comments were mostly split between two topics: more broad commentary pertaining to the classroom (comments 50 through 55), and specific commentary regarding evaluation (comments 56 through 58). Comments (50) through (52) reference the nature of the class in some way:

> (50) It was extremely helpful considering I had never taken a spanish [*sic*] course before. The lessons lined up with the lessons in class and i [*sic*] was able to have a basic understanding of the lesson before going into class.

> (51) I liked how it taught me a lot of vocabulary and the order it taught me grammar in was very similar to the pace my Spanish class was going so it was also a helpful study tool for the class.

> (52) I was really able to get an understanding before going into class and learning the full sentence. This was very helpful since I have never taken a Spanish course before and didn't have any past knowledge of Spanish. I was able to focus on the more difficult things in class rather than trying to learn the vocab.

As can be seen from the commentary, the participants expressed positive views of the app for first-time learners (50, 52), and appeared to find the app a useful tool for preparation (50, 51) and

99 (Tausczik & Pennebaker, 2010)

support for classroom learning (51, 52), and correspondence to the actual Spanish course (50, 51). Particular language learning elements are also mentioned, such as grammar (51) and vocabulary (52). Comment (52) also mentions that the app allowed for a shift in focus in the classroom; in other words, that using the app created a foundation that made better use of class time. While these comments were more general in nature, comments (53) through (55) provide more specific descriptions of the app's impact on the participants' class performance:

> (53) It absolutely did. I got extra practice in speaking, applying, and vocabulary. The app was very similar to the basic material that I was learning in class, so it transferred over perfectly into my class performance. It also made me feel more confident in the work i [*sic*] was doing in class and in participation.
>
> (54) Yes, it taught me new vocabulary that we did not learn in class. It gave me grammar tips and other tools that I was able to use in class and better my Spanish overall. It also helped me to hone my Spanish skills and helped me use Spanish practically.
>
> (55) The app used some different words for phrases than my teacher did in class so when I did the activities on my phone and used the words in class, my teacher was surprised that I had a little bit more vocabulary.

Other themes emerge in the commentary on the classroom, such as confidence and performance during in-class activities (53). Additionally, there were specific language elements that were important to many EG1 participants and supported by the app, exemplified in comment (54): vocabulary and grammar. Comment (55) also provides rare (for the present study) insight into the reactions from the instructors whose students participated in the study, and also suggests that interaction with the app likely resulted in general improvement.

Further discussion of improvement can be seen in comments (56) through (58). Unsurprisingly, as they received extra credit for taking part in the study, grades were of particular interest to the participants:

> (56) I felt like I was able to pronounce things better in class after using the technology for sure, and I liked that the video explanations on grammar

provided me with some background as to why things are phrased as they are in Spanish. It led to me doing better on the quizzes.

(57) It improved my performance in Spanish class because it served as a review for what I was learning in Spanish. I don't get a grade on pronunciation in Spanish so that didn't affect me in class all that much.

(58) No, it was like review after we already took a test. I always felt the activities were behind in the material in class.

Within these comments, pronunciation improvement is more of a focus for some, exemplified by comment (56). In contrast, there was also a negative comment regarding the app's potential influence on evaluated classroom activities (58). As was mentioned in Section 3.7.2, the students were provided a guide (see Appendix H for EG1, Appendix I for EG2) that, to the best of the investigator's ability, closely followed the progression of the course; however, the participants were informed that it was meant to be a *guide*, and they could use the app as it best served them. The most striking comment pertaining specifically to evaluation is comment (57), which describes two polarized opinions pronunciation improvement: practicing with the app improved class performance overall, but since pronunciation is not graded, extra practice had little effect.

Compared to EG1, the comments relating to impact on performance from EG2 had a different focus and were less positive overall. There were fewer comments pertaining to graded activities. Instead, the participants commented on a variety of subjects including accuracy (59), confidence (60), and accent (59, 63):

(59) Yes, because such a big part of learning another language is being able to articulate it to others. This technology allowed me to better understand what Spanish actually sounds like when spoken and what is expected.

(60) Yes, when I first started the class, I was nervous about speaking in class and I have a new-found boldness to speak in class with confidence.

(61) I do think using the technology improved my performance in class. To come down to Gorgas and read the same sentences about once a month was very helpful. I could tell every time I improved and knew how to say and even translate more and more words and phrases.

(62) I think it did a little bit because it is the same stuff we use in class, but not really because of the words I need for class. We also dont [*sic*] do a lot of speaking in relation to our grades.

(63) Not really, I think most of my pronunciation was influenced by my Spanish teacher's accent

While most of the responses were positive (59–61), a few were noteworthy due to their lack of positivity and/or their content. Comments (62) and (63) are the more negative comments from EG2 (in terms of their response to their assigned technology). Comment (62) reiterates an idea that has been mentioned in other comments: that speaking does not impact their course grades; that belief likely would impact their opinion of the technology in a more negative manner, as they might view the practice as superfluous or futile. Comment (63) conveys a different message about class. Despite this comment not being a glowing review of the technology, it does convey that the participant probably pays more attention in class in order to mimic the teacher's accent. Comment (61) referenced a completely different part of the study: the four recordings throughout the semester, indicating that the participant may have been self-monitoring during the recording sessions and perhaps between them as well, to be able to compare sessions four to five weeks apart. Overall, these responses were more positive, but clearly less so than those from EG1.

There was a more dramatic range of responses from the participants in CG when they were asked whether or not interacting with the grammar activities through Blackboard *improved their performance in their Spanish class?* The following responses begin at one end of the spectrum, a rather emphatic *yes*, and range all the way to a strong *no*.

(64) I used it as a way to review before exams, and it really improved my score when I used it regularly. The grammar activities went along with the modules

and chapters I was doing in class so it was great extra study and practice materials.

(65) Yes. Extra practice helped me on my homework. Whenever I got stuck, I was able to go to the course and find the same type of questions I was dealing with.

(66) The content matched up with what we were doing in class. That was nice.

(67) It did not prepare me for exams any better than VHL home work [*sic*] did.

(68) No. I did not gain from it that much. I actually had a friend who was also doing the Blackboard study and we complained about it together.

Comments (64–66) mimic the positive topics and rationales expressed by the previous groups, such as class pacing (64, 66) as well as a focus on grades (64–65). Comments (65) and (67) made comparisons with the technology being used by EG2 (as they were also using it to complete the homework for the course), and the participants do not appear to agree on the effect the grammar activities might have had on course performance. Comment (65), specifically, also likely references the number of attempts the students are given to improve their scores on any given homework activity (three (3) for first-year Spanish courses). Anecdotally, the lack of limitations on numbers of attempts was an unexpected positive opinion expressed to the investigator after a participant was assigned to CG during the meeting for the Baseline Recording (at the beginning of the semester).

Considering the language seen in comments (67) and (68), it is possible that the impact of *negative emotion* (discussed in Section 5.1.1) could be stronger for CG than what was originally inferred from the statistical calculations based on LIWC. Aside from being particularly negative, comment (67) illustrates another facet of opinions regarding technology that should be explored (further): influence from peers. While responses (64) through (66) at least partially demonstrate positive impressions or acknowledgments of the utility of the grammar activities, the examples included here represent a mix of emotions. Additionally, when compared with the emotion in the responses from the two experimental groups, the commentary from CG is, at best, tepid.

5.3.5 Rationales behind Recommendations

The fifth and final topic in the Technology Questionnaire[100] asked the participants whether or not they would recommend the technology, and why. Unlike the responses for previous sections, the differences between the three groups was more clearly defined by the commentary. EG1 continued to have very positive reactions to the technology (though *glitches* are still mentioned), and CG again reflected mixed emotions. EG2, however, appeared to be less enthusiastic than their commentary in other sections would suggest. Thus, the ranking of the technologies from most to least likely to be recommended would be: the app (EG1), the online activities accompanying the course textbook (EG2), the grammar activities (CG).

Regarding recommendations, EG1 participants continued expressing positive reactions to the app. The rationales behind their recommendation of the app, however, varied and provide insight into why they valued the app. Thematically, there were three main groupings: comparisons to other technology, language-specific features, recommendations with caveats. The comments that compared the Spanish SOLO app to other technology varied in what the participants deemed important:

(69) Yes, if they are committed to learning Spanish I would definitely recommend this app. I liked it better than Duolingo and found it convenient to use whenever it decided to work.

(70) I would recommend the technology to someone that is interested in learning Spanish because it is easily accessible and did not seem too expensive for all the lessons that are given to you.

(71) Most definitely. I love that the app combines seeing actual words, visual representations of words, and hearing words all simultaneously. This app would make it much easier for a person who has never learned another language than many other non-interactive programs.

[100] The last question asked if there was anything else the participant wanted to mention (see Appendix G).

(72) I would definitely recommend this app, because it is fun and actually helps you learn Spanish. It is easy to use and it did not really feel like I was doing actual school work. It was also really convenient because it was on my phone, so I could work on it whenever I had free time.

Comment (69) compared the app to other available language-learning technology (as did (71), implicitly), though it also included a terse comment toward the 'glitches'. Comment (70) focused on the cost associated with some of the technology available for use in learning other languages, finding the app to be rather cost effective[101]. Comment (71) shows a preference for the interaction in the app; as the participant mentioned sensory perception (i.e., seeing, hearing), it is possible this recommendation refers to the speech recognition feature, though it is not mentioned specifically, and if so, it is one of the only recommendations to do so. Comments (70) and (72) mention the ease of accessing the app. More of comment (72), however, is spent talking about enjoyment, mentioning the app being fun and not feeling like work; this is echoed in a comment from Section 5.3.1 that refers to the app being game-like. In terms of the investigation conducted in the present study, it is important to note that comment (71) one of the few comments from the participants in EG1 relating to orthography, and one of even fewer that mention orthography and hearing together.

Participants in EG1 also recommended the technology based on specific aspects of language learning, repeating those that have been mentioned in other sections: grammar and vocabulary:

(73) Yes, because it was very helpful in learning grammar and words and how to say them. It is also structured in a logical way to where someone would probably be able to learn basic Spanish solely using this app.

[101] The investigator purchased enough credits (or Solos, as the app called them) for the participants so that the participants could access all the activities at once without having to "earn" more credits (see Section 3.7.2).

(74) Yes because it forces one to say things out loud, pushing people out of their comfort zone. In addition, the visual representation and repetitiveness of the technology allowed for the vocabulary words to really stick

(75) Yes. The technology teaches you the basics. Verbs, how to spell them, meaning, and how to say them.

The comments regarding grammar and vocabulary appear to agree with each other in terms of vocabulary, but have different impressions of grammar. Comment (73) perceived the effectiveness of the app differently relating to how well it helped with learning grammar. All three comments reference vocabulary in some way; in the comments on recommendations overall, vocabulary was mentioned frequently. Comment (74) would recommend the app partially due to being *forced* to speak[102]; this is one of very few recommendations to mention speaking to the app. Comment (75), while it makes reference to 'the basics,' clearly mentions the meaning of words, which directly relates to vocabulary; this comment is another among a very select group of comments to mention both orthography and speaking/hearing.

Participants in EG1 also often recommended the app, but with small caveats in the form of (sometimes implied) qualifying statements:

(76) Yes. It starts off very basic and then gets harder. It helps you to learn how to speak the language correctly.

(77) I would if they were only learning the basics. The app is primarily focused on simple terms and phrases. It is a good beginning for learning the language. It is not meant for advanced Spanish [*sic*] learning however.

(78) I would, but have them start before taking the class. In class, we fly through the material so it is easy to get left behind with the app.

Comments (76) and (77) mention 'basic' in some form; though they vary in how critical they are of the app, they all at least imply that more advanced learners might not have as positive a

[102] The idea of being *forced* could reference the instructions given to participants in EG1 (i.e., that they were required to use the 'microphone button' (i.e., speech recognition functionality) whenever it was an available option (see Section 3.7.2).

reaction to the app (with comment (77) stating it outright). Comment (78) presents a different caveat, and is the only comment from EG1 to recommend the app with the clear expectation that the user would also be enrolled in a Spanish course (one other suggests it as beneficial for those taking the class).

The participants in EG2 expressed both positive and more negative opinions toward recommending their assigned technology. Most of the comments were more positive in nature:

> (79) I would recommend this to someone else because I think it really helped me to hear words and phrases rather than just to read them. Also, the videos about pronunciation that outlined rules for certain letter sounds was very helpful because the lesson applied to other parts of Spanish words that I had never seen before.
>
> (80) I would recomment [*sic*] it because it does have a good basis for helping people with the different styles and how the letters form words.
>
> (81) I would recommend this. The sections I worked on weren't necessarily targeted at new Spanish learners, but assuming there are more introductory courses, absolutely. This allows you to actually understand what you should be saying and teaches basic pronunciation, which would be very beneficial going into more difficult levels of Spanish.
>
> (82) I would recommend this technology to someone learning Spanish. For someone who doesn't know any of the Spanish language, this technology lays out a good learning pattern. It has vocabulary assignments that help you pronounce words and figure out their meaning, and it has grammar assignments that allow you to learn grammar concepts through thinking the questions through, which I think is the best way to learn. However, I would not recommend this be the ONLY thing they use, as some supplemental exercises may also be useful.

Comments (79) and (80) are of particular interest to the present study, as they are among the few in any group to refer to orthography. It is possible that the lessons referred to in (79) occurred earlier on; if that is the case, they had a lasting impression to be mentioned in a questionnaire at the end of the semester. The comments from EG2 also focused more on pronunciation than those from the other experimental group did, illustrated in (79), (81) and (82). Despite the positive recommendations, comments (81) and (82) do express some concerns about recommending the technology as well, indicating that although their feelings overall were positive, they acknowledged that there might also be some drawbacks to these kinds of activities.

Continuing with the topic of reservations about recommending the technology, in contrast to EG1, where recommendations were made with caveats, some EG2 participants implied or stated outright that they would not recommend the technology:

> (83) While it is a very interactive technology, there was a lot that had to be done or already known that the technology did not help with.
>
> (84) I would recommend it but at the same time I would tell them to only use it if they can learn on their own as well.
>
> (85) No because with out [*sic*] the class it would be very hard to get any understanding from it.

Comment (83) does not state it outright, but something being *unhelpful* would likely not make a positive recommendation; this recommendation is one of the rare mentions of problems with the technology itself from EG2. Comment (84) contains an *only if* statement, which limits the amount of people who might view the recommendation as positive. Conversely, comment (85) states outright that the technology would not be recommended. The rationale is insightful, however, and acknowledges that the online site was likely not designed to be used unilaterally. As was the case with EG1, the participants overwhelmingly favored the technology.

The same cannot be said, however, for CG. The only comments included here are ones that could be considered to relate specifically to recommending the technology to someone else:

> (86) I would recommend the technology to someone else, I believe if used in the right ways this could help many people practice Spanish without having as harsh of consequences for getting things wrong and affecting their grade in Spanish.
>
> (87) I would recommend it because repetiton [*sic*] helps learn the language.
>
> (88) I would recommend it to someone that is purely a visual learner. It is a lot of words, and not many pictures or listening, so it might be harder to use in that

aspect. I would not recommend it for being the sole source of Spanish education, but it is a good supplement to a course.

(89) I'm not sure I would recommend this technology over something more adaptive and interactive.

Despite there only being four comments, they are fairly varied, ranging from more positive, (86) and (87), to neutral (88), and then to more negative (89). Comment (86) introduces a class-related rationale for recommending the technology: *harsh consequences for getting things wrong and affecting their grade*. It is probable that this comment relates back to one discussed previously regarding the homework activities in the first-semester Spanish course. If that is the case, then the *harsh consequences* reference the finite amount of attempts allowed on their course homework.

5.4 Language Analysis Discussion

5.4.1 Perceptions of Currently Available Technology for Foreign Language Pronunciation Practice

The present study attempted to address the question: *What are students' attitudes toward and/or perceptions of currently available technology for foreign language pronunciation practice?* Overall, the responses to the Technology Questionnaire indicate that all three of the technologies investigated in the present study were positively received by the participants, to varying degrees. However, sentiment analysis using LIWC and language analysis of the survey responses by topic revealed differences between the three groups in terms of their feelings toward and opinions of the technology they interacted with during the study.

The LIWC categories investigated in the present study were: *positive emotion, negative emotion, pronouns, tense,* and *tentative language.* Additionally, *affect* and *emotional tone* were included as they both subsume *positive* and *negative emotion*. The categories that illustrated language use that likely revealed something significant about the participants' experiences were:

emotional tone, affect, negative emotion, sadness (a special designation within *negative emotion*), and *focus present*. Some of the other categories demonstrated an expected use of language based on the task given to the participants. For example, Tausczik and Pennebaker (2010) note that use of *I*-words denotes attentional focus on the self, as well as a potential for perceived lower status; both of these explanations make sense pertaining to the conditions of the present study, as the majority of the Technology Questionnaire focused on self-reflection, and there was a power dynamic between the investigator and the participants, since they were receiving extra credit for participating, and were aware that the investigator determined how much they would be awarded. Additionally, the use of tentative language was not unexpected, and could (in the participants' minds) relate to the differences in power, as they may have been hesitant to fully express or commit to their opinions regarding the technology for fear of it negatively impacting (Medimorec & Pennycook, 2015) their reward for completing the study.

The findings from the other LIWC categories were less predictable. Beginning with categories relating to emotion, the analysis revealed that the overall emotional tone for all three participant groups was quite high. Additionally, there appeared to be relatively high use of *affect* words across participant groups. However, further analysis of these categories revealed something surprising: the group "outperforming" (i.e., whose responses generated the highest emotional tone scores and the largest mean percentage of use of *affect*) the others was CG, and the group at the opposite end of the spectrum was EG1. Simply reading the participant responses would have led to an interpretation of generally positive feelings from all three groups, but the LIWC analysis revealed a trend inverted from what would generally be expected. Despite using the most advanced of the three technologies included in the present study, EG1 showed a difference in use of *affect* words that approached significance when compared to CG, with CG having a higher mean *affect* percentage than EG1. Similar trends were seen pertaining to *negative emotion*: in this instance, CG utilized more *negative emotion* words than the other two groups, with EG2 using the least and EG1 falling in the middle. This effect was statistically significant between CG and EG2, and approached significance between EG1 and CG. Moreover, considering words tagged by LIWC as *sad(ness)* (a subset of the *negative emotion* words), CG used these words significantly differently than EG1 and EG2, though the frequency of words in this category was very low to begin with. Surprisingly, despite utilizing more words associated with negative emotions than EG1, the effect was not strong enough to affect the emotional tone score of CG.

CG also outperformed EG1 in terms of use of words designated as *focus present*, a category in which higher usage has been associated with positive impressions (Rúas-Araújo et al., 2016). The language use for *focus present* followed the same trends as the others: CG utilized words in this category statistically significantly differently than EG1. Calculating the mean of the responses for each group, it was revealed that, on average, CG utilized more words associated with *focus present* than EG1. This usage, and its interpretation as being more positive (Rúas-Araújo et al., 2016; Tausczik & Pennebaker, 2010), fit the pattern found in the language related to emotionality: in categories connected to positivity, higher usage was found in CG (though it would likely be expected to be found in one of the experimental groups).

That is not to say that the LIWC analysis should be taken as an interpretation that CG had more positive impressions of their technology than the other two groups, but rather that EG1 appears to have had a more negative reaction to their technology than was originally expected. EG2, on the other hand, was more consistently situated between the other groups. An analysis of the lexical content of the responses revealed a pattern that the present study interpreted as a likely explanation for the differences found in the sentiment analysis. Specifically pertaining to technology-related words, the present study found that there was a highly statistically significant difference in their use across all three groups, and more specifically between EG1 and CG as well as between EG1 and EG2 (usage between EG2 and CG was also statistically significant, but with a larger *p*-value). In this instance, EG1 "outperformed" CG and EG2, using more technology-related words and likely indicating what aspects of the app drew their focus. EG1's focus became even clearer when the group of technology-related words was further narrowed down into words that pertain specifically to technology problems, which revealed that EG1 nearly unilaterally used words referencing problems. Though the technology words themselves were not tagged as *negative emotion* by LIWC, the prevalence of these words in the responses from EG1 provides an explanation as to how their responses could have resulted in lower emotional tone (among other categories) than CG and EG2.

Beyond the linguistic and sentiment analyses, the responses themselves also support this interpretation after being classified by topic. The present study investigated the topics: *positive impressions, impressions of feedback, technological issues and/or suggestions for improvement, impact on performance,* and *rationales behind recommendations*. As was indicated by the high emotional tone scores in each group, participants from all three groups held (at least some) positive impressions of their assigned technology, with the responses from EG1 being the most enthusiastic, followed by EG2 and then CG. This trend continued for the positive portion of *impressions of feedback*, with frequent mention of the participants' most favored aspects of the technology. However, an emotional shift began to appear beginning with the responses that detailed negative impressions of feedback. Within these responses, and particularly in the case of EG1, frustration and irritation were clearly communicated. For EG1, many responses mentioned problems with the speech recognition functionality (a feature that many praised previously); these problems often centered on limitations with the technology itself, as the speech recognition mechanism sometimes did not recognize their speech, or marked a correct utterance as wrong, for example. In comparison to EG1, the other two groups had nowhere near as severe of responses.

The topics *impact on performance* and *rationales behind recommendations* were very revealing in terms of the participants' opinions regarding the utility of the technology as it pertains to themselves and by extension, others in similar language-learning situations. Regarding impact on performance, there was a clear divide between the three groups: EG1 mostly felt strongly that the app improved their performance in Spanish class, while EG2 participants were divided, with some believing that it was beneficial and others finding it less so. Most participants in CG, on the other hand, did not find that the grammar activities on Blackboard helped them very much, though multiple comments referenced that the activities helped with their homework as, unlike their homework activities, they had unlimited attempts at the activities for the present study.

The separation between the opinions of the groups was also reflected in the topic *rationales behind recommendations*. Participants in EG1 nearly unilaterally (and enthusiastically) stated that they would recommend the technology to someone else who wanted to learn Spanish. Participants in EG2, however, introduced more caveats or reservations into their recommendations, leading to the conclusion that their impressions of the listen-and-repeat style activities that accompanied their course textbook were less positive than they appeared at first. Participants in CG, in contrast, generally showed more hesitation to recommend the technology, with at least one response nearly being an outright *no*. The rationales for recommending the technology to others seem to reveal the most regarding the participants' attitudes toward their assigned technology, as the technology cannot be considered successful if no one wants to use it.

Overall, the participants appeared to react positively to their assigned technology. However, sentiment and language analyses uncovered revealing differences between the groups, particularly pertaining to emotional language and use of words relating to technology (and problems with technology). Finally, much was revealed in commentary separated by topic, including lukewarm feelings toward the utility of the technology for improving class performance from EG2 and CG, as well as the hesitancy to recommend the technology from the same two groups. Therefore, the present study concludes that despite lower scores in sentiment analysis, the app utilized by EG1 still is the most preferred technology of the three, and without the technology issues that served as distractions, the differences in preference probably would have been more pronounced. The second most favored technology would be the listen-and-repeat style activities utilized by EG2; commentary from EG2 appeared to fall halfway between that of EG1 and CG in terms of the sentiment analysis, and while EG2 did not face the same technological difficulties as EG1, they also did not appear to favor the technology as much as EG1 favored theirs. The grammar activities completed on Blackboard by CG appeared to be the least favored of the three technologies, though some participants did report finding them useful. However, simply because language learning technology is well-liked and/or well-received does not necessarily indicate that it can serve its intended purpose (Youngs et al., 2011).

5.4.2 Connection between Evaluation of Technology and Pronunciation Performance

The present study also attempted to address the question: *Is there a connection between impressions of and/or attitudes toward using the app and improvement in pronunciation?* The findings based on the language analysis of the Technology Questionnaire indicate that there is likely some connection between impressions of technology and pronunciation performance. In contrast to the findings in Chapter 4 (where EG2 outperformed the other two groups), the more extreme results in this chapter pertain mainly to EG1, but there were also some surprising findings from CG as well. Overall, however, the findings of the present study appear to support the claim that positive impressions of technology do not necessarily translate into pronunciation improvement.

Previous research in the CALL/MALL field has concluded that functionality such as speech recognition is something that learners desire to see as an option in the technology they utilize to assist with language learning (Culbertson et al., 2016), and that of the myriad of technology incorporated into foreign language learning (and often the classroom), the most useful has been utilizing speech recognition for pronunciation practice (Golonka et al., 2014). However, research has also concluded that technology can have a positive impact on the language learning experience, but only if the technology is utilized based on solid pedagogical and linguistic theory (Youngs et al., 2011). While *linguistic theory* could possibly be extrapolated out to include speech recognition *accuracy*, the findings of the present study suggest a third area to add to the claim made by Youngs et al. (2011): technological stability. Studies have included technological functionality as a consideration when evaluating an app (Martín Monje et al., 2014), but with advances in technology being utilized for language learning purposes, including ASR (McCrocklin, 2016), technological stability would support pedagogical and linguistic theory (Youngs et al., 2011) by ensuring that the platform being used does not detract from the learning process, and that any special features (i.e., speech recognition) are not only stable in terms of the platform used to access the features, but also are well-informed from a linguistic standpoint (Kolesnikova, 2017).

In the phonetics portion of the present study, the participants in EG1 often outperformed their counterparts in CG, but rarely were so successful in comparison to EG2. In contrast, in the language analysis portion, EG1 appears to have the more extreme results, for better or worse. On the positive side, the participants in EG1 had much more enthusiastic responses when asked whether using the app impacted their performance in their Spanish classes (to which the response was overwhelmingly positive), and when asked if they would recommend the app (which, again, received a highly consistently affirmative response). They also appeared to express positive impressions of the app overall; the speech recognition functionality and its capacity for providing feedback was received positively as well, though with some notable exceptions. However, the general impressions from reading the participants' comments does not appear to align with the analysis of the sentiments expressed through word choice and usage.

The analysis using LIWC revealed that there were specific variables that illustrated (nearly) significant differences in usage between EG1 and CG (i.e., the group that did not interact with any pronunciation elements). Tausczik and Pennebaker (2010) explain that there are certain elements of language use that can indicate underlying thoughts or emotions; of interest pertaining to EG1 are *emotional tone, affect, negative emotion,* and *focus present*. Of these four categories, three of them have positive connotations. However, despite all the positive commentary from EG1, the mean response percentages for EG1 are lower than those of CG in these areas. In terms of *negative emotion*, however, EG1 falls in the middle between CG (which demonstrated the highest average *negative emotion* usage) and EG2 (which demonstrated the lowest). Aside from emotionality, the findings of the present study support previous claims that higher use of present tense is related to lower performance (Robinson et al., 2013), as EG1 generally outperformed CG in terms of pronunciation improvement but used less present tense.

The explanation for these differences in categories that are often associated with emotions, however, is expected in the responses to two questions (one possibly more than the other) as well as within the frequency of the individual words. The main question where an explanation is likely to be found is in the answers to *what did you dislike about the technology?* Many of these responses focused on feedback and were thematically very negative; despite the individual words not being tagged as such in LIWC, the frustration and irritation of the participants was evident in their explanations of the issues they dealt with. The ideas communicated within these responses often mentioned instances where the technology failed to function properly (e.g., the speech recognition function would not recognize their speech, or it would evaluate their utterance as incorrect despite it being correct). It is highly likely that these kinds of experiences were quite distracting in addition to being frustrating, taking the participants' focus away from language learning and pronunciation practice and/or training. Further supporting this interpretation is one specific pattern found within the individual words themselves. When analyzing just the words pertaining to technology, there was a highly statistically significant difference between the three groups. Additionally, when only considering words pertaining to technological issues, it is evident that EG1 focused on the technology and its issues far more than the other two groups, as all but one instance was found within the responses from EG1.

Another possible interpretation of the differences between EG1's and EG2's use of *negative emotion* could be linked to the style of the activities and how supportive the participants believed them to be. As indicated in Table 4, in addition to being able to practice pronunciation, out of the two experimental groups, EG2 had the most speech modeled for them; in other words, in the majority of the activities, the speech EG2 participants were prompted to produce was modeled for them in some way (usually before, but sometimes after the students' attempt). In comparison, some of the activities included in the app (and completed by EG1) would play speech samples (i.e., recordings of native speakers pronouncing the words), but in many activities, participants in EG1 had to produce speech without hearing an example first. One could assume that these differences affected the participants' opinions of the technology; since EG2 likely felt more comfortable and secure due to the pronunciation being modeled for them, they could have had more positive feelings about the technology than EG1, leading to lower use of *negative emotion* words when evaluating the technology. While the idea of evaluating the use of modeling in a similar manner (to that of the present study) is an area for future exploration, it is difficult to make such a claim about the impact of modeling from the results of the present study. Due to amount of focus drawn to the technological issues experienced by EG1, which appeared to negatively influence both their pronunciation improvement and opinions, it is unclear whether EG2 would still have performed differently from EG1 if the technology issues had not occurred. It is possible that the observed outcome of more pronunciation improvement and less use of *negative emotion* was not (completely) caused by EG2 truly outperforming EG1, but rather because EG1 experienced so many distractions that it detracted from what would have been comparable performance otherwise.

CONCLUSION

This chapter begins with a restatement of the present study's research questions, after which the findings and conclusions are summarized in reference to the research questions. Subsequent sections address contributions to various fields, limitations to the present study, considerations for future study, and recommendations for future research. Last, this chapter ends with the final thoughts regarding the impact of the present study.

6.1 Restatement of Research Questions

The present study aimed to investigate pronunciation improvement in beginning learners of Spanish via training utilizing speech recognition functionality as compared to training utilizing currently-available technology. Specifically, the present study focused on problematic consonants (i.e., those for which graphemes and phonemes do not align) and the three voiceless stops (/p t k/). As can be found in Section 1.1, the particular questions the present study aimed to answer were:

(1) To what extent does the use of speech recognition software when practicing pronunciation with a language learning app impact beginning students' improvement of problematic areas of pronunciation in the L2?
- (1a) In which of the problematic pronunciation areas under investigation can significant improvement be found?
- (1b) Does training using speech recognition (and its feedback complement) produce better results than using already available technology?

(2) What are students' attitudes toward and/or perceptions of currently available technology for foreign language pronunciation practice?

(3) Is there a connection between impressions of/attitudes toward using the app and improvement in pronunciation?

Section 6.2 summarizes the findings and conclusions drawn from the present study as they pertain to each of the research questions. The answers to Research Question (1), with its accompanying questions (1a) and (1b), were found in the analysis of the articulatory and acoustic phonetics elements (see Chapter 4), while the answers to Research Questions (2) and (3) mostly came from the analysis of the language found in the answers to the Technology Questionnaire through which the participants expressed their sentiments toward the technology (see Chapter 5). The answer to Research Question (3), however, also required comparison between the answers to the other research questions.

6.2 Summary of Findings and Conclusions

6.2.1 Research Questions (1), (1a), and (1b): Impact on Pronunciation

In response to Research Question (1), the present study found that the use of speech recognition software when practicing pronunciation with a language learning app (i.e., membership in EG1) did produce some improvement over time as compared to the participants' own performance prior to receiving the treatment. For the grapheme-phoneme mismatches, when treating the Baseline Recording and Recording 3 (i.e., pre-test and post-test) as an entire test completed by each participant, there was a statistically significant increase in accuracy over time. Additionally, and in answer to Research Question (1a) specifically, when analyzing accuracy improvement for each grapheme individually, participants in EG1 demonstrated statistically significant improvement in accuracy (as compared to their own performance on the Baseline Recording) for more graphemes (i.e., five: <c{i, e}>, <g{i, e}>, <h>, <ll>, and <ñ>) than EG2 (3) and CG (2).

Despite showing improvement over time in comparison to themselves, membership in EG1 (i.e., utilizing the app containing speech recognition functionality) typically did not significantly increase the odds of a target response. Membership in EG1 increased the odds of a target response for the graphemes <c{i, e}> (significantly), <h>, <ñ>, <j>, and <z>, but decreased the odds for <c{a, o, u}>, <g{a, o, u}>, <g{i, e}>, <ll> (significantly), and <ch>. Analysis based on features (i.e., place or manner of articulation) revealed an increase in the odds of a target response for Fricatives, but a decrease for Stops and Palatals; however, none of these were statistically significant. The regressions performed on the contrastive analysis groupings appeared to show somewhat more positive results; though still not statistically significant, increases in the odds of a target response were found for sounds that do form part of the L1 phonological inventory, as well as for graphemes that have more than one phoneme associated with them in the L1 generally, and graphemes with two options and four options specifically showed an increase in odds.

Pertaining to VOT, participants in EG1 also demonstrated some improvement over time (i.e., a decrease in VOT values) as compared to themselves. There was a significant difference in VOT duration for /p/, /t/, and /k/ in word-medial position, but not for word-initial position. Subsequent comparisons of post-test VOT values across the three participant groups revealed that the groups performed statistically significantly differently from each other for /p/, /t/, and /k/ in word-initial position, but only for /p/ and /t/ in word-medial position. However, EG1 only showed a significant difference in improvement (as compared to CG that completed no pronunciation-related activities) for /t/ in word-medial position. In terms of group membership predicting a change in VOT, membership in EG1 did not produce statistically significant results. For /p/, membership in EG1 predicted a decrease of approximately two milliseconds, and approximately one millisecond for /t/. However, membership in EG1 predicted an increase in VOT of approximately 0.04 milliseconds for /k/. Overall, it appears that interaction with speech recognition functionality positively impacted the participants' pronunciation generally, and statistically significantly when compared to their own performance at the beginning of the semester, but did not typically produce statistically significant differences in comparison to the performance of CG, in which the participants did not practice pronunciation.

Answering Research Question (1b) required comparison between performance using already-available technology (i.e., participants in EG2) and the app utilized in the present study (i.e., participants in EG1). Despite showing some improvement, participants in EG1 often improved statistically significantly less than those in EG2, who interacted with the pronunciation practice activities currently available (i.e., accompanying their course textbook). EG2 also performed statistically significantly differently between their pre-test and post-test when looking at all the graphemes together, but as was mentioned previously, in comparison to five graphemes for EG1, EG2 only demonstrated statistically significant improvement over time (in comparison to themselves) for three graphemes: <g{i, e}>, <h>, and <ñ>. In terms of odds of a target response based on group membership, however, EG2 outperformed EG1, showing a statistically significant increase in the likelihood of a target response for <c{i, e}>, <g{i, e}>, <h>, <ñ>, and <z> (<g{a, o, u}> also showed an increase in odds, but not significantly); specifically, in comparison to EG1, membership in EG2 reversed the directionality of the odds for <g{a, o, u}> and <g{i, e}>. Surprisingly, three of the four graphemes for which EG2 membership produced a decrease in odds produced the same result in EG1: <c{a, o, u}>, <ll> (significantly), and <ch>. EG2 membership decreased the odds of a target response for <j> (whereas EG1 membership increased the odds). When analyzed by feature, EG2 membership produced the same directionality of odds as EG1 for Palatals and Stops (i.e., a decrease) as well as for Fricatives (i.e., an increase); however, in contrast to EG1, the increase in the likelihood of a target response pertaining to Fricatives was statistically significant for EG2. In terms of analysis from the perspective of English as the L1 (i.e., contrastive analysis), membership in EG2 increased the likelihood of a target response in all instances except for graphemes with only one phonemic association in the L1. The effect was statistically significant for: sounds that do form part of the L1 phonological inventory; graphemes that have more than one phoneme associated with them in the L1; and specifically, graphemes that have two and three phonemic associations in the L1.

Beyond mismatches, the analysis of VOT revealed that EG2 often outperformed both EG1 and CG. Pertaining to their own performance over time, participants in EG2 demonstrated statistically significant decreases in VOT over time for all three voiceless stops (/p t k/) in both word-initial and word-medial positions. Additionally, by the end of the semester, EG2 nearly unilaterally outperformed both EG1 and CG. EG2 showed statistically significantly different (i.e., improved) VOT productions than CG for /p/ in both word-initial and word-medial positions, as well as for /t/ in word-medial position, and /k/ in word-initial position. In comparison to EG1, EG2 showed statistically significantly improvement for all three voiceless stops in word-initial position, and for /p/ and /t/ in word-medial position. In terms of predicting an increase or decrease in duration, in contrast to EG1, EG2 group membership predicted a decrease in VOT duration for all three voiceless stops, specifically a decrease of approximately six milliseconds for /p/, approximately eight milliseconds for /t/, and four milliseconds for /k/.

Overall, interaction with the app containing speech recognition functionality did produce improvement in comparison with the other technologies, but less than the pronunciation activities found in the online companion to the *Protagonistas* textbook. The present study found that the app did assist with pronunciation improvement over time, as EG1 significantly improved in their accuracy pertaining to grapheme-phoneme mismatches when all the mismatches were considered together, as well as for some individual graphemes. However, analysis of the post-test recordings revealed that much of EG1's performance was more similar to that of CG (which did not interact with pronunciation technology) than to EG2. Additionally, though group membership in EG1 often did slightly increase the odds of a target response, the effect was not strong enough to support the claim that EG1 membership would better produce a higher likelihood of pronunciation accuracy. Considering VOT, EG1 did demonstrate improvement (i.e., reduction in duration) over time, and group membership in EG1 was shown to predict a decrease in VOT for two of the three voiceless stops, although not statistically significantly. Therefore, the present study concludes that while practicing pronunciation using the app did produce improvement in pronunciation within EG1 itself, the effects were not comparable to already-available technology.

6.2.2 Research Question (2): Attitudes toward Technology

In answering Research Question (2), the present study found deceptively positive attitudes toward the technologies utilized therein. Overall emotional tone (as calculated by LIWC) would indicate positive feelings gleaned from the responses to the questionnaire, but subsequent analyses of LIWC categories indicated that CG used more language associated with positivity than EG1 and EG2. CG also utilized more language associated with *negative emotion*, but that disparity did not have enough of a negative impact on overall emotional tone to move the emotional tone scores to levels lower than those of EG2 or EG1. Analysis of lexical items and word frequency, however, shed more light on the issue: EG1 utilized technology-related words (and within that group, technology-problem related words) with significantly higher frequency than EG2 or CG. The sustained focus on the technology, as suggested by the language use from the questionnaires, did not appear to aid the formation of positive impressions of the technology for EG1. Finally, the analysis of the responses separated by topic revealed trends that likely impacted the sentiment analysis (i.e., LIWC) results. First, feelings of frustration and irritation were clearly conveyed in the responses from EG1 pertaining to negative feedback from the app, and issues with the technology itself (e.g., *glitches, crashing*) were mentioned frequently. While the participants in the other groups did express dislikes pertaining to their assigned technologies, those opinions were not as severe. However, some of the most revealing comments were found within the topics *improvement of performance* and *rationales behind recommendations*. In spite of the findings from the sentiment analysis, the responses from EG1 clearly indicated that they believed the app improved their performance, as well as that they would definitely recommend the app to others. In contrast, the other two groups were less enthusiastic. EG2 appeared to be more neutral in terms of the impact on performance and qualified their recommendations in many cases. CG participants, however, mostly did not believe their technology positively impacted their performance, and were more likely to decline to recommend the technology. Thus, despite the glitches and problems encountered by EG1, the app appears to be the most well-received of the three technologies, as well as the more favored out of the two focused on pronunciation. In comparison to the technology already in use, it appears that the participants using the app found it highly useful and beneficial; however, the preference for the app does not appear to translate into greater improvement in pronunciation performance, at least not in comparison to the technology already in use.

6.2.3 Research Question (3): Connection between Attitude and Pronunciation Improvement

In addition to exploring the opinions of the participants in relation to the technology they used, the present study also attempted to determine whether impressions of and/or attitudes regarding the use of the Spanish SOLO app can be connected with an improvement in pronunciation (Research Question (3)). To answer this question, an analysis focusing primarily on EG1 was conducted, with some comparison to CG, as that group did not complete pronunciation activities at all. From simply reading the written responses to the Technology Questionnaire, it would appear that EG1 had much more positive impressions of the app than CG did of the grammar activities completed on Blackboard. However, the language analysis revealed that there were three LIWC categories that are typically associated with positive emotions or outlooks in which EG1 had average scores lower than those of CG: *emotional tone, affect,* and *focus present*. High scores in *emotional tone* and high frequency usage of words tagged by LIWC as *affect* both represent positive emotional responses. Additionally, high frequency use of *focus present* words has been associated with more positive impressions (Rúas-Araújo et al., 2016; Tausczik & Pennebaker, 2010), but also with lower academic performance (Robinson et al., 2013), which was also reflected in the (not always statistically significant) differences in pronunciation performance between EG1 and CG. Surprisingly, the scores in *negative emotion* placed EG1 between CG and EG2, with CG showing the highest usage of *negative emotion* words and EG2 the lowest. Despite this fact, CG still demonstrated more positive language use than EG1, though it did not translate into better performance.

In order to determine which factors may have led to these differences in language use, it was necessary to closely read the questionnaire answers themselves. In doing so, the present study uncovered some trends in the negative evaluations of the app from EG1 as well as in the use of technology-related words. The negative evaluations of the app (which were often mirrored by the responses to the question that asked for suggestions for improvement) often focused on one specific and one more general area: feedback and problems, though the two areas do overlap. The feedback responses often pointed out issues pertaining to the developmental issues found in the speech recognition functionality; often, participants in EG1 described situations in which they were frustrated or annoyed because the speech recognition did not recognize their utterance, or it did, but classified it as incorrect when the participants knew they had pronounced the word or phrase correctly. The commentary regarding technological problems centered around the functionality of the app in general, as (early on in the experiment) it was discovered that one particular type of activity consistently caused the app to crash. It would appear that this type of activity was not the only element that caused the app to have problems, or to 'glitch,' 'crash,' 'black out,' 'freeze,' or other words that could describe technology problems. All of these words, along with others, were found and used rather frequently in the questionnaire responses, leading to a statistical analysis that attempted to determine whether there was a statistically significant difference in the use of technology-related words between the groups (which there was), as well as a statistically significant difference in the use of words specifically pertaining to technology problems (which there was as well). Specifically, EG1 outperformed the other two groups in terms of their usage of both technology-related words and technology-related words relating specifically to technology problems, so much so that the words related to problems occurred within the EG1 responses all but once.

The combination of the more negative perspective found within the LIWC variables, combined with the analysis of the responses themselves as well as the frequency of words found within them, led the present study to conclude that a plausible explanation for less improvement in pronunciation than EG2, and only marginally more improvement than CG, could be due to their focus on the technology rather than the pronunciation practice/training. For comparison, the group that improved the most in terms of pronunciation, EG2, did not have as positive of impressions overall of their technology (i.e., they liked it, but were less likely to recommend it to others and varied in their opinions regarding whether or not it had improved their performance in their Spanish classes). However, they also did not focus on the technology nearly as much as EG1. Additionally, despite not interacting with pronunciation-related technology during the experiment, CG did show some improvement from pre-test to post-test; for EG1 to have statistically significantly outperformed CG in only a few areas suggests that the technology was enough of a distraction as to lower their potential for improvement closer to levels seen from a group that did not practice/train pronunciation at all. Therefore, the present study concludes that there is likely a connection between the participants' impressions of and/or attitudes toward using the app and pronunciation improvement (or lack thereof). Specifically, positive impressions of the app (i.e., liking it and its features, believing it to have improved class performance, being willing to recommend it to others) do not necessarily translate to improvement in concrete language areas (i.e., pronunciation), particularly if the technology itself becomes a distraction.

Overall, all of the participants in the present study appeared to find at least some aspects of their assigned technologies to be helpful or beneficial to them in some way. Despite the study's focus on pronunciation and its interaction with language learning technology, the commentary from the participants also revealed other elements that are important to understand. First, throughout the three groups, there was consistent mention of the words 'grammar' and 'vocabulary' specifically, as well as a repeated focus on *knowing the right answer*. The pervasiveness of this commentary represents something that is apparently very important to adult language learners, and is an area where technology could be useful. Relatedly, there was much focus on grades and grading, and while 'grammar' and 'vocabulary' were positively associated with grades, commentary regarding grades and pronunciation indicates that pronunciation may not be seen as valuable due to it not being considered something on which students receive grades. Therefore, students' views on pronunciation itself, as well as their interpretations of its importance within the foreign language classroom, should also be explored in order to better integrate pronunciation and the technology that can be utilized for its practice into L2 learning and instruction.

6.3 Contributions to the Field

6.3.1 Investigation of Consonants

One of the biggest contributions of the present study to the field of pronunciation in second language acquisition is the investigation of consonants. Most studies that investigate English and a Romance language focus on vowels, typically due to the sheer difference in the number of vowels in each language (e.g., in comparison with Spanish or Italian), or due to the presence of vowels in one language that do not occur in the other (e.g., studies investigating English and French). Few studies focus on consonantal differences between these languages (Kolesnikova, 2017), and even fewer (if any) take into account the role that orthographic intuition (from the L1) might play when encountering consonants and their orthographic representation in a second language. However, the present study has shown that orthography can impact L2 pronunciation, especially when the L1 and L2 differ in terms of transparency of correspondence between grapheme and phoneme, which is the case for English (Ellis & Hooper, 2002; Timmer & Schiller, 2012) and Spanish (Carreiras et al., 2014; Hualde, 2014; Meschyan & Hernandez, 2005).

Additionally, the present study investigated an oft-studied phonological phenomenon (VOT) in an atypical way. As has been mentioned previously, L2 VOT studies typically investigate a) perception (Alves & Luchini, 2017, 2020; González-Bueno, 1997b; Zampini, 1998), b) explicit instruction (Aliaga-García, 2007, 2017; Alves & Magro, 2011; Camus, 2020; Counselman, 2015; González-Bueno, 1997a; González López & Counselman, 2013; Kissling, 2013; Kupske & Oliveira, 2020; Offerman & Olson, 2016), or c) immersion settings (Hanzawa, 2018; Harada, 2007; Netelenbos et al., 2016). Less common ways of investigating L2 VOT involve longitudinal classroom studies with no instruction (Nagle, 2019), or a snapshot-like comparison between native and non-native speaker durations (Flege & Hillenbrand, 1984).

The present study also investigated VOT somewhere between explicit instruction and a lack thereof. While it was abundantly clear to the participants in both experimental groups that they were supposed to pay attention to their pronunciation, they were not told to focus on the aspiration of voiceless stops. Additionally, the practice and/or training the participants partook in throughout the semester increased their exposure to Spanish pronunciation more than it otherwise would have in typical foreign language classrooms. CG's VOT duration improving pre-test to post-test supports the claim that exposure to the language can impact learners' L2 VOT pronunciation (Nagle, 2019). However, the fact that at least one (and sometimes both) of the experimental groups significantly outperformed CG in each stop in each word location by the end of the semester indicates that VOT can be positively influenced by pronunciation practice even if the activities do not focus on VOT or /p t k/ generally.

6.3.2 L2 Instruction and the Acquisition of Pronunciation

Unlike second language acquisition in children, adults who acquire another language tend to be more cognitively aware, and thus may be more likely to think about the differences between the language they already speak and the one they are learning, as research has shown that metalinguistic awareness can have positive effects on pronunciation (Counselman, 2015; Kissling, 2015; Simon & D'Hulster, 2012). However, despite these differences, foreign language instruction to adults often deemphasizes the L1, perhaps in an attempt to foster more and/or consistent communication in the L2. The findings of the present study suggest that ignoring learners' L1 fails to take advantage of multiple opportunities.

Based on the findings in the present study, contrastive analysis produced similar results to the analysis of individual graphemes. The present study revealed that not every grapheme-phoneme mismatch improves with pronunciation practice alone. How many of the mismatches did (or did not) significantly improve varied by technology, but irrespective of technology, the findings of the present study support the claim that some grapheme-phoneme mismatches may require more attention (including but not limited to explicit instruction) than others. The present study proposed a hierarchy of importance for the grapheme-phoneme mismatches based on statistical likelihood of improvement from treatments such as those utilized in the present study (i.e., no formal pronunciation instruction, but consistent pronunciation practice over the course of a semester). This hierarchy orders the mismatches, beginning with graphemes for which there was a decrease in the odds of a target response, such as <ll>, and ending with the graphemes that did show (significant) increase in the likelihood of a target response. Importantly, the grouping of mismatches designated as being of higher importance based off of the analysis of graphemes individually was subsumed by the grouping created from the analysis based on contrastive analysis (i.e., L1 phoneme options per grapheme, L1 phonological inventory presence). The fact that the contrastive analyses produced a grouping of mismatches designated as higher importance in the hierarchy that subsumed the grouping created by individual analysis demonstrates that contrastive analysis can be utilized to target specific areas that are likely to be problematic to L2 learners.

The hierarchy proposed by the present study has strong implications for pronunciation instruction in the classroom, but it can also function as a support system for instructors. Previous research has shown that instructors are often reluctant to teach pronunciation for a myriad of reasons, including feeling unprepared and/or fearing providing inaccurate information (Derwing, 2008). The hierarchy can assist instructors in a number of ways. First, approaching pronunciation from the perspective of contrastive analysis provides instructors and students with common ground (i.e., how graphemes are pronounced in the L1) as a starting point. Second, though grapheme-phoneme mismatches do not cover all possible pronunciation issues, it creates a clear set to work with, hopefully reducing or eliminating instructor discomfort (Derwing, 2008) pertaining to determining which aspects of pronunciation to focus on. Relatedly, the hierarchy itself can serve as a guide to instructors; as class time is at a premium, the hierarchy can assist instructors in determining which mismatches are the most crucial to explore during class, and which are more probable to improve with consistent pronunciation practice and/or training.

6.3.3 Technology-Assisted Language Learning

The present study also contributes to the field of Computer-Assisted Pronunciation Training (CAPT), first and foremost by demonstrating that accuracy pertaining to grapheme-phoneme mismatches, as well as decreasing the duration of VOT in the voiceless stops /p t k/, can be improved over time by utilizing technology for consistent pronunciation practice and/or training, and that this can be accomplished outside of the classroom. Additionally, the technological enhancements had a positive impact on the participants' experiences, particularly for the participants in the two experimental groups (i.e., those who practiced pronunciation utilizing either speech recognition or listen-and-repeat activities with recordings from native speakers). Fitting with the findings of Steel (2012), availability and exercises pertaining to grammar and vocabulary featured prominently in the responses from both groups. Moreover, the analysis of the responses to the Technology Questionnaire in the present study revealed very positive impressions of the technology overall from both experimental groups, and specifically regarding the opportunities to receive feedback on their pronunciation, as was found by Golonka et al. (2014). Participants from EG1 also felt very strongly that interacting with the app improved their performance in their first-semester Spanish courses; EG2 had somewhat more mixed feelings, but many participants believed they benefited from practicing with the listen-and-repeat style activities.

Although many of the reactions to the technology were positive, the present study also showed similarities to the cautions and potential pitfalls mentioned in previous research. A plausible explanation for the smaller amount of improvement in EG1 is that much of their focus was distracted from their primary task (i.e., interacting with the app and practicing pronunciation) due to issues with the app itself. As was found when analyzing the participant groups' language pertaining to technology, the three groups used language pertaining to technology differently from each other. Specifically, EG1 demonstrated a highly statistically significant difference from both EG2 and CG; although EG2 also used technology-related words significantly differently than CG, the effect was not as strong. Additionally, within the group of technology-related words, the words pertaining to problems with technology were used nearly unilaterally by EG1. Considering the rubric for evaluating language learning mobile applications created by Martín Monje et al. (2014), before even considering any of the special features of the app, it would already be graded poorly in the *technical criteria* area of the rubric based on the glitches within the basic functionality of the app that participants reported would cause the app to crash. Additionally, despite EG1's positive attitude toward the speech recognition functionality of the app, there was also a trend within EG1 pertaining to negative impressions of feedback from the app, citing rationales such as being marked wrong when pronouncing things correctly, or that the app would not recognize the participant's voice. This line of commentary from EG1, combined with their very positive reactions overall toward the speech recognition functionality of the app, suggests that despite its appeal, more improvements need to be made to ASR systems, potentially even requiring specific linguistically-influenced additions to the systems that take language learners and the language learning process into account (Kolesnikova, 2017).

Previous CALL and MALL research has concluded that technology adapted to language learning purposes should be grounded in theory, methodology, and linguistics (Martín Monje et al., 2014; Youngs et al., 2011), and rubrics created to measure language ability have been used to evaluate linguistic content of apps (Martín Monje et al., 2014). Given that language ability rubrics have already been used in this manner, it would not be surprising if, as technology continues to advance, language learning standards were applied to apps (and other technology) as well. Current pedagogical methodology focuses on communication (e.g., Lee & VanPatten, 2003) and current standards emphasize global competence, encompassing ideas including communication, cultures, and comparison (The National Standards Collaborative Board, 2015). Elements of both communicative methodology and current standards could be found in the Spanish SOLO app. The application of methodology and standards used in foreign language classrooms onto apps used for language learning could move the implementation of technology into (or in support of) the classroom forward. For example, the Spanish SOLO app presented the typical modules (e.g., grammar, vocabulary) but each module also contained lessons on the history and/or culture of a Spanish-speaking country. Despite the focus on and explicit instruction of grammar and vocabulary, the content and other activities found in the app (e.g., the conversation builder, ASR) complement current pedagogical approaches and standards. While not all technology has to be implemented during class, aligning with instructional principles and goals could make these kinds of apps (as well as other technology) more attractive to instructors. The next generation of language learning apps could (and/or should) move in this direction.

6.3.4 Language Analysis of L2 Experiences using LIWC

Previous research has utilized LIWC to conduct sentiment analysis on various educational elements, including platforms (Andrews et al., 2009), MOOCs (Almatrafi et al., 2018; Geng et al., 2020) and other online and/or virtual learning environments (Hsiao et al., 2015), computerized tutors (D'Mello & Graesser, 2012), and textbooks (Sell & Farreras, 2017). Relatedly, research has also investigated the use of discussion boards (Almatrafi et al., 2018; Zhu et al., 2019), as well as attempting to use sentiment analysis to determine which messages are more urgent than others (Almatrafi et al., 2018). Fewer studies have focused on L2 learners, and much of those focus on the experiences of English Language Learners (ELLs) (Polat, 2014; Smith-Keiling et al., 2018; Zhang, 2013). Very few studies have investigated learners' experiences with foreign language activities (e.g., study abroad (Savicki & Price, 2015), virtual collaboration (Hsiao et al., 2015)) or instruction (Polat, 2013; Zapata & Ribota, 2021), and even fewer have focused on adult L2 opinions of instruction in Spanish foreign language classes in a university setting (Zapata & Ribota, 2021). The present study, which shares both participant population and L2 with the study conducted by Zapata and Ribota (2021), adds to the research utilizing LIWC to study educational settings, and particularly L2 settings, by not only probing opinions regarding educational technology tools, but also by attempting to connect the experiences of the participants with the technology to concrete evidence of performance. The findings of the present study indicate that LIWC can be used to find patterns in language use in questionnaire responses that, at first glance, appear to be thematically similar (i.e., in this case, giving overall positive impressions of all three technologies), though supplementing the LIWC analysis with one based on word frequency and pattern identification within word usage provided a clearer picture of the impact that the technology likely had on the participants. The language analysis revealed that the group whose performance improved the most (EG2) had more neutral views on their assigned technology, while the opinions from EG1 and CG were more extreme. Additionally, the present study was able to use the analysis from LIWC (as well as the word frequency analysis) to conclude that despite liking the technology they were assigned, the negative feelings expressed by EG1 pertaining to their technology (i.e., the app) and the frequency of words related to problems with the technology, likely served as enough of a distraction as to be detrimental to EG1's improvement in pronunciation, placing them closer to the improvement of CG, which did not interact with any pronunciation elements. The present study's findings demonstrate the possibilities that linking sentiment analysis to analyses of performance creates, particularly in terms of advances in technology that are attractive to L2 learners and language instructors; the combination of analyses provides unique insight into the effectiveness of language learning technology.

6.4 Limitations of the Present Study

6.4.1 Experiment Design

Some limitations of this study stemmed from the experiment design, including the participant population, regulations pertaining to informed consent, and the lack of feedback on the performance during the recordings. One limitation of this study pertains to the population of students enrolled in the first-semester Spanish classes at the time of the study, as well as the restrictions for offering extra credit. At the southeastern university where the study took place, there are placement guidelines which ought to assist students in enrolling in the correct course based on their previous study of Spanish. As was mentioned in Chapter 3, placement into a first-semester Spanish class should only include students who have studied Spanish for a maximum of one year within the past four years, or who have never studied the language at all (Department of Modern Languages and Classics, n.d.). However, despite these efforts, students cannot be forced to change classes even if they report that they have had more exposure to Spanish but have signed up for the first-semester course. Ideally, replications of this study would be able to filter out participants who are not true beginners; in the present study, this was not possible because at this particular southeastern university, if extra credit is offered in first or second-year Spanish courses, it is mandatory to offer the opportunity level-wide.

Related to this issue are two others that were caused by a) the participants being familiar with all three instruments, or b) some of the requirements of the study relying on the honor system. The former was likely caused by the requirement that all participants know about all the elements of the study. Rather than presenting one technology to each course section, all the sections were informed of all three options, and then were placed in a particular group once they enrolled in the study. This prompted one known participant from CG to admit, in the technology survey at the end of the study, that they had downloaded and had been using the Spanish SOLO app for the duration of the study. This participant's data were removed from analysis, which caused the already small number of participants in CG who completed the entire study to decrease even further.

Having knowledge of all three instrument types leads into situation b): in most cases, it was difficult to track participation in general, but it was also necessary to trust that the participants wanted the extra credit enough that they would follow the rules. The Blackboard

LMS performed the best in terms of tracking participation; the metadata for each activity tracked which students completed which activities, and how long they spent on each activity. In contrast, the app and the textbook-accompanying activities monitored time spent on activities to some extent, neither of these were under the investigator's control,. Moreover, the amount of time that participants in EG1 spent using the speech recognition function was difficult to quantify. Being able to track and calculate the amount of time EG1 participants spent using the app and its speech recognition feature (as well as collecting these statistics for the other two groups) would allow for more precise measurements of the technology's effectiveness. Additionally, it would also allow for analysis through a different lens, adding the variable 'time spent practicing' to the models for analyses used in the present study (i.e., logistic and linear regression). As the impact of group membership on pronunciation was small but often significant, differences in time spent practicing could account for more of the data.

Another limitation, and possible avenue for future study, stemmed from comments found in the answers to the Technology Questionnaire. Multiple participants expressed their desire for feedback on their performance during the four recording sessions, and their disappointment that this did not occur. While this would have worked against the aims of the present study, it indicates another avenue to explore pertaining to feedback and pronunciation. Similarly, some participants also expressed that it would have been interesting to see if their performance changed based on the stimuli they read aloud during the recording sessions, indicating that they would have liked to have read different sentences each time, rather than the same ones. The principal investigator consciously chose to keep the stimuli identical throughout all the recordings in order to control for differences based on surrounding words or sounds, which would have been difficult to control for were the stimuli different each time. Keeping everything the same also counteracted potential differences caused by fatigue.

6.4.2 Technological Issues

In addition to the limitations resulting from the experiment design, there were others pertaining to technological issues. The primary issue stemmed from the fact that it was impossible to modify or manipulate the two experimental technologies because they are not proprietary to the principal investigator. In the case of the Spanish SOLO app, the participants reported experiencing glitches such as the app inexplicably crashing in the middle of an activity, which resulted in the principal investigator requesting that all the participants in this group skip that activity going forward. This removed one of the activities in which the participants could use the speech recognition feature; it was also a frequent comment in the Technology Questionnaire, which may have contributed to the interpretation of negative emotion for that group. In the case of the other experimental technology, the listen-then-repeat style activities that accompany the course textbook, the principal investigator had no control over which sounds were focused on, nor how many activities were available. This resulted in some of the activities not focusing specifically on the sounds investigated in the present study; however, they had to be assigned because the participants were required to complete a minimum of two hours per week of interaction with their assigned technology.

Regarding the investigation into and analysis of participants' attitudes toward the technologies included in this study, the analysis was also limited by the disappearance of the Spanish SOLO app. In approximately Summer of 2019, the app was removed from the Apple App Store, and the website was also taken down. There had been a plan in place to conduct a second experiment, in which students who were not enrolled in any foreign language course, nor had been for a few semesters, would interact with the two experimental technologies (the app and the textbook-provided activities) in order to attempt to tease apart attitudes toward and evaluations of these two technologies. As this had been planned, the technology survey given to the participants in the first (semester long) experiment did not contain as targeted of a list of questions, nor did it ask the participants to continue writing until a timer expired. Thus, when the app disappeared and the second experiment could not be conducted, the survey results that had to be used for analysis contained much shorter responses than what had originally been intended. Typically, studies conducted using Linguistic Inquiry and Word Count (LIWC) analyze large data sets and therefore have more normally distributed data. While negative emotion, and specifically words designated as *sad*, were used significantly differently between the three experiment groups, it is possible that the survey data from the experiment specifically designed for the purpose of investigating attitude would have demonstrated more interesting and impactful results.

6.5 Considerations for Future Study

6.5.1 Analysis Based on English/Spanish Cognates

Some of the words in the stimuli in the present study would be considered cognates when comparing English to Spanish. These cognates may have presented more of a challenge to the participants coming from first-semester Spanish foreign language classes due to the combination of the participants' beginning knowledge of Spanish and the similarity to words in the participants' L1 (English). Some cognates included in the stimuli were identical in both languages (e.g., *hotel* 'hotel'; *utopía* 'utopia'; *capital* 'capital'; *Argentina* 'Argentina'), while others were similar enough to trigger recognition (e.g., *oficial* 'official'; *techo* 'roof/ceiling'; *sociología* 'sociology'; *gimnasio* 'gymnasium'; *agencia* 'agency'; *situado* 'situated'; *Egipto*

‘Egypt’). It would be interesting to separate the cognates from the rest of the data set and run the analyses again in order to determine a) how well the participants performed when presented with cognates, b) whether the inclusion of cognates improved or hindered overall performance, and c) how well the pronunciation of grapheme-phoneme mismatches in cognates specifically improve over time.

6.5.2 Experiment Design

One issue that should be considered prior to any future repetition of this study (or ones similar to it) pertains to the recruitment of participants into the three instrument groups. During the recruitment process for the present study, participants were not assigned to a specific instrument group (i.e., technology) until they arrived for the first (Baseline) recording. Then, as was described in Section 3.7.2, their course section was assigned to one of the instruments. During the classroom recruitment visits wherein the study procedures were summarized, every (potential) participant was informed regarding the use of all three instruments in the study itself. However, once a given course section was assigned to a specific instrument, that instrument was the only one that those participants were supposed to interact with. This did not seem to be much of an issue for the two experimental groups, as they interacted with the two instruments that asked them to practice their pronunciation. However, those participants assigned to CG, who performed grammar-related activities with no pronunciation element, had been informed about a study that investigated pronunciation, and yet they were not asked to do anything related to pronunciation. This knowledge, combined with the fact that they knew about the instruments utilized by other participants, either because they had heard about it in the recruitment pitch or because they had friends in other sections who had been assigned to different instruments, led one participant from CG to be excluded from analysis due to that participant’s admission that they had additionally downloaded the Spanish SOLO app and interacted with it throughout the semester. No one else admitted to doing the same; however, it is not outside the realm of possibility that participants in CG explored one (or both) of the experimental instruments without admitting it in the Technology Questionnaire.

The fact that one participant did admit to using one of the experimental instruments despite being in CG brings up two important questions. The first question pertains to whether every (potential) participant should be aware of every aspect of the study. For many students, interactive pronunciation activities are more interesting than cloze-style grammar activities; thus, being informed about three potential instruments but then being assigned to the one with no pronunciation element could have resulted in disappointment. While some participants did express happiness at being able to practice grammar without it negatively impacting their homework grade, the present study did find significant differences in negative emotions in the attitudes toward the control instrument versus the experimental ones (see Chapter 5).

6.5.3 Corpus of Participants' Evaluations of Technology

As was explained in Chapter 5, the quantitative analysis of the participants' responses to the Technology Questionnaire had to implement a non-parametric statistical test in order to account for the short lengths of the responses to the open-ended questions. Many studies conducted using Linguistic Inquiry and Word Count (LIWC) analyze large corpora (pre-made or mined and compiled by the investigators), giving the analysis better statistical power due to the large number of words available for analysis. While the trends explored in Chapter 5 are interesting, were those trends to also be found in a larger dataset, it would further legitimize the findings.

One way to attempt to gather longer participant responses would be to construct an experiment in which participants act much like a focus group testing out a new product. In this experiment, participants would meet with the investigator only once, for approximately an hour; during this time, they would be assigned to one of the two experimental technologies and asked to interact with it for approximately 30 minutes, after which they would be prompted to begin a survey. This survey, comprised mainly of open-ended questions (plus a few background questions), would ask participants to write responses to prompts (i.e., questions). The survey, constructed in Qualtrics (for example), would make use of a setting which, when enabled, automatically moves from one question to the next in the survey after a specified amount of time. Meeting only once would potentially improve recruitment, as it would be less onerous to participate, and having a timer for each open-ended question would make the task of free-writing seem more finite.

6.6 Recommendations for Future Research

6.6.1 Collaboration between Fields

One of the limitations of the present study is the lack of control over the instruments used in the experiment, which could be remedied through collaboration with another field, as it would allow the investigator more freedom and flexibility as well as control over the entire instrument. In the case of the app (used by EG1), participants reported glitches with some of the activities, so much so that the term 'glitch' was used frequently in their survey responses. These glitches were clearly a distraction, and may have contributed to the smaller amount of improvement found in EG1. Collaboration between distinct fields would make it possible to gain more control over the instruments; in this case, it would be most beneficial to collaborate with computer scientists. In the CAPT field, there have been some instances where the researchers have been able to design and create their own programs or applications (Meng et al., 2010; Wong et al., 2010), including ones that utilize ASR (Lumsden et al., 2010). Additionally, having control of the instruments would allow for more accurate tracking of participation hours. In the present study, the hours the participants in the experimental groups spent practicing each week was self-reported, and the app did not track them itself. Better tracking would allow for analysis based on time spent practicing, for example, which would help to pinpoint the optimal amount of practice for improving accuracy of grapheme-phoneme mismatches and duration of VOT.

6.6.2 Increased Focus on Feedback

Another frequent comment made by participants was that they would have appreciated (more) feedback on their pronunciation. Some of the comments came from the participants in EG2, the group that used the online listen-and-repeat style activities that accompanied their textbook. This was somewhat unsurprising since their activities did not provide feedback and put the onus onto them to compare their speech to that of the native speaker in the recording. However, some participants expressed a desire for specific feedback on their pronunciation while completing the practice activities, and a few others requested feedback on their pronunciation of the stimuli during the recording sessions. Much of the previous research regarding explicit instruction of pronunciation trends toward the use of phonetics (Aliaga-García, 2007, 2017; Alves & Luchini, 2017, 2020; Camus 2020; González-Bueno, 1997a; Yang, 2017). Studies have also shown improvement in pronunciation through visual feedback using elements of acoustic phonetics, specifically through the use of spectrogram comparisons (García et al., 2018; Morgan & García, 2016; Offerman & Olson, 2016). While these methods have resulted in improvement, it may not always be possible to instruct students in phonetics, especially in courses not designed with that specific aim in mind, such as beginning level foreign language courses. However, the small amount of feedback provided to EG1 (binary correct-incorrect accompanied by an image and a sound) and the lack of feedback found in the activities completed by EG2 is not enough. Future studies should attempt to find a middle ground between binary non-specific feedback/no feedback and feedback that would require at least some training in phonetics.

Returning to the comment regarding feedback based on the periodic recordings of the participants reading the target sentences aloud, this is another area that should be explored further. A comment found in the questionnaire in the present study suggested meetings between the participants and the investigator to discuss the participants' performance in the recordings that were scheduled periodically throughout the semester. It would require a shift in experiment design, however, as it could introduce bias into the study since the investigator would have to score the recordings before the experiment ended. The manner of providing feedback would also have to be closely controlled; otherwise, it could morph into instruction. It would also be difficult to implement level-wide due to the number of (potential) participants; additionally, providing feedback to all participants within the same time period could be problematic. It could potentially be explored, however, in more of a case-study type of setting such as one or two classrooms of students.

6.6.3 Analysis of L2 Acquisition of Pronunciation as Skill Mastery

The present study analyzed improvement, particularly with grapheme-phoneme mismatches, by treating each recording as if it were a test, producing an accuracy score, both overall (i.e., considering all study graphemes) and for each grapheme individually. This analysis, however, could not fully capture the relationship between skill sets. Some of this analysis was covered in the present study by the place/manner of articulation logistic regression analysis, but odds of a target response does not describe how the acquisition of the sounds might be affected by each other. Cognitive Diagnostic Modeling (CDM), on the other hand, is designed "to provide examinees with information concerning whether or not they have mastered each of a group of specific, discretely defined skills, or attributes" (Huebner, 2010, p. 1). CDM fits well as a plausible next step in the analysis of pronunciation elements such as grapheme-phoneme mismatches because it can also be used to provide instructors with information that would help them make better decisions about what to focus on in the classroom (Ravand & Robitzsch, 2015, p. 1). Moreover, providing feedback to learners regarding specific, clearly defined areas in which they are struggling creates an opportunity for "individualized remediation" (Huebner, 2010, p. 1). This feedback would be specific to the performance of each individual learner and provide specific areas or skills that required more practice/training based on the criteria for mastery used in the diagnostic model. Applying CDM to pronunciation mastery in this way responds nicely to the request for more (and/or better) feedback expressed by participants in the present study.

6.7 Final Remarks

With continual advances in technological development, it is probable that more effort will be made to incorporate new technology into L2 acquisition and the foreign language classroom. The present study simultaneously serves as evidence that consistent pronunciation practice and/or training can positively impact both overt and subtle pronunciation differences between English and Spanish, but also that improvement is highly contingent upon technological stability, the lack of which seemed to have had a negative impact on pronunciation improvement for one of the participant groups. Therefore, it is expected that future implementation of technology for use in L2 acquisition and/or the foreign language classroom may require the adaptation (rather than adoption) of mainstream technology to fit purposes specific to language learning in order for technological advances to be utilized to the fullest. At the same time, the opinions of those that may use the technology ought to be taken into account in determining technological utility. Sentiment and language analyses of L2 impressions of language learning technology (and other aspects of the language learning process) can reveal aspects or features that are deemed essential by learners, but more importantly, connections can be made between impressions of technology and pronunciation improvement.

REFERENCES

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1,* 1–8. https://doi.org/10.1016/S0095-4470(19)31372-5

Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63,* 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

ACTFL. (2012). *Proficiency guidelines – Spanish: Comunicación oral.* Retrieved October 16, 2018 from https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/spanish/comunicación-oral

ACTFL. (2015). *World-readiness standards for learning languages.* Retrieved October 16, 2018 from www.actfl.org/publications/all/national-standards-foreign-language-education

Adda-Decker, M., & Lamel, L. (1999). Pronunciation variants across system configuration, language and speaking style. *Speech Communication, 29*(2), 83–98. https://doi.org/10.1016/S0167-6393(99)00032-1

Agostinelli, C. (2012). Native speaker perceptions of spoken L2 Spanish: The role of pronunciation and implications for pedagogy. In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 147–153). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Aitchison, J. (2003). *Words in the mind* (3rd edition). Blackwell Publishing.

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. *Phonetics Teaching & Learning Conference (PTLC)*. Retrieved from http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_32e.pdf

Aliaga-Garcia, C. (2017). *The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English* (121323) [Doctoral dissertation, Universitat de Barcelona]. Dipòsit Digital de la Universitat de Barcelona. http://hdl.handle.net/2445/121323

Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education, 118,* 1–9. https://doi.org/10.1016/j.compedu.2017.11.002

Alves, U. K., & Luchini, P. L. (2017). Effects of perceptual training on the identification and production of word-initial voiceless stops by Argentinean learners of English. *Ilha do Desterro, 70*(3), 15−32. https://doi.org/10.5007/2175-8026.2017v70n3p15

Alves, U. K., & Luchini, P. L. (2020). ¿Entrenamiento perceptivo o instrucción explícita? Percepción y producción de los patrones de *voice onset time* iniciales del inglés (LE) por estudiantes brasileños. *Forma y Función*, *33*(2), 135−165. https://www.doi.org/10.15446/fyf. v33n2.81067

Alves, U. K., & Magro, V. (2011). Raising awareness of L2 phonology: Explicit instruction and the acquisition of aspirated /p/ by Brazilian Portuguese speakers. *Letras de Hoje. Porto Alegre, 46*(3), 71−80. http://hdl.handle.net/10183/169806

Alvord, S. M., & Rogers, B. M. (2014). Miami-Cuban Spanish vowels in contact. *Sociolinguistic Studies*, *8*(1), 139−170. https://doi.org/10.1558/sols.v8i1.139

American Psychological Association. (2020). Winsorizing. In *APA Dictionary of Psychology online*. Retrieved 1 April 2021 from https://dictionary.apa.org/winsorizing

Andrews, J. J., Bond, G. D., & Speller, L. F. (2009). When course management systems fail: Student and instructor 'on-the-fly' adaptation behaviors. *International Journal of Cognitive Technology*, *14*(1), 45−54.

Anthony, L. E. (1996). The role of phoneme based approaches in the teaching of pronunciation. 岡山理科大学紀要. B, 人文・社会科学 *(The Bulletin of the Okayama University of Science. B, Human Sciences), 32*, 115−122. http://id.nii.ac.jp/1182/00001521/

Anthony, L. (2018). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570. https://doi.org/10.1016/j.wocn.2011.03.001

Arnold, N., & Ducate, L. (Eds.). (2011). *Present and future promises of CALL: From theory and research to new directions in language teaching* (2nd edition). CALICO.

Arroju, M., Hassan, A., & Farnadi, G. (2015). Age, gender and personality recognition using tweets in a multilingual setting. In L. Cappellato, N. Ferro, G. J. F. Jones, & E. San Juan (Eds.), *6th Conference and Labs of the Evaluation Forum (CLEF 2015): Experimental IR meets multilinguality, multimodality, and interaction* (pp. 23-31). Ghent: Universiteit Gent.

Arteaga, D. (2000). Articulatory phonetics in the first-year Spanish classroom. *Modern Language Journal, 84*(3), 339−354. https://doi.org/10.1111/0026-7902.00073

Aspillera, P. S., & Hernandez, Y. C. (2014). *Basic Tagalog for foreigners and non-Tagalogs: (mp3 downloadable audio included)*. Tuttle Publishing.

Audacity Team. (2008). Audacity (Version 2.3.0) [Computer program]. Retrieved September 30, 2018 from http://audacityteam.org/

Austen, J. (1853). *Pride and Prejudice*. RD Bentley.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317–342. https://doi.org/10.1177/0023830908099068

Bankí, J. R. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4; Proceedings of the 4th international symposium on bilingualism* (pp. 240–249). Cascadilla Press. www.cascadilla.com/isb4.html

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). Multilingual Matters. https://doi.org/10.21832/9781847691118-013

Baugh, A. C., & Cable, T. (2012). *A history of the English language* (6th edition). Routledge.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. J. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins Publishing.

Blanco, J.A. (2016). *Portales: Spanish online.* Vista Higher Learning.

Blanco, J. A., & Horwitz, E. K. (2001). *Vistas: Introducción a la lengua española*. Vista Higher Learning.

Boersma, P. & Weenink, D. (2018). *Praat: Doing phonetics by computer* (6.0.43). Retrieved September 8, 2018 from http://www.praat.org/

Bond, M., & Pennebaker, J. W. (2012). Automated computer-based feedback in expressive writing. *Computers in Human Behavior*, *28*(3), 1014–1018. https://doi.org/10.1016/j.chb.2012.01.003

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy.* Cambridge University Press.

Camus, P. (2020). The effects of explicit pronunciation instruction on the production of second language Spanish voiceless stops: A classroom study. *Instructed Second Language Acquisition, 3*(1), 81–103. https://doi.org/10.1558/isla.37279

Carreiras, M., Duñabeitia, J., & Molinaro, N. (2014). Reading words and sentences in Spanish. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 803–826). John Wiley and Sons. http://doi.org/10.1002/9781118228098.ch38

Casas Solà, A. (2014). *An acoustic study of word-initial stops by Catalan/Spanish speakers* (2500245) [Bachelor's thesis, Universitat Autònoma de Barcelona]. Dipòsit digital de documents de la UAB.

Cavus, N. (2016). Development of an intelligent (sic) mobile application for teaching English pronunciation. *Procedia Computer Science*, *102*(August), 365–369. https://doi.org/10.1016/j.procs.2016.09.413

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first language speech production. *Journal of Phonetics*, *40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Chiranjeevi, P., Santosh, D. T., & Vardhan, B. V. (2018). Finding and classifying the deceptive spam reviews using LIWC dictionary variables and decision tree classifier. *International Journal of Engineering and Technology*, *10*, 308–315. https://doi.org/10.21817/ijet/2018/v10i1/181001131

Cho, T. & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207–229. https://doi.org/10.1006/jpho.1999.0094

Chuang, Y-T. (2015). SSCLS: A smartphone-supported collaborative learning system. *Telematics and Informatics, 32,* 463–474. http://dx.doi.org/10.1016/j.tele.2014.10.004

Cichoki, W., House, A. B., Kinlock, A. M., & Lister, A. C. (1999). Cantonese speakers and the acquisition of French consonants. In J. Leather (Ed.), *Phonological issues in language learning* (pp. 95–121). Blackwell Publishers.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676. https://doi.org/10.1121/1.2000774

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, *39*(2), 237–245. https://doi.org/10.1016/j.wocn.2011.02.006

Cobb, K., & Simonet, M. (2015). Adult second language learning of Spanish vowels. *Hispania*, *98*(1), 47–60. https://doi.org/10.1353/hpn.2015.0026

Colantoni, L., & Steele, J. (2006). Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters. In *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 59–73). Cascadilla. Retrieved from http://www.lingref.com/cpp/casp/7/paper1275.pdf

Cook, V. (2001). *Second language learning and language teaching* (3rd edition). Oxford University Press.

Counselman, D. (2015). Directing attention to pronunciation in the second language classroom. *Hispania, 98*(1), 31–46. https://www.jstor.org/stable/24368850

Crawley, M. J. (2013). *The R book* (2nd edition). John Wiley & Sons, Ltd.

Culbertson, G. R., Andersen, E. L., White, W. M., Zhang, D., & Jung, M. F. (2016). Crystallize: An immersive, collaborative game for second language learning. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16* (pp. 635–646.) https://doi.org/10.1145/2818048.2820020

Cutshall, S. (2012) More than a decade of standards: Integrating "comparisons" in your language instruction. *The Language Educator (Oct. 2012)*, 32–37. Retrieved from https://www.actfl.org/sites/default/files/publications/standards/Communities.pdf

D'Mello, S. K., & Graesser, A. (2012). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, *5*(4), 304–317. https://doi.org/10.1109/TLT.2012.10

Daniels, P. T., & Bright, W. (1996). *The world's writing systems*. Oxford University Press.

Darcy, I.; Ewert, D; & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 93–108). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

de Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. Routledge.

de la Fuente Iglesias, M., & Castillejo, S. P. (2020). Spanish mid vowels as sociolinguistic variables in Galicia. *Spanish in Context*, *17*(3), 464–487. https://doi.org/10.1075/sic.18027.igl

del Pilar Salas-Zarate, M., Paredes-Valverde, M. A., Limon-Romero, J., Tlapa, D. A., & Báez-López, Y. A. (2016). Sentiment classification of Spanish reviews: An approach based on feature selection and machine learning methods. *Journal of Universal Computer Science*, *22*(5), 691–708.

Delahunty, G. P., & Garvey, J. J. (2010). *The English language: From sound to sense* (M. Palmquist, Ed.). The WAC Clearinghouse.

Demouy, V., & Kukulska-Hulme, A. (2010). On the spot: Using mobile devices for listening and speaking practice on a French language programme. *Open Learning*, *25*(3), 217–232. https://doi.org/10.1080/02680513.2010.511955

Department of Modern Languages and Classics. (n.d.). *Spanish course placement*. The University of Alabama. Retrieved October 1, 2018 from https://mlc.ua.edu/spanish/undergraduate/course-placement/

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–370). John Benjamins.

Deterding, D. (2004). How many vowel sounds are there in English. *STETS Language & Communication Review*, *19*(10), 19–21.

Díaz-Campos, M. (2014). Segmental phonology in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 146–165). West Sussex, UK: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118584347

Distant Tribe LLC. (2016). *SOLO language.* Retrieved April 8, 2021 from https://web.archive.org/web/20180424224909/http://www.sololanguage.com/

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Donato, C., Bordage, N., & Rustin, P. (2012). French for Spanish Speakers through intercomprehension: A method of multiple-language acquisition for the Romance languages with implications for the future. *Adfl*, *42*(1), 49–60. https://doi.org/10.1632/adfl.42.1.49

Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL: The Journal of EUROCALL*, *27*(2), 197–216.

Eckman, F. R. (1987), Markedness and the contrastive analysis hypothesis. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*. Newbury House Publishers. https://doi.org/10.1017/S0272263100007671

Eddington, D. (2015). *Statistics for linguists: A step-by-step guide for novices*. Cambridge Scholars Publishing.

Eddington, D., & Elzinga, D. (2008). The phonetic context of American English flapping: Quantitative evidence. *Language and Speech*, *51*(3), 245–266. https://doi.org/10.1177/0023830908098542

Edwards, J. (1995). *Multilingualism*. New York, NY: Penguin Books.

Egnoto, M. J., Griffin, D. J., & Qiao, F. (2018). Grandstanding or foreshadowing: Analysing the University of Alabama active shooter threats with intergroup threat theory. *Dynamics of Asymmetric Conflict*, *11*(3), 171−185. https://doi.org/10.1080/17467586.2018.1432867

Eisenbach, M., & Eisenbach, A. (2020). *jsSyntaxTree*. IronCreek Software. http://ironcreek.net/syntaxtree/

Ellis, N. & Hooper, A. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics,* 22(4), 571−599. https://doi.org/10.1017/S0142716401004052

Estigarribia, B. (2017). A grammar sketch of Paraguayan Guaraní. In B. Estigarribia & J. Pinta (Eds.), *Guarani linguistics in the 21st century* (pp. 7−85). Brill Academic Publishers. https://doi.org/10.1163/9789004322578_003

Face, T. L. (2006). Intervocalic rhotic pronunciation by adult learners of Spanish as a second language. In *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 47−58). Cascadilla Proceedings Project.

Fitt, S. E. (1998). *Processing unfamiliar words: A study in the perception and production of native and foreign placenames* [Doctoral dissertation, University of Edinburgh]. Edinburgh Research Archive.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Mouton de Gruyter.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*(3), 708−721. https://doi.org/10.1121/1.391257

Flege, J. E., Munro, M., & Fox, R. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*(6), 3623–3641. https://doi.org/10.1121/1.409931

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Fosler-Lussier, E., & Morgan, N. (1998). Effects of speaking rate and word frequency on conversational pronunciations. *Modeling Pronunciation Variation for Automatic Speech Recognition*, 35–40. Retrieved March 28, 2021 from http://www.isca-speech.org/archive_open/mpv_98/mpv8_035.html

Foulkes, P., Docherty, G. J., & Jones, M. J. (2011). Analyzing stops. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 131–148). Routledge.

Fukada, T., & Sagisaka, Y. (1997). Automatic generation of a pronunciation dictionary based on a pronunciation network. In *fifth European conference on speech communication and technology*, 2471–2474. https://www.isca-speech.org/archive/archive_papers/eurospeech_1997/e97_2471.pdf

Fukada, T., Yoshimura, T., & Sagisaka, Y. (1999). Automatic generation of multiple pronunciations based on neural networks. *Speech Communication*, *27*(1), 63–73. https://doi.org/10.1016/S0167-6393(98)00066-1

Garcia, C., Kolat, M., & Morgan, T. A. (2018). Self-correction of second-language pronunciation via online, real-time, visual feedback. In *Proceedings of the 9th pronunciation in second language learning and teaching conference* (pp. 54–65). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2018/09/PSLLT-Proceedings-9_9-14-18_2.pdf

Garrido, M. (2007). Diphthongization of mid/low vowel sequences in Colombian Spanish. In *Selected proceedings of the third workshop on Spanish sociolinguistics* (pp. 30–37). Cascadilla Proceedings Project.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd edition). Routledge.

Gee, H. (1999, December 21). Say what you see. *Nature*. Retrieved from https://www.nature.com/news/1999/991223/full/news991223-6.html doi:10.1038/news991223-6

Geng, S., Niu, B., Feng, Y., & Huang, M. (2020). Understanding the focal points and sentiment of learners in MOOC reviews: A machine learning and SC-LIWC-based approach. *British Journal of Educational Technology*, *51*(5), 1785–1803. https://doi.org/10.1111/bjet.12999

Glass, C. R. (2017). Self-expression, social roles, and faculty members' attitudes towards online teaching. *Innovative Higher Education*, *42*(3), 239–252. https://doi.org/10.1007/s10755-016-9379-2

Gleason, J. (2012). Beaches and peaches: Common pronunciation errors among L1 Spanish speakers of English. In J. Levis & K. LeVelle (Eds.), In *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 205–215). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Glen, S. (5 December 2014). *Independent variable (treatment variable) definition and uses.* StatisticsHowTo.com: Elementary statistics for the rest of us! Retrieved April 1, 2021 from https://www.statisticshowto.com/independent-variable-definition/

Glen, S. (9 December 2014). *Criterion variable: Definition, use and examples.* StatisticsHowTo.com: Elementary statistics for the rest of us! Retrieved April 1, 2021 from https://www.statisticshowto.com/criterion-variable-2/

Glen, S. (15 February 2016). *Standardized beta coefficient: Definition & example.* StatisticsHowTo.com: Elementary statistics for the rest of us! Retrieved April 5, 2021 from https://www.statisticshowto.com/standardized-beta-coefficient/

Glen, S. (17 May 2016). *Winsorize: Definition, examples in easy steps.* StatisticsHowTo.com: Elementary statistics for the rest of us! Retrieved November 11, 2020 from https://statisticshowto.com/winsorize/

Glen. S. (9 November 2016). *Beta weight: Definition, uses.* StatisticsHowTo.com: Elementary statistics for the rest of us! Retrieved April 1, 2021 from https://www.statisticshowto.com/beta-weight/

Goikoetxea, E. (2006). Reading errors in first- and second-grade readers of a shallow orthography: Evidence from Spanish. *British Journal of Educational Psychology, 76,* 333–350. https://doi.org/10.1348/000709905X52490

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Gong, J., Cooke, M., & García Lecumberri, M. L. (2015). A quantitative model of first language influence in second language consonant learning. *Speech Communication*, *69*, 17–30. https://doi.org/10.1016/j.specom.2015.02.003

González-Bueno, M. (1997a). The effects of formal instruction on the acquisition of Spanish phonology. In W. R. Glass & A. T. Pérez-Leroux (Eds.), *Contemporary Perspectives on the Acquisition of Spanish* (pp. 57–75). Cascadilla Press.

González-Bueno, M. (1997b). Voice onset time in the perception of foreign accent by native listeners of Spanish. *International Review of Applied Linguistics in Language Teaching 35*(4), 251–262.

González López, V., & Counselman, D. (2013). L2 acquisition and category formation of Spanish voiceless stops by monolingual English novice learners. *Selected proceedings of the 16th Hispanic linguistics symposium*, (pp. 118–127). Cascadilla.

Gordon, J. (2012). Extra-linguistic factors in the teaching and learning of pronunciation in an ESL class. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 134–146). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation, 2*(1), 56–92. https://doi.org/10.1075/jslp.2.1.03gor

Goronzy, S., Rapp, S., & Kompe, R. (2004). Generating non-native pronunciation variants for lexicon adaptation. *Speech Communication*, *42*(1), 109–123. https://doi.org/10.1016/j.specom.2003.09.003

Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*(2), 159–176. https://doi.org/10.1016/S0167-6393(99)00050-3

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on production in Quichua-Spanish bilinguals. *Journal of Phonetics*, *28*(1), 27–42. https://doi.org/10.1006/jpho.2000.0104

Habash, N., Diab, M., & Rambow, O. (2012). Conventional orthography for dialectal Arabic. *Proceedings of the eighth international conference on language resources and evaluation (LREC-2012)*, (January 2012), 711–718. Retrieved March 24, 2021 from http://www.lrec-conf.org/proceedings/lrec2012/pdf/579_Paper.pdf

Hanzawa, K. (2018). The development of voice onset time (VOT) in a content-based instruction university program by Japanese learners of English: A longitudinal study. *Canadian Modern Language Review*, *74*(4), 502–522. https://doi.org/10.3138/cmlr.2018-0196

Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *IRAL, International Review of Applied Linguistics in Language Teaching*, *45*(4), 353–378. https://doi.org/10.1515/IRAL.2007.015

Heathman, A. (22 Nov. 2017). *The smartphone turns 25: here are the five major milestones of the device.* Verdict Media Limited (verdict.co.uk). Retrieved May 27, 2021 from https://www.verdict.co.uk/smartphone-invented-25-years/

Hempel, J. (2016, Dec. 19). *Voice is the next big platform and Alexa will own it.* Wired. Retrieved from https://www.wired.com/2016/12/voice-is-the-next-big-platform-and-alexa-will-own-it/

Henriksen, N. (2013). Suprasegmental phenomena in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 166−182). John Wiley & Sons, Ltd.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, *119*(5), 3072–3084. https://doi.org/10.1121/1.2184289

Hosch, W. L. (2020). Smartphone. In R. Curley, R. Lewis, G. Lotha, & S. Sinha (Eds.), *Encyclopedia Britannica*. Retrieved May 27, 2021 from https://www.britannica.com/technology/smartphone

Hsiao, I. Y. T., Yang, S. J. H., & Chia-Jui, C. (2015). The effects of collaborative models in Second Life on French learning. *Educational Technology Research and Development*, 63, 645−670. https://doi.org/10.1007/s11423-015-9379-4

Hualde, J. I. (2003). Segmental phonology. In J. I. Hualde & J. Ortiz de Urbana (Eds.), *A grammar of Basque* (pp. 15−110). Mouton de Gruyter.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

Hualde, J. I. (2014). *Los sonidos del español: Spanish Language edition*. Cambridge University Press.

Hualde, J. I., Olarrea, A., Escobar, A. M., & Travis, C. E. (2010). *Introducción a la lingüística hispánica* (2nd ed.). Cambridge University Press.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*(3), 1−7. https://doi.org/10.7275/7fdd-6897

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105–120. https://doi.org/10.1016/j.wocn.2015.06.007

Hughes, A. (1983). Second learning and communicative language teaching. In K. Johnson & D. Porter (Eds.), *Perspectives in communicative language teaching* (pp. 1−19). Academic Press.

Hurtado, L. M., & Estrada, C. (2010). Factors influencing the second language acquisition of Spanish vibrants. *Modern Language Journal*, *94*(1), 74–86. https://doi.org/10.1111/j.1540-4781.2009.00984.x

Ijalba. E, & Obler, L. K. (2015). First language grapheme-phoneme transparency effects in adult second-language learning. *Reading in a Foreign Language, 27*(1), 47–70.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

iOS Feature Availability. (n.d.). *Siri*. Apple.com. Retrieved September 22, 2018 from https://www.apple.com/ios/feature availability/#siri

Isbister, T., Kaati, L., & Cohen, K. (2017). Gender classification with data independent features in multiple languages. *European Intelligence and Security Informatics Conference, EISIC 2017, Athens, Greece, September 11-13, 2017*, 54–60. https://doi.org/10.1109/EISIC.2017.16

Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America, 125*(4), 2348–2360. https://doi.org/10.1121/1.3082103

Jaidka, K., Chhaya, N., Mumick, S., Killingsworth, M., Halevy, A., & Ungar, L. (2020). Beyond positive emotion: Deconstructing happy moments based on writing prompts. *Proceedings of the International AAAI Conference on Web and Social Media, 14*(1), 294–302. Retrieved March 28, 2021 from https://ojs.aaai.org/index.php/ICWSM/article/view/7300

Jensen, K. J., & Riis, S. (2000). Self-organizing letter code-book for text-to-phoneme neural network model. *Sixth International Conference on Spoken Language Processing*, 3, 318–321.

Johnson, B. (21 Oct. 2008). Apple profits rocket on iPhone boom*. Guardian News & Media Limited (theguardian.com)*. Retrieved May 27, 2021 from https://www.theguardian.com/technology/2008/oct/21/apple-apple

Jordan, K. N., & Pennebaker, J. W. (2016, August 1). Accepting the nomination: A comparison of the speeches of Trump and Clinton. *Wordwatchers*. Retrieved March 28, 2021 from https://wordwatchers.wordpress.com/2016/08/01/accepting-the-nomination-a-comparison-of-the-speeches-of-trump-and-clinton/

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. [Doctoral dissertation, MIT]. Published 1980, New York: Garland.

Kampstra, P. (2008). Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software*, *28*(1), 1–9.

Kanters, S., Cucchiarini, C., Strik, H. (2009). The goodness of pronunciation algorithm: A detailed performance study. *SLaTE-2009*, 49–52. http://hdl.handle.net/2066/76491

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*(Nov), 1−17. https://doi.org/10.3389/fpsyg.2014.01246

Keating, P. (1984). Phonetic and phonological representations of stop consonant voicing. *Language*, *60*, 286–319. https://doi.org/10.2307/413642

Keating, P. A., Mikoś, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America, 70*(5), 1261−1271. https://doi.org/10.1121/1.387139

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory*. Academic Press.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *Modern Language Journal*, *97*(3), 720–744. https://doi.org/10.1111/j.1540-4781.2013.12029.x

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research, 19*(3), 254–275. https://doi.org/10.1177/1362168814541735

Kolesnikova, O. (2017). Comparative analysis of American English and Mexican Spanish consonants for Computer Assisted Pronunciation Training. *Revista Signos*, *50*(94), 195–216. https://doi.org/10.4067/S0718-09342017000200195

Konopka, K., & Pierrehumbert, J. (2008). Vowels in contact: Mexican heritage English in Chicago. In *Proceedings from The Sixteenth Annual Symposium about Language and Society-Austin. Texas Linguistic Forum, 52* (pp. 94−103). University of Texas. Citeseer. http://salsa.ling.utexas.edu/proceedings/2008/index.html

Krashen, S. D. (1988). *Second language acquisition and second language learning*. Prentice Hall International (UK) Ltd.

Kupske, F. F., & Oliveira, M. S. de. (2020). O desenvolvimento do padrão de voice onset time das oclusivas surdas iniciais do inglês por aprendizes soteropolitanos: Efeitos da instrução explícita. *Ilha do Desterro, 73*(3), 185−204. https://doi.org/10.5007/2175-8026.2020v73n3p185

Kyngäs, H., & Rissanen, M. (2001). Support as a crucial predictor of good compliance of adolescents with a chronic disease. *Journal of Clinical Nursing*, *10*(6), 767–774. https://doi.org/10.1046/j.1365-2702.2001.00538.x

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich.

Leather, J. (1999). Second-language speech research: An introduction. In J. Leather (Ed.)., *Phonological issues in language learning* (pp. 1−56). Blackwell.

Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen* (2nd edition). McGraw-Hill.

Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses, The. (2015). The Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology (Bernard Comrie, Martin Haspelmath) & The Department of Linguistics of the University of Leipzig (Balthasar Bickel) (Eds.). University of Leipzig. Retrieved May 28, 2021 from http://www.eva.mpg.de/lingua/resources/glossing-rules.php

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, *27*, 184–202. https://doi.org/10.1017/S0267190508070098

Levis, J. & Barriuso, T. A. (2012). Nonnative speakers' pronunciation errors in spoken and read English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 187–194). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th edition). Oxford University Press.

Lillis, D. (2020). *Generalized linear models (GLMs) in R, part 4: Options, link functions, and interpretation.* The Analysis Factor, LLC. Retrieved December 19, 2020 from https://www.theanalysisfactor.com/generalized-linear-models-glm-r-part4/

Linden Research, Inc. (2018). *Using voice chat.* Secondlife.com. Retrieved September 22, 2018 from https://community.secondlife.com/knowledgebase/english/using-voice-chat-r77/

Lipski, J. (1994). *Latin American Spanish*. Addison Wesley Longman.

Lipski, J. M. (2011). Socio-phonological variation in Latin American Spanish. In M. Díaz-Campos (Ed.), *The handbook of Hispanic sociolinguistics* (pp. 72–97). https://doi.org/10.1002/9781444393446.ch4

Lisker, L, & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, Y., Holden, D., & Zheng, D. (2016). Analyzing students' language learning experience in an augmented reality mobile game: An exploration of an emergent learning environment. *Procedia – Social and Behavioral Sciences, 228,* 369–374. https://doi.org/10.1016/j.sbspro.2016.07.055

Liu, M., Xu, B., Huang, T., Deng, Y., & Li, C. (2000). Mandarin accent adaptation based on context-independent/context-dependent pronunciation modeling. *ICASSP, IEEE international conference on acoustics, speech and signal processing - Proceedings*, *2*, 1025–1028. https://doi.org/10.1109/ICASSP.2000.859137

Llanes, A., Mora, J. C., & Serrano, R. (2017). Differential effects of SA and intensive AH courses on teenagers' L2 pronunciation. *International Journal of Applied Linguistics, 27*(2), 470–490. https://doi.org/10.1111/ijal.12151

Logistic Regression. (n.d.). UCLA: Statistical Consulting Group. Retrieved December 19, 2020 from https://stats.idre.ucla.edu/spss/output/logistic-regression/

Long, M. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.

Lord, G., & Fionda, M. I. (2014). Teaching pronunciation in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 514–529). Wiley Blackwell.

Lozano-Argüelles, C., Sagarra, N., & Casillas, J. (2020). Slowly but surely: Interpreting facilitates L2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition, 23*(4), 752–762. https://doi.org/10.1017/S1366728919000634

Lumsden, J., Leung, R., D'Amours, D., & McDonald, D. (2010). ALEX©: A mobile adult literacy experiential learning application. *International Journal of Mobile Learning and Organisation*, *4*(2), 172–191. https://doi.org/10.1504/IJMLO.2010.032635

MacDonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect, 17*(3), 3–18.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America, 110*(1), 516–528. https://doi.org/10.1121/1.1377287

Mahrt, T. (2016). *PraatIO*. Retrieved March 28, 2021 from https://pypi.org/project/praatio/

Major, R. C. (2008). Transfer in second language phonology: A review. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition*, (Vol. 36, pp. 63–94). John Benjamins.

Martín Monje, E., Arús-Hita, J., Rodríguez-Arancón, P., Calle- Martínez, C. (2014). REALL: Rubric for the evaluation of apps in language learning. *Actas de las Jornadas Internacionales Tecnología Móvil e Innovación en el Aula: Nuevos Retos y Realidades Educativas*. Retrieved March 28, 2021 from http://eprints.ucm.es/25096/

Martinsen, R. A., Alvord, S. M., & Tanner, J. (2014). Perceived foreign accent: Extended stays abroad, level of instruction, and motivation. *Foreign Language Annals*, *47*(1), 66–78. https://doi.org/10.1111/flan.12076

Masip, J., Bethencourt, G. L., Sánchez-San Segundo, M., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology, 53*(2), 103-111. https://doi.org/103-111. 10.1111/j.1467-9450.2011.00931.x

Matte, C. (4 Feb 2019). Spanish SOLO: Learn Spanish app review. *Common Sense Media (commonsensemedia.org)*. Retrieved May 21, 2021 from https://www.commonsensemedia.org/app-reviews/spanish-solo-learn-spanish

McBride, K. (2015). Which features of Spanish learners' pronunciation most impact listener evaluations?. *Hispania*, *98*(1), 14–30. https://doi.org/10.1353/hpn.2015.0001

McCrocklin, S. (2012). The role of word stress in English as a lingua franca. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011.* (pp. 249−256). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

McHaney, R., Tako, A., & Robinson, S. (2018). Using LIWC to choose simulation approaches: A feasibility study. *Decision Support Systems, 111,* 1−12. https://doi.org/10.1016/j.dss.2018.04.002

Medimorec, S., Pennycook, G. (2015). The language of denial: Text analysis reveals differences in language use between climate change proponents and skeptics. *Climatic Change, 133*, 597–605. https://doi.org/10.1007/s10584-015-1475-2

Meng, H., Lo, W. K., Harrison, A. M., Lee, P., Wong, K. H., Leung, W. K., & Meng, F. (2010). Development of automatic speech recognition and synthesis technologies to support Chinese learners of English: The CUHK experience. In A. Ortega (Ed.), *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (pp. 811−820).

Meschyan, G., & Hernandez, A. E. (2005). Impact of language proficiency and orthographic transparency on bilingual word reading: An fMRI investigation. *NeuroImage*, *29*(4), 1135–1140. https://doi.org/10.1016/j.neuroimage.2005.08.055

Miller, J. S. (2012). Teaching pronunciation with phonetics in a beginner French course: Impact of sound perception. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference*, *Sept. 2011.* (pp. 109−123). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Mirisis, C. A. (2020). Production and perception of Spanish voiced stops and approximants by L2 learners. *Journal of Second and Multiple Language Acquisition, 8*(4), 79–110. Retrieved April 6, 2021 from https://www.science-res.com/index.php/jsmula/article/view/9

Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). Oxford University Press.

Mixdorff, H., Bach, N. H., Fujisaki, H., & Luong, M. C. (2003). Quantitative analysis and synthesis of syllabic tones in Vietnamese. In *Eighth European Conference on Speech Communication and Technology* (pp. 177–180). Retrieved from https://www.isca-speech.org/archive/eurospeech_2003/e03_0177.html

Mohr, J. (2004). *Randomly scrambled lists of numbers*. Augustana.ualberta.ca. Retrieved from http://www.augustana.ab.ca/~mohrj/algorithms/scramble.html

Moquin, L., & Salmons, J. (2020). Scandinavian-American English over time: Stereotypes and regionalization. *Ampersand, 7*, 1–8. https://doi.org/10.1016/j.amper.2020.100064

Morgan, T. (2006). On the teaching of Spanish pronunciation. *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, 119–130. Retrieved from http://www.lingref.com/cpp/lasp/2/paper1330.pdf

Morgan, T. A., & García, C. (2016). Seeing is believing: Student self-correction of Spanish pronunciation via online, real-time, visual feedback. Retrieved from R. Campos-Astorkiza, T. Morgan, & H. Nibert. *IGNITE: CASPSLaP Resources.* https://u.osu.edu/ignite2016

Morin, R. (2007). A neglected aspect of the standards: Preparing foreign language Spanish teachers to teach pronunciation. *Foreign Language Annals*, *40*(2), 342–360. https://doi.org/10.1111/j.1944-9720.2007.tb03206.x

Müller Pograjc, B., & Markič, J. (2017). Nasal vowels and diphthongs in European Portuguese: A problem for Slovene speakers. *Linguistica*, *57*(1), 243–254. https://doi.org/10.4312/linguistica.57.1.243-254

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). John Benjamins.

Nadeu, M. (2014). Stress-and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, *46*, 1–22. https://doi.org/10.1016/j.wocn.2014.05.003

Nagle, C. L. (2017). Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation*, *3*(2), 218–241. https://doi.org/10.1075/jslp.3.2.03nag

Nagle, C. L. (2018). Examining the temporal structure of the perception-production link in second language acquisition: A longitudinal study. *Language Learning, 68*(1), 234–270. https://doi.org/10.1111/lang.12275

Nagle, C. L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics, 40*(1), 86–107. https://doi.org/10.1093/applin/amx011

National Standards Collaborative Board, The. (2015). *World-Readiness Standards for Learning Languages*, 4th ed. Author.

Nemec, E. C., & Dintzner, M. (2016). Comparison of audio versus written feedback on writing assignments. *Currents in Pharmacy Teaching and Learning*, *8*(2), 155–159. https://doi.org/10.1016/j.cptl.2015.12.009

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism, 20*(3), 346–357. https://doi.org/10.1177/1367006914564566

Nicholes, J. (2017). Measuring writing engagement and emotional tone in L2 creative writing: Implications for interdisciplinarity. *Journal of Creative Writing Studies*, *2*(1), 1–21. https://scholarworks.rit.edu/jcws/vol2/iss1/2

Noble, J., & Lacasa, J. (2010). *Introduction to Quechua: Language of the Andes*. Dog Ear Publishing.

Norkevičius, G., Raškinis, G., & Kazlauskienė, A. (2005). Knowledge-based grapheme-to-phoneme conversion of Lithuanian words. In G. Kokkinakis, N. Fakotakis, E. Dermatas, & R. Potapova (Eds.), *SPECOM 2005: Speech and computer: 10th international conference proceedings, 17-19 October, 2005, University of Patras, Greece, 1,* (pp. 235–238). Moscow State Linguistics University.

O'Rourke, E. (2010). Dialect differences and the bilingual vowel space in Peruvian Spanish. In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology* (pp. 20–30). Somerville, MA: Cascadilla Proceedings Project.

Odden, D. (2005). *Introducing Phonology*. Cambridge University Press.

Odlin, T. (2014). Rediscovering prediction. In ZH. Han & E. Tarone (Eds.), *Interlanguage: Forty years later* (pp. 27–46). John Benjamins.

Odriozola, I., Navas, E., Hernáez, I., Sainz, I., Saratxaga, I., & Sánchez, J. (2012). D.: Using an ASR database to design a pronunciation evaluation system in Basque. In N. Calzolari, K. Choukri, T. Declerck, M. Ugur Dogan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), *Proceedings of the 8th international conference on language resources and evaluation (LREC. 2012)* (pp. 4122–4126).

Offerman, H. M. (2020). *The effects of pronunciation instruction on L2 production and L2 perception in Spanish: A comparative analysis* [Doctoral dissertation, Purdue University Graduate School]. Figshare.

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System*, *59*, 45–60. https://doi.org/10.1016/j.system.2016.03.003

Office for Academic Affairs - Testing Services. (2018). *Foreign language assessment*. The University of Alabama. Retrieved October 1, 2018 from https://testing.ua.edu/foreign-language-assessment

Oliver, K. M., Houchins, J. K., Moore, R. L., Wang, C. (2021). Informing makerspace outcomes through a linguistic analysis of written and video-recorded project assessments. *International Journal of Science and Mathematics Education, 19*, 333–354. https://doi.org/10.1007/s10763-020-10060-2

Olsen, M. K. (2012). The L2 acquisition of spanish rhotics by L1 English speakers: The effect of L1 articulatory routines and phonetic context for allophonic variation. *Hispania*, *95*(1), 65–82. https://doi.org/10.1353/hpn.2012.0008

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*(6), 407–420. https://doi.org/10.1016/j.wocn.2013.07.005

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, *97*(1), 47–68. https://doi.org/10.1353/hpn.2014.0030

Oxford University Press. (n.d.-a). Arced. In *Oxford English Dictionary online (oed.com).* Retrieved March 13, 2021 from https://www.oed.com/view/Entry/10246?redirectedFrom=arced#eid

Oxford University Press. (n.d.-b). Arcing. In *Oxford English Dictionary online (oed.com).* Retrieved March 13, 2021 from https://www.oed.com/view/Entry/10246?redirectedFrom=arcing#eid40300338

Oxford University Press. (n.d.-c). Chutzpah. In *Oxford English Dictionary online (oed.com).* Retrieved September 20, 2018 from http://www.oed.com/view/Entry/32880?redirectedFrom=chutzpah#eid

Oxford University Press. (n.d.-d). Clique. In *Oxford English Dictionary online (oed.com).* Retrieved September 20, 2018 from http://www.oed.com/view/Entry/34455?rskey=TbJltV&result=1&isAdvanced=false#eid

Oxford University Press. (n.d.-e). Equipage. In *Oxford English Dictionary online (oed.com).* Retrieved May 21, 2021 from https://www.oed.com/view/Entry/63780?rskey=viyH94&result=1#eid

Oxford University Press. (n.d.-f). Jalapeño. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/243609?redirectedFrom=jalapeño#eid

Oxford University Press. (n.d.-g). Mannequin. In *Oxford English Dictionary online (oed.com)*. Retrieved May 21, 2021 from https://www.oed.com//view/Entry/113568?redirectedFrom=mannequin#eid

Oxford University Press. (n.d.-h). Mosquito. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/122564?redirectedFrom=mosquito#eid

Oxford University Press. (n.d.-i). Patzer. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/139008?redirectedFrom=patzer#eid

Oxford University Press. (n.d.-j). Polyphony. In *Oxford English Dictionary online (oed.com)*. Retrieved April 8, 2021 from https://www.oed.com/view/Entry/147305?redirectedFrom=polyphony#eid

Oxford University Press. (n.d.-k). Provincial. In *Oxford English Dictionary online (oed.com)*. Retrieved March 13, 2021 from https://www.oed.com/view/Entry/153461?redirectedFrom=provincial#eid

Oxford University Press. (n.d.-l). Quesadilla. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/156330?redirectedFrom=quesadilla#eid

Oxford University Press. (n.d.-m). Queso. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/57158670?redirectedFrom=queso#eid

Oxford University Press. (n.d.-n). Questionnaire. In *Oxford English Dictionary online (oed.com)*. Retrieved May 21, 2021 from https://www.oed.com/view/Entry/156364?redirectedFrom=questionnaire#eid

Oxford University Press. (n.d.-o). Quiche. In *Oxford English Dictionary online (oed.com)*. Retrieved September 20, 2018 from http://www.oed.com/view/Entry/156415?rskey=7rnzkU&result=2&isAdvanced=false#eid

Oxford University Press. (n.d.-p). Quixotic. In *Oxford English Dictionary online (oed.com)*. Retrieved April 5, 2021 from https://www.oed.com/view/Entry/156828?redirectedFrom=quixotic#eid

Oxford University Press. (n.d.-q). Sequin. In *Oxford English Dictionary online (oed.com)*. Retrieved May 21, 2021 from https://www.oed.com/view/Entry/176323?rskey=vbaqEg& result=1&isAdvanced=false#eid

Oxford University Press. (n.d.-r). Smartphone. In *Oxford English Dictionary online (oed.com).* Retrieved May 27, 2021 from https://www.oed.com/view/Entry/381083?redirectedFrom=smartphone#eid

Oxford University Press. (n.d.-s). Technique. In *Oxford English Dictionary online (oed.com).* Retrieved September 20, 2018 from http://www.oed.com/view/Entry/198458?redirectedFrom=technique#eid

Oxford University Press. (n.d.-t). Tequila. In *Oxford English Dictionary online (oed.com).* Retrieved May 21, 2021 from https://www.oed.com/view/Entry/199314?redirectedFrom=tequila#eid

Paradis, C. (2006). The unnatural /Cju/ (< foreign /Cy/) sequence in Russian loanwords: A problem for the perceptual view. *Lingua, 116*(7), 976–995. https://doi.org/10.1016/j.lingua.2005.06.005

Paradis, C. & La Charité, D. (1997). Preservation and minimality in loanword adaptation. *Journal of Linguistics, 33*, 379–430. https://www.jstor.org/stable/4176422

Pasfield-Neofitou, S., Huang, H., & Grant, S. (2015). Lost in Second Life: Virtual embodiment and language learning via multimodal communication. *Educational Technology Research and Development, 63*(5), 709–726. https://doi.org/10.1007/s11423-015-9384-7

Pennebaker, J. W. (2013). *The secret life of pronouns*. Bloomsbury Press.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). *Linguistic Inquiry and Word Count: LIWC2015* (1.4.0). Pennebaker Conglomerates, Inc. http://liwc.wpengine.com

Pennebaker, J.W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Pennebaker Conglomerates. (2015). *How it works*. LIWC. Retrieved March 28, 2021 from https://liwc.wpengine.com/how-it-works/

Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences,* 367–370. http://www.internationalphoneticassociation.org/icphs/icphs2003

Pereltsvaig, A. (2011 Dec. 14). *The history of English orthography*. Languages of the world. Retrieved from https://www.languagesoftheworld.info/student-papers/the-history-of-english-orthography.html

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish-English spontaneous code-switching. *Journal of Phonetics, 52,* 121–137. https://doi.org/10.1016/j.wocn.2015.07.004

Polat, B. (2013). Language experience interviews: What can they tell us about individual differences?. *System*, *41*(1), 70–83. https://doi.org/10.1016/j.system.2013.01.021

Polat, B. (2014). *Words of experience: Semantic content analysis and individual differences among successful second language learners* [Doctoral dissertation, Georgia State University] Scholarworks. https://scholarworks.gsu.edu/alesl_diss/28

Porter, S. M. (2010). *Orthographic influence on the perception and production of Spanish loans in English* (Publication No. 1490327) [Master's thesis, California State University, Long Beach]. ProQuest LLC.

Pullum, G. K., & Ladusaw, W. A. (1996). *Phonetic symbol guide* (2nd edition). The University of Chicago Press.

R Core Team. (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Rafat, Y. (2015). The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative rhotics. *Applied Psycholinguistics*, *36*(1), 43–66. https://doi.org/10.1017/S0142716414000423

Ranganathan, P., Pramesh, C.S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research, 8*(3), 148–151. https://doi.org/10.4103/picr.PICR_87_17

Rao, R., Fuchs, Z., Polinsky, M., & Parra, M. L. (2020). The sound pattern of heritage Spanish: An exploratory study on the effects of a classroom experience. *Languages*, *5*(4), 1–37. https://doi.org/10.3390/languages5040072

Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction* (2nd edition). Bloomsbury Academic.

Ravand, H. & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation, 20*(1), 1–12. https://doi.org/10.7275/5g6f-ak15

Regueira, X. L. (1996). Galician. *Journal of the International Phonetic Association, 26*(2), 119–122. https://www.jstor.org/stable/44526207

Reinhardt, J., & Sykes, J.M. (2014). Special issue commentary: Digital game activity in L2 teaching and learning. *Language Learning & Technology, 18*(2), 2–8. http://llt.msu.edu/issues/june2014/commentary.pdf

Repp, B. H., & Lin, H-B. (1991). Effects of preceding context on the voice-onset-time category boundary. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 289–302. https://doi.org/10.1037/0096-1523.17.1.289

Richards, M. (2012). Helping Chinese learners distinguish English /l/ and /n/. In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference*, *Sept. 2011*. (pp. 161–167). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology, 32*(4), 469–479. https://doi.org/10.1177/0261927X13476869

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The evaluative lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods*, *50*(4), 1327–1344. https://doi.org/10.3758/s13428-017-0975-6

Rúas-Araújo, J., Puentes-Rivera, I., & Míguez-González, M. I. (2016). Capacidad predictiva de Twitter, impacto electoral y actividad en las elecciones al Parlamento de Galicia: Un análisis con la herramienta LIWC. *Observatorio*, *10*(2), 55–87.

Saalfeld, A. K. (2012). Teaching L2 Spanish stress. *Foreign Language Annals, 45*(2), 283–303. https://doi.org/10.111/j.1944-9720.2012.01191.x

Salas-Zárate M.P., Paredes-Valverde M.A., Rodríguez-García M.Á., Valencia-García R., Alor-Hernández G. (2017). Sentiment analysis based on psychological and linguistic features for Spanish language. In G. Alor-Hernández & R. Valencia-García (Eds.), *Current trends on knowledge-based systems*. *Intelligent Systems Reference Library, 120*, (pp. 73–92). Springer, Cham. https://doi.org/10.1007/978-3-319-51905-0_4

Saraçlar, M. (2001). *Pronunciation modeling for conversational speech recognition* (AAT 85538686) [Doctoral dissertation, The Johns Hopkins University]. ProQuest Digital Dissertations.

Savicki, V., & Price, M. V. (2015). Student reflective writing: Cognition and affect before, during, and after study abroad. *Journal of College Student Development*, *56*(6), 587–601. https://doi.org/10.1353/csd.2015.0063

Schairer, K. E. (1992). Native speaker reaction to non-native speech. *The Modern Language Journal, 76*(3), 309–319. https://www.jstor.org/stable/330161

Schwegler, A. & Kempff, J. (2007). *Fonética y fonología españolas* (3rd edition). John Wiley & Sons, Inc.

Sell, J., & Farreras, I. G. (2017). LIWC-ing at a century of introductory college textbooks: Have the sentiments changed?. *Procedia Computer Science*, *118*, 108–112. https://doi.org/10.1016/j.procs.2017.11.151

Sercu, L. (2013). Weblogs in foreign language education: Real and promised benefits. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *7th International technology, education and development conference, Valencia (Spain), 4-6 March 2013.* (pp. 4355–4366). IATED.

Silverman, D. (1992). Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology*, *9*(2), 289–328. https://www.jstor.org/stable/4420058

Simon, E., Chambless, D., & Kickhöfel Alves, U. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences*, *32*(3), 380–394. https://doi.org/10.1016/j.langsci.2009.07.001

Simon, E., & D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences*, *34*(3), 269–283. https://doi.org/10.1016/j.langsci.2011.10.002

Simonet, M. (2014a). The L2 acquisition of Spanish phonetics and phonology. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 729–746). John Wiley and Sons.

Simonet, M. (2014b). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, *43*(1), 26–37. https://doi.org/10.1016/j.wocn.2014.01.004

Simonović, M. (2015). *Lexicon immigration service: Prolegomena to a theory of loanword integration* (Vol. 393) [Doctoral dissertation, Utrecht University]. LOT Dissertation Series. Retrieved from https://dspace.library.uu.nl/bitstream/handle/1874/312539/Simonovic.pdf?sequence=1&isAllowed=y

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Smith-Keiling, B. L., & Hyun, H. I. F. (2019). Applying a computer-assisted tool for semantic analysis of writing: Uses for STEM and ELL. *Journal of Microbiology & Biology Education*, *20*(1), 1–6. https://doi.org/10.1128/jmbe.v20i1.1709

Smith-Keiling, B. L., Swanson, L. K., & Dehnbostel, J. M. (2018). Interventions for supporting and assessing science writing communication: Cases of Asian English language learners. *Journal of Microbiology & Biology Education*, *19*(1), 1–15. https://doi.org/10.1128/jmbe.v19i1.1522

Starkweather, J. (2018, Nov.). *Reference category and interpreting regression coefficients in R.* Research Matters, Benchmarks Online. Retrieved December 19, 2020 from https://it.unt.edu/interpreting-glm-coefficients

Steel, C. (2012, November). Fitting learning into life: Language students' perspectives on benefits of using mobile apps. In M. Brown, M. Hartnett & T. Stewart (Eds.), Proceedings *ascilite Wellington 2012* (pp. 875–880).

Stefanich, S., & Cabrelli, J. (2020). Shared or separate representations? The Spanish palatal nasal in early Spanish/English bilinguals. *Languages*, *5*(4), 1–21. https://doi.org/10.3390/languages5040050

Stockwell, G., & Hubbard, P. (2013). Some emerging principles for Mobile-assisted Language Learning. *The International Research Foundation for English Language Education*, 1–15. https://doi.org/10.5539/ijel.v3n1p19

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age of learning on voice onset time: Categorical perception of Swedish stops by near-native L2 speakers. *Language and Speech*, *57*(4), 425–450. https://doi.org/10.1177/0023830913508760

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). Effects of age and speaking rate on voice onset time. *Studies in Second Language Acquisition*, *37*(1), 71–100. https://doi.org/10.1017/S0272263114000151

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition*, ((36), pp. 153−192). Benjamins.

Strik, H., Truong, K., de Wet, F., & Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication*, *51*(10), 845–852. https://doi.org/10.1016/j.specom.2009.05.007

Tausczik, Y. L., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24−54. https://doi.org/10.1177/0261927X09351676

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744–765. https://doi.org/10.11139/cj.28.3.744-765

Thorne, S. L. (2000). Second language acquisition theory and the truth(s) about relativity. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning.* (pp. 219−244). Oxford University Press.

Timmer, K., & Schiller, N. O. (2012). The role of orthography and phonology in English: An ERP study on first and second language reading aloud. *Brain Research*, *1483*, 39–53. https://doi.org/10.1016/j.brainres.2012.09.004

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment*, *25*(4), 1069−1078. https://doi.org/10.1037/a0033007

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal*, *90*(4), 519–535. https://doi.org/10.1111/j.1540-4781.2006.00464.x

Tsai, M., & Lee, L.-S. (2003). Pronunciation variation analysis based on acoustic and phonemic distance measures with application examples on Mandarin Chinese. *Proceedings of the 2003 IEEE workshop on automatic speech recognition and understanding (ASRU '03)*, 117–122. https://doi.org/10.1109/ASRU.2003.1318414

Turton, D. (2014). Some /l/s are darker than others: Accounting for variation in English /l/ with ultrasound tongue imaging. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 189–198. https://repository.upenn.edu/pwpl/vol20/iss2/21

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, *45*(1), 26–42. https://doi.org/10.1016/j.wocn.2014.02.008

Underwood, J., Cuadrado, C., Gaudioso, B., Melero, P., & Sacristán, E. (2012). *Protagonistas. A Communicative Approach*. Vista Higher Learning.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Valenti, A., Chita-Tegmark, M., Gold, M., Law, T., & Scheutz, M. (2019). In their own words: A companion robot for detecting the emotional state of persons with Parkinson's Disease. In M. Salichs et al. (Eds.), *Social Robotics. ICSR 2019. Lecture notes in computer science, 11876*, (pp. 443–452). Springer, Cham. https://doi.org/10.1007/978-3-030-35888-4_41

van Santen, J. P., & Sproat, R. W. (1999). High-accuracy automatic segmentation. *Eurospeech*, 2809–2812. http://crow.ee.washington.edu/people/bulyko/papers/segm.pdf

Viberg, O., & Grönlund, Å. (2012). Mobile assisted language learning: A literature review. *In Proceedings of the 11th international conference on mobile and contextual learning*, 1–8. http://du.diva-portal.org/smash/record.jsf?pid=diva2:549644

Vilares, D., Thelwall, M., & Alonso, M. A. (2015). The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*, *41*(6), 799–813. https://doi.org/10.1177/0165551515598926

Walla P. (2018) Affective processing guides behavior and emotions communicate feelings: Towards a guideline for the NeuroIS community. In: F. Davis, R. Riedl, J. vom Brocke, PM. Léger, & A. Randolph (Eds.), *Information systems and neuroscience. Lecture notes in information systems and organisation,* (Vol 25, pp. 141–150). Springer, Cham. https://doi.org/10.1007/978-3-319-67431-5_16

Wang, L., Chen, H., Li, S., & Meng, H. M. (2012). Phoneme-level articulatory animation in pronunciation training. *Speech Communication*, *54*(7), 845–856. https://doi.org/10.1016/j.specom.2012.02.003

Wang, Y. B., & Lee, L. S. (2012). Improved approaches of modeling and detecting error patterns with empirical analysis for computer-aided pronunciation training. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 5049–5052. https://doi.org/10.1109/ICASSP.2012.6289055

Whaley, J., & Kibby, M. (1981). The Relative Importance of Reliance on Intraword Characteristics and Interword Constraints for Beginning Reading Achievement. *The Journal of Educational Research, 74*(5), 315–320. Retrieved April 8, 2021, from http://www.jstor.org/stable/27539830

Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, *39*(2), 388–398. https://doi.org/10.1121/1.1909900

Wong, K. H., Leung, W. K., Lo, W. K., & Meng, H. (2010). Development of an articulatory visual-speech synthesizer to support language learning. In *2010 7th international symposium on Chinese spoken language processing* (pp. 139–143). IEEE.

Wong, W., & VanPatten, B. (2003). The evidence is IN: Drills are OUT. *Foreign Language Annals, 36*(3), 403–423. https://doi.org/10.1111/j.1944-9720.2003.tb02123.x

Yamashita, J. (2013). Word recognition subcomponents and passage level reading in a foreign language. *Reading in a Foreign Language, 25*(1), 52–71. Retrieved April 6, 2021 from http://nflrc.hawaii.edu/rfl

Yang, I. Y. (2017). Instructional effects on voice onset time perception in L2 speech segmentation. *Language Research, 53*(3), 561–585. http://hdl.handle.net/10371/138518

Youngs, B., Ducate, L., & Arnold, N. (2011). Linking second language acquisition, CALL, and language pedagogy. *Present and future promises of CALL: From theory and research to new directions in language teaching*, *5* (2nd edition), pp. 23–60. CALICO.

Zabrodskaja, A. (2007). Russian-Estonian code-switching in the university. *Arizona Working Papers in SLA Teaching*, *14*, 123–139. http://w3.coh.arizona.edu/awp/

Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania, 77*(3), 470–481. https://doi.org/10.2307/344974

Zampini, M. L. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*, *3*(3), 85–100.

Zampini, M. L. (2008). L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219–249). Benjamins.

Zampini, M. L. (2014). Voice onset time in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 111–129). John Wiley & Sons, Ltd.

Zapata, G. C., & Ribota, A. (2021). The instructional benefits of identity texts and learning by design for learner motivation in required second language classes. *Pedagogies: An International Journal, 16*(1), 1–18. https://doi.org/10.1080/1554480X.2020.1738937

Zhang, H. (2013). Pedagogical challenges of spoken English learning in the Second Life virtual world: A case study. *British Journal of Educational Technology, 44*(2), 243–254. https://doi.org/10.1111/j.1467-8535.2012.01312.x

Zhao, J., Yuan, H., Leung, W. K., Meng, H., Liu, J., & Xia, S. (2013, May). Audiovisual synthesis of exaggerated speech for corrective feedback in computer-assisted pronunciation training. In *2013 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 8218–8222). IEEE. https://doi.org/10.1109/ICASSP.2013.6639267

Zhu, M., Herring, S. C., & Bonk, C. J. (2019). Exploring presence in online learning through three forms of computer-mediated discourse analysis. *Distance Education*, *40*(2), 205–225. https://doi.org/10.1080/01587919.2019.1600365

Zielinski, B. (2012). The social impact of pronunciation difficulties: Confidence and willingness to speak. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference, Sept. 2011*. (pp. 18–26). Iowa State University. Retrieved from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/06/Proceedings_3rd_PSLLT.pdf

APPENDICES

APPENDIX A

Supplemental Tables

Table A1

Number of Tokens in Stimuli by Grapheme, with Target Pronunciation

Grapheme(s)	Position	Target Pronunciation	Total
<p>	WI	[p^h]	31
	WM	[p^h]	30
<t>	WI	[t^h]	37
	WM	[t^h]	35
<c{a, o, u}>	WI (*n* = 33)	[k^h{a, o, u}]	140
	WM (*n* = 30)		
<c{i, e}>	n/a	[s{i, ɛ}]	44
<g{a, o, u}>	n/a	[g{a, o, u}]	42
<g{i, e}>	n/a	[h{i, ɛ}]	30
<h>	n/a	Ø	38
<ll>	n/a	[j]	32
<ñ>	n/a	[n.j]	30
<j>	n/a	[h]	40
<ch>	n/a	[t͡ʃ]	33
<z>	n/a	[s]	34
		Total:	596

Table A2

Breakdown of Participants by Class Section and Assigned Technology

	Spanish SOLO app		Online Homework		Blackboard Learn	
Section	Y/N	#	Y/N	#	Y/N	#
001	N	0	N	0	Y	3
002	Y	4	N	0	N	0
003	N	0	N	0	Y	5
004	Y	2	N	0	Y	1
005	N	0	Y	3	N	0
006	Y	2	N	0	N	0

007	N	0	Y	7	N	0
008	N	0	Y	3	N	0
009	N	0	N	0	Y	1
010	N	0	Y	6	N	0
011	Y	2	N	0	N	0
012	N	0	Y	1	N	0
013	N	0	N	0	Y	1
014	Y	2	N	0	N	0
015	Y	3	N	0	N	0
016	Y	5	N	0	N	0
017	N	0	N	0	Y	1
018	Y	2	N	0	Y	2
Total/Tech		22		20		14

Table A3

Examples of Regional Variation used as Criteria for Potential Exclusion from Analysis[103]

Grapheme(s)	Canonical Spanish Pronunciation	Example(s) of Regional Variation	Articulatory Features
<c{i , e}>	[s]	distinción: [θ]	voiceless interdental fricative
<g{i, e}>	[x]	[ç]	voiceless palatal fricative
<ll>	[ʝ]	[ʒ] lleísmo: [ʎ]	voiced pre-palatal fricative voiced palatal lateral
<j>	[x]	[ç]	voiceless palatal fricative
<ch>	[t͡ʃ]	[ʃ] [c] [ɟ] [t͡s]	voiceless pre-palatal fricative voiceless palatal stop voiced palatal stop voiceless alveolar affricate
<z>	[s]	ceceo: [θ]	voiceless interdental fricative
<rr>	[r]	[ʒ] [ʀ]	voiced pre-palatal fricative voiced uvular trill
<s>	[s]	e.g., *este* ‘this’ [e^{h}.te]	aspiration

[103] The information in this table was drawn from Hualde (2014), Hualde et al. (2010), Lipski (1994), and Pullum and Ladusaw (1996); it is not intended to be an exhaustive list of Spanish allophonic regional variation.

Table A4

Overall Accuracy (Raw Frequency, N=596) per Participant per Group, Pre-Test to Post-Test

EG1			EG2			CG		
Participant	Pre	Post	Participant	Pre	Post	Participant	Pre	Post
1A	369	405	2A	326	367	CA	371	382
1B	367	397	2B	309	317	CB	377	389
1C	289	293	2C	370	397	CC	390	410
1D	418	431	2D	416	449	CD	329	381
1E	371	405	2E	357	367	CE	343	362
1F	332	363	2F	377	395	CF	341	355
1G	415	417	2G	426	415			
1H	426	436	2H	428	444			
1I	269	310	2I	340	384			
1J	424	434	2J	383	421			
1K	354	375	2K	413	424			
1L	318	361	2L	368	415			
1M	352	367						

Table A5

Breakdown of Pronunciation Activities in the Online Companion to the Textbook, Protagonistas

Activities explicitly focusing on consonants[104]		Times study graphemes were explicitly in directions		Times other graphemes were explicitly in directions	
c before vowels	1	<j>	7	<p>[105]	6
g before vowels	1	<c>	6	<d>	6
b, v	3	<g>	6	<t>[106]	5
ch, ll, ñ	2	<ch>	6	<n>	4
h, j, g	4	<h>	4	<m>	4
j, g	1	<ñ>	2	<q>	4
r	2	<ll>	2	<r>	4
r, rr	2	<z>	1	<b>	4
c before consonant, q	3			<v>	3
c, q, k	1			<rr>	2
b, p	1			<k>	1
ch, p	4				
c, z	1				
d, t	4				

[104] Activities found in online component of *Protagonsitas* (Underwood et al., 2012)
[105] The grapheme <p> was part of the acoustic portion of the present study, but not the articulatory portion, hence its classification as "other".
[106] The grapheme <t> was part of the acoustic portion of the present study, but not the articulatory portion, hence its classification as "other".

p, t	1		
m, n	4		
d, j word final	2		
Total:	37 (44%)	34 (44%)	43 (56%)

Table A6

Additional Commentary, by Participant Group: Positive Impressions of Technology

	Experimental Group 1
(90)	The technology gave me oppurtunities [*sic*] to go back and improve my skills as many times as I wanted.
(91)	It was very easy to use and I liked having it on my phone where I could access it anytime I needed it.
(92)	I liked that i [*sic*] was able to learn words and also work on my accent in Spanish. The best feature was when you had to order the sentence correctly, because it had you practice grammar, vocabulary, and pronunciation all in the same activity.
	Experimental Group 2
(93)	I really liked the ability to practice with any activity that was available online. Without this study, I would not have known about these activities and extra study practice available to me. I also like how the activities are not graded or necessarily even seen by my teacher, so I can do them on my own time, and as needed.
(94)	I think that the technology is an awesome resource for those who have trouble in pronunciation areas! it helped me a ton!
(95)	The ones were [*sic*] there was a word and you could listen to the correct pronunciation and then record yourself saying it back.
	Control Group
(96)	It was very easy to access due to it being on BlackBoard [*sic*]. It was right next to all of my other classes that I accessed everyday [*sic*], so it was easy find.
(97)	Yes, because extra practice never hurt anyone. The lessons are more helpful than the homework because you get unlimited tries, so you have a chance to work on what you have problems with.
(98)	There was a lot of content to chose [*sic*] from, I liked that.

Table A7

Additional Commentary, by Participant Group: Impressions of Feedback from Technology

	Experimental Group 1
(99)	The best feature was the ability to use the speaking aspect in all of the activities. (even ones that were not focused on speaking necessarily). I liked this because it gave multiple opportunities to practice and get feedback on whether you were correct or not.
(100)	At certain times, the technology did not recognize my pronunciations of the words, even when they were correct. It was frustrating, but rare.
(101)	Sometimes I knew I was pronouncing the word correctly but I would still get it wrong.
	Experimental Group 2
(102)	It allowed me to know when I was wrong and helped me improve my pronunciation.
(103)	In VHL most of the activities I did allowed me to compare what I said to how it should all be pronounced and that helped a lot.

Table A8

Additional Commentary, by Participant Group: Technological Issues

	Experimental Group 1
(104)	The technology i [*sic*] used would sometimes crash and shut me out and not let me finish my lesson.
(105)	I would just like to mention as i [*sic*] have already, that the app did not work on multiple devices which made it hard when i [*sic*] got a new phone because that was what i [*sic*] had the app on. Also, the app would often become slow and crash on me.
(106)	I didn't like that sometimes the app would quit on me in the middle of a lesson.
(107)	It would sometimes lock me out and i [*sic*] would have to start over.
	Experimental Group 2
(108)	What I disliked was that sometimes the technology would not allow me to record what I needed to. It only happened once or twice, but I felt as if I was not doing my part and therefore altering the study results.
(109)	I thought it was all very beneficial, but if I had to choose one thing I would say sometimes it was difficult to respond in full sentences when I didn't know the vocab very well yet.

(110)	Some of the potential activities included more than just pronunciation. For these activities, I would have to say multiple sentences before I could hear the correct pronunciation. I could not correct my pronunciation immediately.

Table A9

Additional Commentary, by Participant Group: Suggestions for Improvement

	Experimental Group 1
(111)	I would update the technology so it would not crash as often and so it would be able to be used on multiple devices.[107]
(112)	Fix the fill in the blank glitch and provide a way for you to target words that you repeatedly mess up.
(113)	Fix the crashing, improve the listening feature even further.
(114)	More matching up and listening opposed to reading the question then matching.
	Experimental Group 2
(115)	I would maybe add pictures with the words that the participant is supposed to repeat in order to help the participant learn the vocab words while saying them. Maybe even add what the word is in English beside it.
(116)	To improve the technology, I would allow multiple attempts like I previously said. Also, I would add a grammar review. The lessons available are only on vocabulary, and a grammar lesson would help on the grammar activities.
(117)	Some of the instructions on the activities could be more clear.
	Control Group
(118)	I would add a little more variety to the tests, like VHL does. I used VHL for my SP-101 class, so it might have been even better practice to use VHL for it.
(119)	I'd make the head sets bigger, or make it a single ear piece. Also, I'd include the English translation somewhere (even if it is just an optional thing at the end). I'm not the biggest fan of only knowing about half of what I am saying.

[107] The Spanish SOLO app was only available for iOS devices (see Section 3.4.1 for further explanation).

Table A10

Additional Commentary, by Participant Group: Impact on Performance

	Experimental Group 1
(120)	YES!! This app allowed me to practice all the time. Because I had to speak aloud for some of the activities it better prepared me for speaking activities in class. Most of the app correlated perfectly with my class, as well.
(121)	I think this technology helped me have more confidence in replying in Spanish in class. Although I still stutter on some words I definitely think that I have learned to just try the words instead of holding back because of this app.
(122)	Yes. We usually do not work on vocab in class, so it helped me learn the verbs. It also helped me understand verbs like "ser" and "estar"[108]. Those confused me in class.
(123)	Yes. Specifically with my speaking pruebas[109]. It helped me with my pronunciation which helped me to preform [*sic*] better on my speaking quizzes.
	Experimental Group 2
(124)	I think it helped for sure. I was able to pronounce words more accurately and was able to comprehend topics more.
(125)	I definitely do. I noticed my speaking became much more natural when recording, although I improved my pronunciation a lot because I understand better where the stressed syllables should be. I think I have a better handle on accents and new vocabulary words.
(126)	It allowed me to get more practice speaking Spanish than I normally would in a classroom setting.
	Control Group
(127)	Slightly. It was a good review, and review is always healthy. Other than that it had no substantial effect on my spanish [*sic*] class performance.
(128)	Not really. Just that this has definitely been a unique experience, and I at least feel like my diction has been enhanced.

[108] *ser* 'to be'; *estar* 'to be'
[109] *pruebas* 'quizzes'

Table A11

Additional Commentary, by Participant Group: Rationales behind Recommendations

	Experimental Group 1
(129)	Yes! It was extremely helpful with vocabulary. Where it doesn't teach you much about forming sentences it gives you a great beginner understanding.
(130)	Yes I would mention this technology. It has more practical vocabulary that you would actually use day to day, and it has so many more features than other apps that make learning Spanish more enjoyable.
(131)	Definitely. Especially if they were a beginner in Spanish. It really helped me understand my basics which in turn helped me understand the more we learned.
	Experimental Group 2
(132)	I would recommend the technology I used in this study to someone interested in learning Spanish. This was very beneficial and was probably a huge help to my grades. The audio learning was the best part. It's much better having audible homework than sitting in a lecture.
(133)	Absolutely. The website had many other activities as well as the ones where you talk in a microphone.

APPENDIX B

Tokens

Word	Transcription	Gloss	Count	Grapheme(s)	
				1st	2nd (3rd)
<c{aou}>					
abanicar	[a.βa.ni.ˈkaɾ]	‘to strike out’	1	<ca>	
acá	[a.ˈka]	‘(over) here’	4	<ca>	
acaba (de)	[a.ˈka.βa]	‘just’ 3SG-PRS	1	<ca>	
acaban (de)	[a.ˈka.βan]	‘just’ 3PL-PRS	1	<ca>	
acabas (de)	[a.ˈka.βas]	‘just’ 2SG-PRS	1	<ca>	
acabo (de)	[a.ˈka.βo]	‘just’ 1SG-PRS	2	<ca>	
acaso	[a.ˈka.so]	‘perhaps’	2	<ca>	
(ella se) aplicó	[a.pli.ˈko]	‘applied (herself) 3SG-PRET	1	<co>	
blanco	[ˈblaŋ.ko]	‘white’	1	<co>	
cabaña	[ka.ˈβa.ɲa]	‘cabana’	1	<ca>	<ñ>
cacao	[ka.ˈka.o]	‘cocoa’	1	<ca>	<ca>
cada	[ˈka.ða]	‘each’	6	<ca>	
calcetines	[kal.se.ˈti.nes]	‘socks’	1	<ca>	<ce> <t>
calle	[ˈka.ʝe]	‘street’	1	<ca>	<ll>
campeonato	[kam.pe.o.ˈna.to]	‘championship’	1	<ca>	
capaz	[ka.ˈpas]	‘capable’	1	<ca>	<p> <z>
capital	[ka.pi.ˈtal]	‘capital’	1	<ca>	<t>
Carmen	[ˈkaɾ.men]	(name) ‘Carmen’	1	<ca>	
carrozas	[ka.ˈro.sas]	‘floats’ (n.)	1	<ca>	<z>
cartas	[ˈkaɾ.tas]	‘letters’	1	<ca>	
casi	[ˈka.si]	‘almost’	1	<ca>	
castaña	[kas.ˈta.ɲa]	‘brown’	1	<ca>	<ñ>
colegio	[ko.ˈle.xi̯o]	‘school’	1	<co>	
colores	[ko.ˈlo.ɾes]	‘colors’	1	<co>	
cómo/como	[ˈko.mo]	‘how/like’	2/2	<co>	

comparto	[kom.ˈpaɾ.to]	‘share’ 1SG-PRS	1	<co>	
comprarme	[kom.ˈpɾaɾ.me]	‘to buy myself’	1	<co>	
comprarnos	[kom.ˈpɾaɾ.nos]	‘to buy ourselves’	1	<co>	
computación	[kom.pu.ta.ˈsi̯on]	‘computer science’	1	<co>	<ci>
computadoras	[kom.pu.ta.ˈðo.ɾas]	‘computers	1	<co>	
con	[ˈkon]	‘with	13	<co>	
conejo	[ko.ˈne.xo]	‘rabbit’	1	<co>	<j>
conocerte	[ko.no.ˈseɾ.te]	‘to meet you’	1	<co>	<ce>
considero	[kon.si.ˈðe.ɾo]	‘consider’ 1SG-PRS	1	<co>	
construida	[kons.ˈtɾu̯i.ða]	‘built’ (adj.)	1	<co>	
construyeron	[kons. tɾu.ˈʝe.ɾon]	‘built’ 3PL-PRET	1	<co>	
copa	[ˈko.pa]	‘cup’	1	<co>	
coro	[ˈko.ɾo]	‘chorus’	1	<co>	
corregir	[ko.re.ˈxiɾ]	‘to correct’	1	<co>	<gi>
correr	[ko.ˈreɾ]	‘to run’	1	<co>	
cosas	[ˈko.sas]	‘things’	1	<co>	
costa/Costa	[ˈkos.ta]	‘coast’ (with Rica) ‘Costa Rica’	2	<co>	
cualquier	[ku̯al.ˈki̯eɾ]	‘any’	1	<cu>	
cuándo/cuando	[ˈku̯an̪.do]	‘when’	4/1	<cu>	
cuántos	[ˈku̯an̪.tos]	‘how many’	2	<cu>	
cuarto	[ˈku̯aɾ.to]	‘quarter’	3	<cu>	
cuatro	[ˈku̯a.tɾo]	‘four’	3	<cu>	
cubanoamericanos	[ku.βa.no̯a.me.ɾi.ˈka.nos]	‘Cuban-Americans’ M.PL	1	<cu>	<ca>
cueva	[ˈku̯e.βa]	‘cave’	1	<cu>	
cuidado	[ku̯i.ˈða.ðo]	‘care’ (with ten) ‘be careful’	1	<cu>	

cumpleaños	[kum.ˈple̯a.ɲos]	‘birthday’	1	<cu>	<ñ>
(se) destacó	[des.ta.ˈko]	‘stood out’ 3SG-PRET	1	<co>	
documentos	[do.ku.ˈmen̪.tos]	‘documents’	1	<cu>	
ecológico	[e.ko.ˈlo.xi.ko]	‘ecological’	1	<co>	<gi> <co>
Ecuador	[e.ku̯a.ˈðoɾ]	‘Ecuador’	2	<cu>	
encontrar	[eŋ.kon̪.ˈtɾaɾ]	‘to find’	1	<co>	
enfocarnos	[eɱ.fo.ˈkaɾ.nos]	‘to focus ourselves’	1	<ca>	
escoge	[es.ˈko.xe]	‘choose’ 2SG-IMP	1	<co>	<ge>
espectáculo	[es.pek.ˈta.ku.lo]	‘show’ (n.)	1	<cu>	
facultad	[fa.kul̪.ˈtað]	‘faculty’	1	<cu>	
física	[ˈfi.si.ka]	‘physics’	1	<ca>	
frecuencia	[fre.ˈku̯en.si̯a]	‘frequency’	2	<cu>	<ci>
masticar	[mas.ti.ˈkaɾ]	‘to chew’	1	<ca>	
México	[ˈme.xi.ko]	‘Mexico’	1	<co>	
música	[ˈmu.si.ka]	‘music’	1	<ca>	
ocupados	[o.ku.ˈpa.ðos]	‘busy’ M.PL	1	<cu>	<p>
picar	[pi.ˈkaɾ]	‘to needle’	1	<ca>	
practicar	[pɾak.ti.ˈkaɾ]	‘to practice’	4	<ca>	
química	[ˈki.mi.ka]	‘chemistry’	1	<ca>	
recaes	[re.ˈka.es]	‘relapse’ 2SG-PRS	1	<ca>	
recogerlos	[re.ko.ˈxeɾ.los]	‘to pick them up’	1	<co>	<ge>
Rica	[ˈri.ka]	(with Costa) ‘Costa Rica’	1	<ca>	
Ricardo	[ri.ˈkaɾ.ðo]	(name) ‘Richard’	1	<ca>	
Rico	[ˈri.ko]	(with Puerto) ‘Puerto Rico’	2	<co>	
sacó	[sa.ˈko]	‘got’ 3SG-PRET	1	<co>	
sacudo	[sa.ˈku.ðo]	‘dust’ 1SG-PRS	1	<cu>	

Salamanca	[sa.la.ˈmaŋ.ka]	(name) ‘Salamanca’	2	<ca>	
típicos	[ˈti.pi.kos]	‘typical’ M.PL	1	<co>	
tocado	[to.ˈka.ðo]	‘played’ PTCP	1	<co>	
turísticas	[tu.ˈɾis.ti.kas]	‘tourist’ (adj.), F.PL	1	<ca>	
<ch>					
anoche	[a.ˈno.t͡ʃe]	‘last night’	1	<ch>	
chaqueta	[t͡ʃa.ˈke.ta]	‘jacket’	1	<ch>	
chica	[ˈt͡ʃi.ka]	‘girl’ F.SG	1	<ch>	<ca>
chicos	[ˈt͡ʃi.kos]	‘boys’ M.PL	2	<ch>	<co>
chino	[ˈt͡ʃi.no]	‘Chinese’	1	<ch>	
derecha	[de.ˈɾe.t͡ʃa]	‘right’	1	<ch>	
mochila	[mo.ˈt͡ʃi.la]	‘backpack’	1	<ch>	
mucha	[ˈmu.t͡ʃa]	‘many’ F.SG	3	<ch>	
muchachos	[mu.ˈt͡ʃa.t͡ʃos]	‘boys/kids’ M.PL	1	<ch>	<ch>
muchas	[ˈmu.t͡ʃas]	‘many’ F.PL	3	<ch>	
mucho	[ˈmu.t͡ʃo]	‘many’ M.SG	7	<ch>	
muchos	[ˈmu.t͡ʃos]	‘many’ M.PL	2	<ch>	
noche	[ˈno.t͡ʃe]	‘night’	3	<ch>	
ochenta	[o.ˈt͡ʃen̪.ta]	‘eighty’	1	<ch>	
Quichua	[ˈki.t͡ʃu̯a]	‘Quichua’	1	<ch>	
Sánchez	[ˈsanʲ.t͡ʃes]	(name) ‘Sánchez’	1	<ch>	<z>
<c{i,e}>					
asociado	[a.so.ˈsi̯a.ðo]	‘associated’ (adj.), M.SG	1	<ci>	
celebra	[se.ˈle.bɾa]	‘celebrate’ 3SG-PRS	1	<ce>	
cerca	[ˈseɾ.ka]	‘near’	2	<ce>	<ca>
ciclismo	[si.ˈklis. mo]	‘cycling’	1	<ci>	
cien	[ˈsi̯en]	‘one hundred’	1	<ci>	
ciencias	[ˈsi̯en.si̯as]	‘sciences’	1	<ci>	<ci>

ciento	[ˈsi̯en̪.to]	‘hundred'; (with por) ‘percent’	1	<ci>	
cierra	[ˈsi̯e.ra]	‘close’ 3SG-PRS	1	<ci>	
cinco	[ˈsiŋ.ko]	‘five’	1	<ci>	<co>
cincuenta	[siŋ.ˈku̯en̪.ta]	‘fifty’	1	<ci>	<cu>
ciudad	[si̯u.ˈðað]	‘city’	4	<ci>	
diciembre	[di.ˈsi̯em.bɾe]	‘December’	1	<ci>	
edificios	[e.ði.ˈfi.si̯os]	‘buildings’	1	<ci>	
electricidad	[e.lek.tɾi.si.ˈðað]	‘electricity’	1	<ci>	
Francisco	[fɾan.ˈsis.ko]	(name) ‘Frances’	1	<ci>	<co>
municipal	[mu.ni.si.ˈpal]	‘municipal’	1	<ci>	<p>
necesitamos	[ne.se.si.ˈta.mos]	‘need’ 1PL-PRS	3	<ce>	<t>
necesitas	[ne.se.ˈsi.tas]	‘needs’ 2SG- PRS	1	<ce>	
novecientos	[no.βeˈsi̯en̪.tos]	‘nine hundred’	1	<ci>	
oficial	[o.fi.ˈsi̯al]	‘official’	1	<ci>	
parece	[pa.ˈɾe.se]	‘seems’ 3SG- PRS	1	<ce>	
participa	[paɾ.ti.ˈsi.pa]	‘participates’ 3SG-PRS	1	<ci>	
porcentaje	[poɾ.sen̪.ˈta.xe]	‘percentage’	1	<ce>	<j>
recibe	[re.ˈsi.βe]	‘receive,’ 3SG- PRS	1	<ci>	
residencias	[re.si.ˈðen.si̯as]	‘residences’	1	<ci>	
sociología	[so.si̯o.lo.ˈxi.a]	‘sociology’	1	<ci>	<gi>
valenciana	[ba.len.ˈsi̯a.na]	‘Valencian’ F.SG	1	<ci>	
<g{a,o,u}>					
agosto	[a.ˈɣos.to]	‘August’	1	<go>	
amiga	[a.ˈmi.ɣa]	‘friend, F.SG’	1	<ga>	
antigua	[an̪.ˈti.ɣu̯a]	‘ancient’	1	<gu>	
figuras	[fi.ˈɣu.ɾas]	‘figures’ (n.)	1	<gu>	
Galápagos	[ga.ˈla.pa.ɣos]	‘Galapagos’	1	<ga>	<go>
ganado	[ga.ˈna.ðo]	‘won’ PTCP	1	<ga>	

ganar	[ga.ˈnaɾ]	‘to win’	1	<ga>	
Gómez	[ˈgo.mes]	(name) ‘Gómez’	1	<go>	<z>
gordo	[ˈgoɾ.ðo]	‘fat’ M.SG	1	<go>	
guapísimo	[gu̯a.ˈpi.si.mo]	‘very handsome’ M.SG	1	<gu>	<p>
guitarra	[gi.ˈta.ra]	‘guitar’	1	<gu>	<t>
gusta	[ˈgus.ta]	‘like’ 3SG-PRS	2	<gu>	
gustan	[ˈgus.tan]	‘like’ 3PL-PRS	3	<gu>	
gustaría	[gus.ta.ˈɾi.a]	‘like’ 3SG-COND	1	<gu>	
gusto	[ˈgus.to]	‘pleasure’; (with mucho) ‘nice to meet you’	1	<gu>	
igualmente	[i.ɣu̯al.ˈmen̪.te]	‘equally’	1	<gu>	
lengua	[ˈleŋ.gu̯a]	‘language’	1	<gu>	
lenguajes	[leŋ.ˈgu̯a.xes]	‘languages’	1	<gu>	<j>
lenguas	[ˈleŋ.gu̯as]	‘languages’	1	<gu>	
ninguna	[niŋ.ˈgu.na]	‘not any’ F.SG	1	<gu>	
obligatorio	[o.βli.ɣa.ˈto.ɾi̯o]	‘obligatory’ M.SG	1	<ga>	<t>
orgullo	[oɾ.ˈɣu.ʝo]	‘pride’	1	<gu>	<ll>
portugués	[poɾ.tu.ˈɣu̯es]	‘Portuguese,’ M.SG	1	<gu>	
siguiente	[si.ˈɣi̯en̪.te]	‘subsequent’	1	<gu>	
tortugas	[toɾ.ˈtu.ɣas]	‘turtles’	1	<ga>	
Uruguay	[u.ɾu.ˈɣu̯ai̯]	‘Uruguay’	1	<gu>	
<g{i,e}>					
agencia	[a.ˈxen.si̯a]	‘agency’	1	<ge>	<ci>
alergias	[a.leɾ.ˈxi̯as]	‘allergies’	1	<gi>	
Argentina	[aɾ.xen̪.ˈti.na]	‘Argentina’	1	<ge>	
biología	[bi̯o.lo.ˈxi.a]	‘biology’	3	<gi>	
dirige	[di.ˈɾi.xe]	‘conduct’ 3SG-PRS	1	<ge>	

Egipto	[e.ˈxip.to]	‘Egypt’	1	<gi>	
eliges	[e.ˈli.xes]	‘choose’ 3SG-PRS	1	<ge>	
energía	[e.neɾ.ˈxi.a]	‘energy’	1	<gi>	
frágil	[ˈfɾa.xil]	‘fragile’	1	<gi>	
gente	[ˈxen̪.te]	‘people’	2	<ge>	
geografía	[xe.o.gɾa.ˈfi.a]	‘geography’	1	<ge>	
gigantes	[xi.ˈɣan̪.tes]	‘giant’ PL	2	<gi>	<ga>
gigantesca	[xi.ɣan̪.ˈtes.ka]	‘gigantic’ F.SG	1	<gi>	<ga> <ca>
gimnasio	[xim.ˈna.si̯o]	‘gymnasium’	1	<gi>	
gira	[ˈxi.ɾa]	‘turn’ 2SG-IMP	1	<gi>	
girasoles	[xi.ɾa.ˈso.les]	‘sunflowers’	1	<gi>	
gitanos	[xi.ˈta.nos]	‘gypsies’ M.PL	1	<gi>	<t>
imagina	[i.ma.ˈxi.na]	‘imagine’ 2SG-IMP	1	<gi>	
origen	[o.ˈɾi.xen]	‘origin’	1	<ge>	
religiosos	[re.li.ˈxi̯o.sos]	‘religious’ M.PL	1	<gi>	
<h>					
ahora	[a.ˈo.ɾa]	‘now’	1	<h>	
ha	[ˈa]	(with part.) ‘have’ 3SG-PRS	1	<h>	
Habana	[a.ˈβa.na]	‘Habana’	1	<h>	
habla	[ˈa.βla]	‘speaks’ 3SG-PRS	1	<h>	
hablan	[ˈa.βlan]	‘speak’ 3PL-PRS	2	<h>	
hablar	[a.ˈβlaɾ]	‘to speak’	1	<h>	
hacer	[a.ˈseɾ]	‘to do/make’	3	<h>	<ce>
hambre	[ˈam.βɾe]	‘hunger’	1	<h>	
han	[ˈan]	(with part.) ‘have’ 3PL-PRS	1	<h>	
hasta	[ˈas.ta]	‘until’	1	<h>	
hay	[ˈai̯]	‘there is/are’	7	<h>	
hechos	[ˈe.t͡ʃos]	‘facts’	1	<h>	<ch>

hermanas	[eɾ.ˈma.nas]	‘sisters’ F.PL	1	<h>	
hermano	[eɾ.ˈma.no]	‘brother’ M.SG	1	<h>	
hermanos	[eɾ.ˈma.nos]	‘brothers’ M.PL	1	<h>	
hija	[ˈi.xa]	‘daughter’ F.SG	2	<h>	<j>
historia	[is.ˈto.ɾi̯a]	‘history’	2	<h>	
históricos	[is.ˈto.ɾi.kos]	‘historic(al)’ M.PL	1	<h>	<co>
(se) hizo	[ˈi.so]	‘became’ 3SG-PRS	1	<h> <z>	<z>
hola	[ˈo.la]	‘hello’	1	<h>	
hombre	[ˈom.βɾe]	‘man’	2	<h>	
hora	[ˈo.ɾa]	‘hour’	1	<h>	
hospital	[os.pi.ˈtal]	‘hospital’	1	<h>	<t>
hotel	[o.ˈtel]	‘hotel’	1	<h>	<t>
hoy	[ˈoi̯]	‘today’	1	<h>	
<j>					
ajedrez	[a.xe.ˈðɾes]	‘chess’	1	<j>	<z>
Alejandro	[a.le.ˈxan̪.dɾo]	(name) ‘Alexander’	1	<j>	
debajo	[de.ˈβa.xo]	‘below’	1	<j>	
dibuja	[di.ˈβu.xa]	‘draws’ 3SG-PRS	1	<j>	
dibujar	[di.βu.ˈxaɾ]	‘to draw’	1	<j>	
extranjeras	[eks.tɾaŋ.ˈxe.ɾas]	‘foreign’ F.PL	1	<j>	
extranjeros	[eks.tɾaŋ.ˈxe.ɾos]	‘foreign’ M.PL	1	<j>	
Javier	[xa.ˈβi̯eɾ]	(name) ‘Xavier’	1	<j>	
joven	[ˈxo.βen]	‘youth’	1	<j>	
joyas	[ˈxo.ʝas]	‘jewels’	1	<j>	
Juan	[ˈxu̯an]	(name) ‘John’	1	<j>	
Juárez	[ˈxu̯a.ɾes]	(name) ‘Juárez’	1	<j>	<z>
jueves	[ˈxu̯e.βes]	‘Thursday’	3	<j>	
Julia	[ˈxu.li̯a]	(name) ‘Julia’	1	<j>	
julio	[ˈxu.li̯o]	‘July’	1	<j>	

junio	[ˈxu.ni̯o]	‘June’	1	<j>	
juntos	[ˈxun̪.tos]	‘together’ M.PL	1	<j>	
lejos	[ˈle.xos]	‘far’	1	<j>	
mejor	[me.ˈxoɾ]	‘best’	1	<j>	
mujer	[mu.ˈxeɾ]	‘woman’	1	<j>	
trabaja	[tɾa.ˈβa.xa]	‘works’ 3SG-PRS	1	<j>	
trabajador	[tɾa.βa.xa.ˈðoɾ]	‘hardworking’ M.SG	1	<j>	
trabajan	[tɾa.ˈβa.xan]	‘work’ 3PL-PRS	1	<j>	
traje	[ˈtɾa.xe]	‘suit’	1	<j>	
viaja	[ˈbi̯a.xa]	‘travels’ 3SG-PRS	1	<j>	
viajes	[ˈbi̯a.xes]	‘trips’	1	<j>	
viejo	[ˈbi̯e.xo]	‘old’ M.SG	2	<j>	
<ll>					
allá	[aˈʝa]	‘there’	2	<ll>	
amarillos	[a.ma.ˈɾi.ʝos]	‘yellow’ M.PL	1	<ll>	
bellas	[ˈbe.ʝas]	‘pretty’ F.PL	2	<ll>	
ella	[ˈe.ʝa]	‘she’	8	<ll>	
ellos	[ˈe.ʝos]	‘they’ M.PL	3	<ll>	
llama	[ˈʝa.ma]	‘calls’ 3SG-PRS	1	<ll>	
llamas	[ˈʝa.mas]	‘call’ 2SG-PRS	1	<ll>	
llamo	[ˈʝa.mo]	‘call’ 1SG-PRS	2	<ll>	
llegas	[ˈʝe.ɣas]	‘arrive’ 2SG-PRS	1	<ll>	
llevarnos	[ʝe.ˈβaɾ.nos]	(with bien) ‘to get along well with each other’	1	<ll>	
llevas	[ˈʝe.βas]	‘wear’ 2SG-PRS	1	<ll>	
llorar	[ʝo.ˈɾaɾ]	‘to cry’	1	<ll>	
llueve	[ˈʝu̯e.βe]	‘rains’ 3SG-PRS	1	<ll>	
paella	[pa.ˈe.ʝa]	(name) ‘Paella’	2	<ll>	
pollo	[ˈpo.ʝo]	‘chicken’	1	<ll>	

sillas	[ˈsi.ɟas]	‘chairs’	1	<ll>	
<ñ>					
añade	[a.ˈɲa.ðe]	‘add’ 2SG-IMP	1	<ñ>	
año	[ˈa.ɲo]	‘year’	2	<ñ>	
años	[ˈa.ɲos]	‘years’	4	<ñ>	
baño	[ˈba.ɲo]	‘bath’; (with traje de) ‘bathing suit’	1	<ñ>	
brasileño	[bɾa.si.ˈle.ɲo]	‘Brazilian’ M.SG	1	<ñ>	
diseños	[di.ˈse.ɲos]	‘designs’	1	<ñ>	
enseñan	[en.ˈse.ɲan]	‘teach’ 3PL-PRS	1	<ñ>	
España	[es.ˈpa.ɲa]	‘Spain’	1	<ñ>	
español	[es.pa.ˈɲol]	‘Spanish’	3	<ñ>	
mañana	[ma.ˈɲa.na]	‘tomorrow’	3	<ñ>	
montañas	[mon̪.ˈta.ɲas]	‘mountains’	1	<ñ>	
montañoso	[mon̪.ta.ˈɲo.so]	‘mountainous’ M.SG	1	<ñ>	
pequeña	[pe.ˈke.ɲa]	‘small’ F.SG	1	<ñ>	
pequeño	[pe.ˈke.ɲo]	‘small’ M.SG	1	<ñ>	
puertorriqueño	[pu̯eɾ.to.ri.ˈke.ɲo]	‘Puerto Rican’ M.SG	1	<ñ>	
señor	[se.ˈɲoɾ]	‘sir/Mr.’	2	<ñ>	
sueño	[ˈsu̯e.ɲo]	‘dream’ (n.)	1	<ñ>	
<p>[110]					
aeropuerto	[a.e.ɾo.ˈpu̯eɾ.to]	‘airport’	1	<p>	
agrupó	[a.ɡɾu.ˈpo]	‘grouped’ 3SG-PRET	1	<p>	
antepago	[an̪.te.ˈpa.ɣo]	‘prepay’ 1SG-PRS	1	<p>	<go>
apaga	[a.ˈpa.ɣa]	‘turn off’ 2SG-IMP	1	<p>	<ga>
aparte	[a.ˈpaɾ.te]	‘apart’	1	<p>	
apuntes	[a.ˈpun̪.tes]	‘notes’	1	<p>	

[110] Count for tokens containing word-initial /p/ represents only those that were measured.

deportes	[de.ˈpoɾ.tes]	‘sports’	2	<p>	
equipaje	[e.ki.ˈpa.xe]	‘luggage’	1	<p>	<j>
europea	[eu̯.ɾo.ˈpe.a]	‘European’ F.SG	1	<p>	
mariposas	[ma.ɾi.ˈpo.sas]	‘butterflies’	1	<p>	
opino	[o.ˈpi.no]	‘to think’	1	<p>	
opuesta	[o.ˈpu̯es.ta]	‘opposite’	1	<p>	
papá	[pa.ˈpa]	‘father’	1	<p>	
papás	[pa.ˈpas]	‘fathers, parents’	1	<p>	
para	[ˈpa.ɾa]	‘for’	4	<p>	
pasa	[ˈpa.sa]	‘happens’ 3SG-PRS	1	<p>	
paso	[ˈpa.so]	‘pass’ 1SG-PRS	1	<p>	
paz	[ˈpas]	‘peace’	1	<p>	<z>
pero	[ˈpe.ɾo]	‘but’	3	<p>	
piensas	[ˈpi̯en.sas]	‘think’ 2SG-PRS	2	<p>	
pinta	[ˈpin̪.ta]	‘paints’ 3SG-PRS	1	<p>	
poco	[ˈpo.ko]	‘little’	1	<p>	<co>
por/por (qué)	[ˈpoɾ]	‘for’; (with qué) ‘why’	7	<p>	
porque	[ˈpoɾ.ke]	‘because’	3	<p>	
(se) prepara	[pɾe.ˈpa.ɾa]	‘(is) prepared’ 3SG-PRS	2	<p>	
pueblo	[ˈpu̯e.βlo]	‘town’	3	<p>	
puede	[ˈpu̯e.ðe]	‘can/is able to’ 3SG-PRS	1	<p>	
puedes	[ˈpu̯e.ðes]	‘can/are able to’ 2SG-PRS	2	<p>	
puerta	[ˈpu̯eɾ.ta]	‘door’	1	<p>	
Puerto (Rico)	[ˈpu̯eɾ.to]	‘port’ (with Rico) ‘Puerto Rico’	1	<p>	
pues	[ˈpu̯es]	‘well’	1	<p>	

repaso	[re.ˈpa.so]	‘review’ (n.)	1	<p>	
repita	[re.ˈpi.ta]	‘repeat’ 3SG-SBJV	1	<p>	
separan	[se.ˈpa.ɾan]	‘separate’ 3PL-PRS	1	<p>	
supones	[su.ˈpo.nes]	‘suppose’ 2SG-PRS	1	<p>	
supongo	[suˈpoŋ.go]	‘suppose’ 1SG-PRS	1	<p>	<go>
trepar	[tɾe.ˈpaɾ]	‘to climb’	1	<p>	
utopía	[u.to.ˈpi.a]	‘utopia’	1	<p>	
<t>[111]					
admitir	[að.mi.ˈtiɾ]	‘to admit’	1	<t>	
biblioteca	[bi.βli̯o.ˈte.ka]	‘library’	2	<t>	<ca>
debatir	[de.βa.ˈtiɾ]	‘to argue/debate’	1	<t>	
detalle	[de.ˈta.ʝe]	‘detail’ (n.)	1	<t>	<ll>
escritorio	[es.kɾi.ˈto.ɾi̯o]	‘desk’	1	<t>	
latino	[la.ˈti.no]	‘Latin’ M.SG	1	<t>	
literatura	[li.te.ɾa.ˈtu.ɾa]	‘literature’	1	<t>	
omitir	[o.mi.ˈtiɾ]	‘to omit’	1	<t>	
otoño	[o.ˈto.ɲo]	‘autumn’	1	<t>	<ñ>
patinas	[pa.ˈti.nas]	‘skate’ 2SG-PRS	1	<t>	
quitar	[ki.ˈtaɾ]	‘to remove’	1	<t>	
repetir	[re.pe.ˈtiɾ]	‘to repeat’	1	<t>	
setenta	[se.ˈten̪.ta]	‘seventy’	1	<t>	
situada	[si.ˈtu̯a.ða]	‘situated’ F.SG	1	<t>	
situado	[si.ˈtu̯a.ðo]	‘situated’ M.SG	1	<t>	
tal	[ˈtal]	(with qué) ‘how are [things]?’	2	<t>	
talla	[ˈta.ʝa]	‘size’	1	<t>	<ll>
tan	[ˈtan]	‘as/so’	1	<t>	

[111] Count for tokens containing word-initial /t/ represents only those that were measured.

tarde	[ˈtaɾ.ðe]	‘afternoon, evening’	2	<t>	
te	[ˈte]	‘you’	4	<t>	
techo	[ˈte.t͡ʃo]	‘roof’	1	<t>	<ch>
ten	[ˈten]	(with cuidado) ‘be (careful)’ 2SG-IMP	1	<t>	
tengo	[ˈteŋ.go]	‘have’ 1SG-PRS	4	<t>	<go>
tiempo	[ˈti̯em.po]	‘time/weather’	2	<t>	
tienda	[ˈti̯en̪.da]	‘store’	1	<t>	
tiene	[ˈti̯e.ne]	‘has’ 3SG-PRS	3	<t>	
tienen	[ˈti̯e.nen]	‘have’ 3PL-PRS	1	<t>	
tienes	[ˈti̯e.nes]	‘have’ 2SG-PRS	2	<t>	
tío	[ˈti.o]	‘uncle’	1	<t>	
tiritabas	[ti.ɾi.ˈta.βas]	‘shivered’ 2SG-IPFV	1	<t>	
toda	[ˈto.ða]	‘all’ F.SG	1	<t>	
todo	[ˈto.ðo]	‘all’ M.SG	1	<t>	
todos	[ˈto.ðos]	‘all’ M.PL	1	<t>	
toman	[ˈto.man]	‘take’ 3PL-PRS	1	<t>	
tomas	[ˈto.mas]	‘take’ 2SG-PRS	2	<t>	
tomo	[ˈto.mo]	‘take’ 1SG-PRS	3	<t>	
tratamos	[tɾa.ˈta.mos]	‘try’ 1PL-PRS	1	<t>	
tú	[ˈtu]	‘you’ (SG)	2	<t>	
tu	[ˈtu]	‘your’ (SG)	1	<t>	
visitar	[bi.si.ˈtaɾ]	‘to visit’	3	<t>	
visitarnos	[bi.si.ˈtaɾ.nos]	‘to visit us’	1	<t>	
visitaron	[bi.si.ˈta.ɾon]	‘visited’ 3PL-PRET	1	<t>	
visitaste	[bi.si.ˈtas.te]	‘visited’ 2SG-PRET	1	<t>	
<z>					
arroz	[a.ˈros]	‘rice’	1	<z>	
azafrán	[a.sa.ˈfɾan]	‘saffron’	1	<z>	
azteca	[as.ˈte.ka]	‘Aztec’ F.SG	1	<z>	<ca>

aztecas	[as.ˈte.kas]	‘Aztec’ F.PL	1	<z>	<ca>
azul	[a.ˈsul]	‘blue’	1	<z>	
cruza	[ˈkɾu.sa]	‘crosses’ 3SG-PRS	1	<z>	
diez	[ˈdi̯es]	‘ten’	3	<z>	
empieza	[em.ˈpi̯e.sa]	‘begins’ 3SG-PRS	2	<z>	
empiezan	[em.ˈpi̯e.san]	‘begin’ 3PL-PRS	1	<z>	
fortaleza	[foɾ.ta.ˈle.sa]	‘fortress’	1	<z>	
izquierda	[is.ˈki̯eɾ.ða]	‘left’ (n.)	1	<z>	
luz	[ˈlus]	‘light’	1	<z>	
Mazatlán	[ma.sat.ˈlan]	(name) ‘Mazatlán’	1	<z>	
pizarra	[pi.ˈsa.ra]	‘chalkboard’	1	<z>	
quizás	[ki.ˈsas]	‘maybe’	1	<z>	
razón	[ra.ˈson]	‘reason’ (n.)	1	<z>	
(se) realiza	[re.a.ˈli.sa]	‘is fulfilled’ 3SG-PRS	1	<z>	
(una) vez	[ˈbes]	(with una) ‘once’	2	<z>	
zapatos	[sa.ˈpa.tos]	‘shoes’ M.PL	2	<z>	<p>
Zaragosa	[sa.ɾa.ˈɣo.sa]	Zaragosa	1	<z>	<go>
zonas	[ˈso.nas]	‘zones’ F.PL	1	<z>	

APPENDIX C

Recording Instructions

Read the following sentences aloud.

When you finish reading all the sentences on a slide, click the arrow, arrow key, or mouse to move to the next slide.

When you have finished all the slides, you'll see: [END]; at this time, alert the investigator, who will stop the recording.

Then, the investigator will give you further instructions about your participation in the study and the next steps for the study.

APPENDIX D

Stimuli

(In order of appearance; adapted/adopted from *Vistas* (Blanco & Horwitz, 2001))

Slide 1: (Instructions – See Appendix C)

Slide 2:

¿Por cuántos años ha tocado la guitarra ella? ¿Dirige un coro?

Ella quiere que la profesora repita el repaso porque ella quiere tomar apuntes y corregir sus errores.

Hola, ¿qué tal? ¿Te recaes? ¿Por qué tiritabas?

Ella se llama Carmen. Ella habla chino y portugués brasileño.

Mucho gusto a conocerte. Igualmente.

Hasta mañana, señor Sánchez.

Acaba de admitir que hay cuatro chicos pero no hay ninguna chica.

Slide 3:

Imagina que son las cuatro y cuarto de la tarde.

Una de las joyas es la Pequeña Habana, donde viven muchos cubanoamericanos.

Cada junio desde mil novecientos cincuenta y uno, mucha gente de origen puertorriqueño celebra su orgullo latino con un desfile en la ciudad.

El espectáculo tiene carrozas y música. Mucha gente participa con frecuencia.

Allá hay diez residencias bellas. Escoge la mejor.

¿Cuándo tomas biología? Tomo biología los jueves a las diez y cuarto.

Alejandro viaja a Zaragosa mañana para practicar su español.

Slide 4:

¿Qué tal las clases? Tú tomas geografía y yo tomo historia y literatura. Tomo muchas clases.

Hay chicos en la [*sic*] clases de historia y sociología, casi un ochenta y cinco por ciento.

Y tú, joven, ¿cómo te llamas? Me llamo Javier Gómez y soy de San Juan, Puerto Rico.

¿Te gustan las computadoras, Javier? No me gustan, pero me gusta mucho más el arte y sobre todo me gusta dibujar.

Ellos y sus hermanas trabajan en el aeropuerto cada noche todos los jueves.

Ahora necesitamos hablar con frecuencia con el señor Francisco.

Slide 5:

Ellos enseñan física, química, y computación en el colegio en la costa de Costa Rica.

¿Por qué necesitas picar y debatir tanto con ella?

¿Cuántos hechos hay que hacer y repetir hoy?

¿Cómo llegas a la tienda cada tarde?

Acaso hay un libro azul y una mochila castaña en el escritorio allá. ¿Puedes recogerlos?

En la biblioteca, la pizarra está debajo del blanco techo con cien sillas.

¿La biblioteca está lejos de esta calle?

Slide 6:

No, está cerca. Gira a la derecha y está a la izquierda de la Facultad de Bellas Artes.

Las clases de español para estudiantes extranjeros empiezan en julio y terminan en agosto. La clase siguiente empieza en diciembre.

Mario y Julia hablan de los gitanos y toman clases de ciencias y lenguajes extranjeras juntos todos los martes y jueves.

La antigua Universidad de Salamanca está situada en la ciudad de Salamanca, que es famosa por sus gigantes edificios históricos.

La paella es uno de los platos más típicos de España. Siempre se prepara con arroz y azafrán.

Slide 7:

La paella valenciana se prepara con pollo y conejo.

La hija de mi tío es mi prima. Ella vive en Egipto y le gustan los girasoles y las mariposas.

Es una mujer que trabaja en un hospital para los muchachos.

Es un hombre que dibuja y pinta mucho sin omitir cualquier detalle. Sacó un premio.

¿Qué pasa con Inés y Javier? ¿Van a mantenerse aparte? ¿Se separan?

¡Qué alto es tu papá! ¿Acaso tiene alergias? Sacudo para quitar el polvo.

Sólo tengo un hermano mayor, Pablo. Sus papás visitaron cada año.

Slide 8:

Él se destacó porque es viejo pero es un hombre trabajador. Es muy capaz.

Cuando una puerta se cierra, otra se abre.

¿Recibe muchas cartas Usted? ¿Las agrupó por destino?

¿Eliges llorar cuando tienes sueño o trepar un árbol?

Usted tiene que visitarnos acá en ese país europea.

Ten cuidado con la pintura de la ribera opuesta.

¿Qué talla de chaqueta llevas?

Slide 9:

Tú y Ricardo tienen mucha hambre, pero yo tengo sed. Deben masticar más mientras bebo cacao.

Me llamo Armando y tengo setenta años pero no me considero viejo.

Paso mucho tiempo con mi hija porque es frágil.

Comparto mis cosas con mis hermanos. Nosotros tratamos de llevarnos bien y enfocarnos.

Opino que el novio de mi amiga es un poco gordo pero guapísimo.

¿Te gustaría visitar la ciudad?

Slide 10:

La lengua oficial del Ecuador es el español, pero también se hablan otras lenguas como quichua.

El sistema montañoso cruza y divide el Ecuador en dos.

Muchas personas visitan las islas Galápagos porque son un tesoro ecológico. Son famosas por sus tortugas gigantes.

Cada pueblo usa colores, figuras, y diseños diferentes.

Pues, en el parque acá hay una estatua de Benito Juárez. Está en el parque municipal, cerca del gimnasio.

Acabo de practicar el ciclismo cuando patinas en línea.

Slide 11:

Esta noche acaban de correr ellos, pero yo no tengo energía para practicar deportes y abanicar.

Tienes razón que debemos practicar deportes.

¿Deseas visitar Mazatlán o La Paz para tu cumpleaños?

¿A qué hora piensas volver? Quizás el partido empieza a las diez y cuarto.

¿Piensas que nuestro equipo puede ganar?

Ana estudia biología, y anoche se aplicó mucho en los estudios.

Supongo que puedes estudiar o jugar ajedrez.

Slide 12:

¿Qué acabas de hacer esta noche?

¿Qué tiempo va a hacer mañana?

¿Supones que llueve mucho en tu pueblo en otoño?

¡Apaga la luz! Antepago por la electricidad.

Cada cuatro años se realiza la Copa Mundial de Fútbol.

Argentina y Uruguay han ganado este campeonato más de una vez.

Los aztecas construyeron pirámides con templos religiosos acá.

Slide 13:

La ciudad de México está construida en el sitio de la capital azteca y muchos turistas quieren visitar sus ruinas.

Enrique tiene una cabaña en las montañas que parece como utopía.

¿Visitaste una agencia de viajes una vez el año pasado?

Nosotros estamos muy ocupados cuando necesitamos encontrar el equipaje.

La Cueva de los Tres Pueblos es una gigantesca bóveda, tan grande que toda la fortaleza del Morro.

El hotel está situado en un pequeño pueblo.

Slide 14:

En las zonas turísticas este porcentaje es mucho más alto.

El inglés es obligatorio para documentos.

Puerto Rico se hizo un estado libre asociado hace setenta años.

Necesitamos comprarnos un par de zapatos. Añade un traje de baño y los calcetines a la lista.

Acabo de comprarme zapatos amarillos acá.

[END]

APPENDIX E

Background Questionnaire

Please enter the code the researcher gave you here: _____________

Background

The questions in this section will ask for your demographic information.

1. Age (select one): 18-21 22-25 26-30 31-35 36-40 40+
2. What gender do you identify as?
3. Where are you from? If you moved a lot growing up, where did you live the longest?
4. What language/s did you learn first (i.e., what language/s did you speak before age 5)?
5. Have you ever visited a Spanish-speaking country? YES NO

   If yes, which country/countries?

   If yes, how long did you stay in each country?
6. What language/s have you studied other than Spanish? Please explain below; if you have studied no other languages, please write "none".
7. **How often** do you regularly encounter Spanish outside of class (not including online homework activities or the technology you have used during this study)?

   Not at all once a day several times a day once a week several times a week

   once a month several times a month
8. **How** do you typically encounter Spanish outside of class (not including online homework activities or the technology you have used during this study)? Please select all that apply.

   Not at all Hear it Read it Speak it Write it
9. Why do you want to study Spanish?

10. On a scale of 1-10, 1 being not at all and 10 being completely, how comfortable are you using iPad and/or iPhone apps?
11. On a scale of 1-10, 1 being not at all and 10 being completely, how comfortable are you using the VHL site (the online homework site for Spanish 101)?
12. On a scale of 1-10, 1 being not at all and 10 being completely, how comfortable are you using Blackboard?

(Submission confirmation message):

Thank you for your participation in this study. Please close this window and check back in with the investigator.

APPENDIX F

Weekly Check-In Survey for EG1 and EG2

Please enter the code the researcher gave you here: _____________

Use this survey to log your participation each week.

1. Enter the approximate amount of time you spent interacting with the technology this week: ________ .
2. Which activities did you complete this week? Please list them by name (and by chapter if you're using the VHL site).
3. Based on your interaction with the technology, list some words/phrases that you think you pronounced well.
4. Based on your interaction with the technology, list some words/phrases that gave you trouble.
5. Based on the activities you completed this week, what do you think you need to practice more?

APPENDIX G

Technology Questionnaire

Please enter the code the researcher gave you here: _____________

Brief Technology Evaluation

The questions in this section will ask you for your opinion about the technology you used during the study. Please answer honestly.

1. What did you like about the technology you used during this study?
2. What did you dislike about the technology you used during this study?
3. Do you think using the technology improved your performance in your Spanish class? Please explain.
4. Would you recommend the technology you used during this study to someone interested in learning Spanish? Why or why not?
5. Is there anything else you'd like to mention about the technology you used using this study and/or your experiences using it?

(Submission confirmation message):

Thank you for your participation in this study. Please close this window and check back in with the investigator.

APPENDIX H

Suggested Activity List – EG1 (Spanish SOLO App)

Week #:	Chapter/s Covered:	Suggested Activities:
2	Unit 1	• Get Started • Numbers and Time: Numbers • "To Be": Subject Pronouns • "To Be": Gender & Agreement • "To Be": The Verb SER • Greetings: The Alphabet • Greetings: Familiar Introductions • Greetings: Formal Introductions
3	Unit 1	• Numbers and Time: Everyday Numbers • Numbers and Time: Numbers to 100 • "To Be": Using SER with Adjectives • "To Be": Practice: SER with Adjectives • Essential Verbs, -AR: Regular –AR Verbs • Essential Verbs, -AR: Practice: Regular –AR Verbs
4	Unit 2	• Essential Verbs, -AR: The Verb ESTAR • Essential Verbs, -AR: Practice: The Verb ESTAR • Essential Verbs, -AR: SER vs ESTAR • Essential Verbs, -ER & -IR: Regular –ER & -IR Verbs • Essential Verbs, -ER & -IR: The Verb TENER
5	Unit 2	• Essential Verbs, -AR: Regular –AR Verbs • Essential Verbs, -AR: Practice: Regular –AR Verbs • Essential Verbs, -AR: The Verb ESTAR • Essential Verbs, -AR: Practice: The Verb ESTAR
6	Unit 2/Unit 3	• Review suggested activities from Unit 2, or start activities listed for Unit 3 (Week 7).
7	Unit 3	• Essential Verbs, -ER & -IR: Days & Months • Essential Verbs, -ER & -IR: Practice: Days & Months • Numbers & Time: Telling Time • "To Be": The Verb SER
8	Unit 3	• Essential Verbs, -ER & -IR: The Verb TENER • "To Be": Gender & Agreement • Essential Verbs, -AR: Regular –AR Verbs
9	Unit 4	• Review suggested activities from Unit 3, and/or start activities listed for Unit 4.
10	Unit 4	• "To Be": Using SER with Adjectives • "To Be": Practice: SER with Adjectives • Essential Verbs, -AR: Practice: Regular –AR Verbs
11	Unit 4/Unit 5	• Review suggested activities from Unit 4, or start activities listed for Unit 5 (Week 12).

12	Unit 5	• Numbers and Time: Numbers • Essential Verbs, -ER & -IR: Days & Months • Essential Verbs, -ER & -IR: Practice: Days & Months
13	Unit 5	• Family & Friends: Family* • Family & Friends: Modern Family* • Essential Verbs, -AR: The Verb ESTAR • Essential Verbs, -AR: Practice: The Verb ESTAR
14	Unit 6	• Review suggested activities from Unit 5, or start activities from Unit 6 (Week 15).
15	Unit 6	• "To Be": Using SER with Adjectives • "To Be": Practice: SER with Adjectives • "To Be": Gender & Agreement
16	Unit 6	• Essential Verbs, -ER & -IR: The Verb HACER • "To Be": The Verb SER • Essential Verbs, -AR: SER vs ESTAR

* This activity is part of Beginner II; you need to change tracks to access it.

Note: This is not a comprehensive list of activities in the app, especially as they periodically add new activities. This is a list of activities that are most closely related to what will be covered in class at that time. What you choose to complete is up to you, as long as you complete the minimum of 2 hours per week.

APPENDIX I

Suggested Activity List – EG2 (Online Textbook Companion Activities)

Week #:	Chapter/s Covered:	Suggested Activities:
2	Unit 1	Pronunciación Tab: • Lección 1A: The Spanish Alphabet • Presentation + 2 Online Workbook Activities Imágenes y Palabras Tab: • 2 – Es una isla Yo Soy de Madrid Tab: • Ciudades • 13 - ¿De dónde eres? • 17 – Conversación Benicio del Toro Tab: • [c] • [g] • 22 – Es una ciudad
3	Unit 1	Pronunciación Tab: • Lección 1B: Intonation • Presentation + 1 Online Workbook Activity Hablamos Tab: • 3 – Él es alemán Profesiones y Trabajos Tab: • 16 – Yo soy estudiante Joan, un Catalán Políglota Tab: • Entonación
4	Unit 2	Pronunciación Tab: • Lección 2A: Spanish [b] and [v] • Presentation + 3 Online Workbook Activities ¡Viva Guatemala! Tab: • 10 – Es interesante Cerca de Casa Tab: • 13 – Mi casa tiene… • 21 – Responder Juan Tomás, un Taxista de Panamá Tab: • [b] y [v]
5	Unit 2	Pronunciación Tab: • Lección 2B: Word Stress and Accent Marks • Presentation + 3 Online Workbook Activities Café con Leche Tab: • 2 – El desayuno • 9 – Desayunar Rosa, Camarera del Bar El Rey de la Tapa Tab: • Sílabas acentuadas

5	Unit 2 (cont.)	• 21 – Cuántas veces
6	Unit 2/Unit 3	Review activities from Units 1 and 2
7	Unit 3	Pronunciación Tab: • Lección 3A: Spanish Vowels • Presentation + 3 Online Workbook Activities Agenda Semanal Tab: • 17 – Los planes de Carlos Pablo Linares, Dos Ritmos de Vida Tab: • [ch], [ll], y [ñ] • Pronunciación
8	Unit 3	Pronunciación Tab: • Lección 3B: The Letters [h], [j], and [g] • Presentation + 3 Online Workbook Activities Prendas de Moda Tab: • 13 – ¿De qué color? Olga Piedrahita, La Moda en Colombia Tab: • [j] y [g]
9	Unit 4	Pronunciación Tab: • Lección 4A: The Consonant [r] • Presentation + 2 Online Workbook Activities Hacer las Compras Tab: • 15 – Las cantidades • 22 - ¿Qué les doy? Paco, Amigo del Rastro Tab: • [r] y [rr]
10	Unit 4	Pronunciación Tab: • Lección 4B: C Before a Consonant, and Q • Presentation + 2 Online Workbook Activities Apartamento de Alquiler Tab: • 2 – Escoger • 10 - ¿Qué hay? Carmen Abreu, Fundadora de Caribe Propiedades Tab: • [c], [qu], y [k]
11	Unit 4/Unit 5	Review activities from Units 3 and 4
12	Unit 5	Pronunciación Tab: • Lección 5A: Diphthongs and Linking • Presentation + 3 Online Workbook Activities ¡Felicidades, Manuela! Tab: • 8 – En mil novecientos Diario de Viaje Tab: • 18 - ¿Cuándo fue la última vez? Salma Hayek, Más que una Cara Bonita Tab: • [b] y [p]

13	Unit 5	Pronunciación Tab: • Lección 5B: [ch] and [p] • Presentation + 3 Online Workbook Activities Fotos de Familia Tab: • 10 – Preguntas Mis Amigos, Mi Otra Familia Tab: • 21 - ¿Qué hizo el año pasado? Juanita y Bernardo, Dos Abuelos Muy Activos Tab: • [c] y [z]
14	Unit 6	Review activities from Units 1-5
15	Unit 6	Pronunciación Tab: • Lección 6A: The Consonants [d] and [t] • Presentation + 3 Online Workbook Activities Tiempo Libre Tab: • 9 – Preguntas Un Día Perfecto Tab: • 20 – Consejos Domingo Buendía, Maestro de Yoga Tab: • [p] y [t]
16	Unit 6	Pronunciación Tab: • Lección 6B: [m] and [n] • Presentation + 3 Online Workbook Activities Planes de Viaje Tab: • 9 – Planes Reservas Tab: • 18 – Me gustaría, quería Julia y Felipe, Amor a Distancia Tab: • [d] y [j] finales • Repetir

Note: These are the suggested activities to complete each week according to the Course Calendar for Spanish 101. However, you are free to interact with the technology by completing whichever of the pronunciation activities you wish, as long as you complete the minimum of 2 hours per week.

Technology Specifications: To complete these activities, your computer must have a working microphone and speakers.

THE UNIVERSITY OF
ALABAMA®
Office of the Vice President for
Research & Economic Development
Office for Research Compliance

September 1, 2020

Stacey Jacobson, MA
Department of Modern Languages and Classics
College of Arts & Sciences
The University of Alabama
Box 870246

Re: IRB # 17-OR-287-R3 "Choose Your Destiny: Pronunciation, Perception, and Digital Gaming in Second Language Acquisition"

Dear Ms. Jacobson:

The University of Alabama Institutional Review Board has granted approval for your renewal application.
Your renewal application has been given expedited approval according to 45 CFR part 46. Approval has been given under expedited review category 7 as outlined below:

(7) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

The approval for your application will lapse on August 31, 2021. If your research will continue beyond this date, please submit a continuing review to the IRB as required by University policy before the lapse. Please note, any modifications made in research design, methodology, or procedures must be submitted to and approved by the IRB before implementation. Please submit a final report form when the study is complete.

Good luck with your research.