THE EFFECTS OF PRONUNCIATION INSTRUCTION ON L2 PRODUCTION AND L2 PERCEPTION IN SPANISH: A COMPARATIVE ANALYSIS

by

Heather M. Offerman

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Languages and Cultures

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL

STATEMENT OF COMMITTEE APPROVAL

Dr. Daniel J. Olson, Chair

School of Languages & Cultures

Dr. Olga Dmitrieva

School of Languages & Cultures

Dr. Jessica Sturm

School of Languages & Cultures

Dr. Mariko M. Wei

School of Languages & Cultures

Approved by:

Dr. Jennifer M. William

This dissertation is dedicated to my family, close friends, advisor, and advisory committee – without your support and continual encouragement, I would not have succeeded.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Olson, for all the help he has provided me over the years. I started as a nervous and timid master's student, and I constantly sought help from you, even before you were my advisor. Once I finished my M.A. and decided to take an abrupt break during the PhD, you encouraged me to take that break and discover if I truly wanted to be in academia. Upon returning from my two-year absence, you made sure that I was able to pick up where I had left off, then made sure that my transition was as smooth as possible. Once I chose my dissertation topic, you provided nothing but support, and you spent countless hours revising chapters, giving feedback, and helping me while I agonized over data coding and data analysis. Thank you for always cheering me on, pushing me to do my best, and for molding me into the researcher that I am today. I also appreciate your kindness and humility; I hope to embody all the same qualities as you once I begin to advise my own students.

To Dr. Sturm, thank you for your constant support and treating me like an equal. This gave me so much confidence, and I am grateful that you embraced me as a colleague even though I did not have the title to be treated as such. To Dr. Dmitrieva, your expertise and input was always on point and vital to the research I was attempting to do – thank you for your graciousness, interest in my study, and your invaluable feedback. To Dr. Wei, I am so thankful to have had you as a committee member, not only because of your talent in SLA and study design, but also because of how kind and encouraging you were along the way; it was always clear that you wanted me to succeed. To my extremely reliable and very hard-working R.A., Ellen Deemer, this dissertation would have been long delayed had you not been so consistent and dependable. I cannot express how much I appreciate the work and time you contributed to this dissertation.

Thank you to my sister, Kristin, for being such a huge support in my graduate career. I will always cherish our time together. Thank you to my niece and nephew, Alaina and Archer, for the sweetness and joy you brought in my life, which helped me in staying motivated to press on, even during the worst of days. To my brother, Ryan, I loved the time we were able to spend together in not living too far from each other – you always listened and reminded me that life is too short to worry so much about things out of my control. To my parents, Dad and Kristy, you were a tremendous support to me throughout my whole graduate career. Thank you for always supporting me and reminding me of the end goal, while always believing that I could do it. Thank you to my

mom – I know I would not have made it to where I have today without you being an example of hard work and dedication. Thank you for believing in me when I was young, and I wish you could see how far I've come; thank you for believing I had the intelligence and strength to achieve great things, despite what the teachers said.

To some dear friends and colleagues that helped me make it to the finish line: Boris, Katie, Kathryn, Maria, Tatiana, Joni, Joyce, Daniel, Miluska, Débora, Filza, Jaya, Carlos, Jessica, Sydney, Taryn, Madison, Angie, Laura, Amy, Tyler, Ignacio, Alex, Chuck, Dan, Beth, Noah, Thomas, Destinee, Chuck, Tim, Alba, Andi, Bryan, and Josh. I'm sure there are others whose names I missed, but thank you all for the emotional support, professional support, care, and comradery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Having historically received less attention than other linguistic structures (Derwing & Munro, 2005), second language (L2) pronunciation instruction represents an emergent area of research in the field (Thomson & Derwing, 2015). While several methods have been shown to be effective for improving L2 segmental production, including explicit instruction (Aliaga-García & Mora, 2009; Lord, 2010; Saito & Lyster, 2012) and inductive visual feedback instruction (Offerman & Olson, 2016; Olson, 2014b; Olson, 2019), there is a notable lack of empirically-based research comparing approaches (Derwing & Munro, 2015; Lee et al., 2015). Moreover, research has largely ignored the effects of instruction on L2 perception, due in part to the tacit assumption that L2 perception precedes L2 production (Levy & Law, 2010). Responding to these gaps, this study provides a large-scale comparative analysis of three types of pronunciation instruction (explicit instruction [EI], visual feedback [VF], and a combination instruction [CI] approach) on L2 segmental production and perception in Spanish. Production-oriented analyses focus on the change in voice onset time (VOT), and perceptual analyses focus on an L2 discrimination task (AXB task) and a nativeness judgment task (Liker-scale rating task). Differences in VOT for the stops /p,t,k/ in word-initial position exist for English (long-lag VOT) and Spanish (short-lag VOT) (Lisker & Abramson, 1964), causing notable accentedness for English-speaking L2 learners of Spanish (Lord, 2005), thus serving as the basis for L2 learner performance. Considering results for the production portion of the study, all three experimental groups were found to outperform the control (CO) group, along with each experimental group significantly improving individually from pretest to posttest. For the perception portion, participants did not display difficultly in discriminating between long-lag and short-lag productions at the pretest, and as such, showed no improvement following instruction. In contrast, results from the nativeness judgment task showed that participants were not able to categorize sounds as native-like (Spanish) or non-native-like (English), and significant improvement following training was found only for the CI group. Additionally, previous L2 perception theories largely focus on category formation and discrimination of sounds, such as the SLM (Flege, 1987), the PAM (Best, 1994), and the PAM-L2 (Best & Tyler, 2007), while this study considers perception as it applies to both discrimination and the social categorization of sounds in the L1 and L2. For the production and perception portions, the CI group largely outperformed all groups. It is proposed that the combination of two different

modalities, auditory and visual (Baran-Łucarz, 2012), provides learners more resources for noticing (Schmidt, 1990) differences between their L1 and L2 for production and perception purposes. Moreover, the CI treatment is proposed to be most beneficial for teaching L2 pronunciation, although more research is to be done with comparative analyses for different segments in the L2, as it has been previously noted that not all pronunciation treatments are equally beneficial to L2 learners for different segmental features (Ruellot, 2011). This study adds to theoretical understanding of L2 phonetic acquisition, in both production and perception, while empirically testing pedagogical approaches in a classroom setting.

CHAPTER 1. INTRODUCTION

In recent years, much discussion has developed regarding the current state of pronunciation instruction for second language (L2) learners, concluding that, based on current pedagogical theory and practice, pronunciation has received considerably less attention than other aspects of L2 teaching (Derwing & Munro, 2005). As the body of research has developed over the last three decades, two broad approaches have emerged. The first consists of explicit instruction, including the use of IPA symbols (Lord, 2010; Miller, 2012), explanation of place and manner of articulation (Elliott, 1997), use of repetition (Yoshida & Fukada, 2014), as well as implementation of corrective feedback and focus-on-form instructional methods (Saito & Lyster, 2012). A second type of instruction that has recently been examined is the use of visual feedback, utilizing speech analysis software to allow students to examine their own productions and compare them (visually) with native speaker (NS) productions (Motohashi-Saigo & Hardison, 2009; Offerman & Olson, 2016; Olson, 2019). While both of these methods have been shown to be broadly effective (deBot, 1983; Elliott, 1997; Lord, 2005; Lord, 2010, Offerman & Olson, 2016; Saito & Lyster, 2012), there have yet to be any systematic studies that compare the effectiveness of the two approaches (Derwing & Munro, 2015; Lee et al., 2015). Moreover, these approaches are often considered in isolation, with some visual feedback studies eschewing any form of explicit instruction (Offerman & Olson, 2016). As such, any comprehensive comparative approach should also consider coupling the two methods, mixing both explicit pronunciation instruction and visual feedback treatments.

As the strategies and methods for L2 pronunciation instruction have focused nearly exclusively on production-oriented paradigms, it is also important to consider how L2 perception is influenced, if at all, by these types of training. There exists the common belief that L2 perception precedes L2 production in L2 acquisition (Levy & Law, 2010), but some L2 pronunciation research posits that there is not clear evidence that perception necessarily precedes production (Flege et al., 1997; Zampini, 1998), along with some data indicating that perception may not precede production for certain phonological categories for L2 learners (Zampini, 1998).

The present study seeks to address these gaps by considering the impact of three different types of phonetic training on L2 production and perception: explicit instruction, visual feedback, and a combination of the two methods. Specifically, three broad questions are addressed. First, does each type of treatment serve to improve L2 *production*, and within a comparative framework,

do any of the three treatments offer an advantage? Second, does each type of treatment serve to improve L2 *perception*, and within a comparative framework, do any of the three treatments offer an advantage? And finally, what is the relationship between L2 production and L2 perception?

To answer these questions, a series of pronunciation trainings were conducted with four groups of L1 English intermediate-level learners of Spanish (L2). Each of the groups received one type of training: explicit instruction, visual feedback, a combination instruction approach, or a cultural activity of similar duration (i.e., control group). Training focused on the word-initial voiceless stops, /p,t,k/, leveraging the cross-linguistic differences in voice onset time between English (long-lag) and Spanish (short-lag), as there are notable differences in production by native English speakers producing these segments in their L2 Spanish (Lord, 2005). The over-arching design relies on a pretest, three instructional treatments, and a posttest, with participants performing four production-oriented tasks, ranging from more controlled productions to more spontaneous speech. For the perception-oriented tasks, all participants in each group also took part in two types of perceptual tests preceding all production treatments, as well as after all production treatments.

Beyond responding to the call in the literature for a systematic comparative analysis of different types of pronunciation training (Derwing & Munro, 2015; Lee et al., 2015), as well as further addressing the question of whether or not L2 perception precedes L2 production, the current study adds to ongoing discussions regarding the acquisition of L2 pronunciation and the links between adult L2 production and L2 perception.

CHAPTER 2. REVIEW OF THE LITERATURE

This chapter details current production-oriented pronunciation teaching methods, perception theories and testing methods, as well as the specific differences in VOT values for English and Spanish as they relate to voiceless plosives /p,t,k/ in word-initial position. Section 2.1 details the history of L2 pronunciation teaching methods, intelligibility, comprehensibility and the importance of pronunciation teaching, and student desire for pronunciation teaching. Section 2.2 discusses current methods used for L2 pronunciation teaching focusing on production, with Section 2.3 elaborating on three relevant L2 perceptual theories, concluding with a discussion of the relationship between L2 perception and L2 production. Section 2.4 introduces the differences in VOT values between English and Spanish, along with a closer look at the distribution of short-lag and long-lag voiceless stops in Spanish and English. Section 2.5 concludes this chapter and presents the research questions and hypotheses as they relate to gaps found in the literature.

2.1 Teaching L2 Pronunciation: History, Issues & Current Materials

This section details the history of L2 pronunciation teaching and current practices. Although pronunciation instruction is a growing field, it is still lacking in concrete methods and empirical evidence that conveys which methods are best for implementing into classroom instruction. The history of pronunciation teaching is detailed, along with several notable gaps in the extant research, namely empirically-tested instructional methods and a comparative analysis of varied methods. Concluding this section, gaps in L2 pronunciation research are also addressed.

2.1.1 Theory & Background of L2 Pronunciation Teaching

The role of L2 pronunciation has evolved considerably over the last several decades regarding teaching practices (Oxford et al., 1989; Saville-Troike, 1973). Beginning with the Audio-Lingual Method (ALM), L2 learners were guided to focus specifically on pronunciation by means of drills, with little attention being given to fluency, meaning, and communicative skills (for a review, see Isaacs, 2009). In response to this form-focused approach, a resulting shift in priorities towards communicative skills lead to the Communicative Teaching Method (CTM)

becoming the primary form of L2 instruction (Derwing & Munro, 2005; Isaacs, 2009; Saville-Troike, 1973). CTM has been described as a 'naturalistic' approach (Isaacs, 2009), meaning that is does not focus on teaching through explicit methods, but rather having learners develop L2 skills through input and meaning-focused activities, with the overall goal of communication and less focus on forms. Once the communicative method became the favored method of L2 instruction, focusing more on meaning and fluency (Oxford et al., 1989), L2 pronunciation was presumed to be learned more implicitly (Arteaga, 2000; Isaacs, 2009).

With the decline of the ALM after its peak in the 1960s (Pennycock, 1989) pronunciation teaching became less popular. Yet over the last three decades many researchers have argued that pronunciation instruction is essential to L2 learners' language development (Derwing & Munro, 2005; Elliott, 1997; Levis & Grant, 2003; Simões, 1996). The perceived foreign accents of L2 learners by native speakers (NSs) (Flege & Bohn, 1989; Piske et al., 2001) has led to an argument that pronunciation is a crucial part of intelligibility (Derwing & Munro, 1997; Derwing & Munro, 2009; Derwing et al., 1998). More specifically, researchers have suggested that pronunciation instruction is a necessary component for L2 learners to produce intelligible, comprehensible, and relatively un-accented speech (Derwing & Munro, 1997; Derwing & Munro, 2005; Derwing & Munro, 2009; Derwing et al., 1998; Levis, 2005; Levis & Grant, 2003; Lord, 2010; Munro & Derwing, 1999; Simões, 1996).

2.1.2 Intelligibility, Comprehensibility & Accentedness in L2 Pronunciation

Derwing et al. (1998) define intelligibility as "how much of an utterance a listener successfully processes" (p.396). Comprehensibility refers to the level of ease the interlocutor experiences when listening to L2 learner speech (Derwing et al., 1998). Simões (1996) explains further that there are multiple words containing minimal pairs that L2 speakers may confuse in their production, changing the meaning of the intended production. Instances with segmental confusion may cause further or more intelligibility and comprehensibility costs on the interlocutor. In other words, intelligibility refers to the overall message of string of L2 speech being understood by the interlocutor; comprehensibility differs in that it refers to the ease with which these strings of speech are understood by interlocutors. Therefore, while L2 speech may be intelligible, it may not be easily comprehensible, causing more time costs in processing L2 speech by the NS interlocutor.

As the importance of maintaining intelligible and comprehensible communication throughout speech has been discussed as the primary goal, the aspect of accentedness should also be considered when looking at NNS speech. As defined by Derwing & Munro (2009), accentedness is defined as "how different a pattern of speech sounds to a local variety" (p. 478). Although accentedness does not always indicate degree of intelligibility or comprehensibility for a NS listener (Derwing & Munro, 1997; Derwing & Munro, 2009), it can influence NS perceptions about the NNS, with accented speech resulting in negative judgments of the linguistic abilities and intelligence of L2 learners (Derwing & Munro, 2009) [1]. While some may not consider this to be the main goal of pronunciation teaching, it should be considered along with intelligibility and comprehensibility when developing materials and conducting pronunciation instruction.

2.1.3 Instructional Materials for L2 Pronunciation

Although the above literature shows that pronunciation instruction is crucial to L2 development, a number of researchers have noted a general lack of instructional material provided or implemented into L2 classrooms (Arteaga, 2000; Carbó et al., 2003; Derwing & Munro, 2005; Foote et al., 2011; Gilakjani, 2011; MacDonald, 2002; Morin, 2007; Offerman & Olson, 2016; Olson, 2014a). In line with this finding of a lack of materials, Derwing and Munro (2015) also found that methods that have currently implemented L2 pronunciation teaching have not undergone rigorous, empirical validation. As such, they suggest that in order to provide sound, L2 pronunciation teaching methods, methods should be empirically tested and driven by quantifiable improvements in learner productions (Derwing & Munro, 2015).

Arteaga (2000) specifically notes that there is a lack of materials in L2 Spanish textbooks at the beginner level, and if they are present, there is no concrete manner of instruction provided for L2 teachers. Morin (2007) and MacDonald (2002) expand on this idea by arguing that even in a field as large as English as a Second Language (ESL), there are some materials being developed, but there is no explicit guidance for teachers on how to use provided materials or when to

[1] VOT production by English-speaking learners of Spanish, the focus of the current study, is more of a concern for accentedness than intelligibility (Lord, 2005). However, as this feature is viewed as accented speech when produced by NSs of English (Hammond, 2001; Lord, 2005), it is possible that it may play a role in comprehensibility. It is worth noting that VOT does play a role in intelligibility in other language pairings (Pater, 2003). Moreover, VOT is represented an ideal test case in the current study for a variety of reasons, among them being the demonstrated success of the visual feedback training.

implement them into the curriculum. MacDonald (2002) also found that many teachers believed that the curriculum they had been provided with did not place much, if any, importance on pronunciation instruction (p. 6).

In addition to these findings, Olson (2014a) also found in a questionnaire distributed to L2 instructors that most of these instructors tend to approach L2 pronunciation instruction in a very ad-hoc manner. In other words, L2 instructor responses in this study indicate that they are unsure of when or how to teach pronunciation, mostly attempting to teach pronunciation when L2 learners' production errors are obvious to the instructor. It was found that these instructors focused an average of less than eight minutes total per week on pronunciation, adding that they believed that they spent much less time than necessary on teaching pronunciation (i.e. it was thought that eight minutes is not a sufficient amount of time per week). These instructors also indicated that a significant portion of their lack of comfort with teaching pronunciation is due to the lack of materials found in current curriculum provided to them in language textbooks.

Regarding learner thoughts towards pronunciation instruction, it has also been noted by several researchers that L2 learners want to receive pronunciation instruction (Elliott, 1995; Levis & Grant, 2003), and they also view pronunciation instruction as important to their learning of the target L2 (Huensch, 2017; Lord, 2008; Offerman, 2014; Olson, 2014b; Sturm et al., 2019). For example, Lord (2008) found that participants valued the pronunciation gains made during a semester, and that they also felt it was important for future L2 learners to take part in similar pronunciation activities to become more aware of one's own abilities, as well as how to improve these abilities. Sturm et al. (2019) additionally found that L2 learners felt that pronunciation instruction was important for their L2 learning, and that they even wanted their instructors to correct incorrect L2 productions.

Despite the lack of materials and implementation of pronunciation instruction, there has been a growing number of studies in the last several decades dedicated to researching and developing L2 pronunciation instruction (Aliaga-García & Mora, 2009; Derwing & Munro, 2009; Diaz-Campos, 2004; deBot, 1983; Foote et al., 2011; Gilakjani, 2011; Gonzalez-Bueno, 1997; Kissling, 2013; Lord, 2005; Miller, 2012; Morin, 2007; Offerman & Olson, 2016; Olson, 2014b; Rajab, 2013; Saito & Lyster, 2012; Saito & Lyster, 2012b; Sturm, 2013a; Wang & Munro, 2004) (for review, see Thomson & Derwing, 2015). As a number of different studies and approaches for pronunciation teaching have been developed and detailed, there has been a recent call to move

towards a comparative analysis of the different approaches, with the goal of maximizing the benefits for L2 learners (Lee et al., 2015). As Derwing and Munro (2015) point out, there exists a wide scope of instruction for L2 pronunciation, also stating that "A true classroom study over a period of several weeks comparing different methods of teaching several aspects of pronunciation…would make a significant new contribution to the field" (p. 92).

The following section details different methods of teaching L2 pronunciation. The first subsection reviews types of explicit pronunciation instruction, and the second subsection details various visual feedback studies. Both subsections primarily focus on methods for teaching segmentals rather than suprasegmentals, as the following study will focus on segmentals. These two types of L2 pronunciation instruction have been selected since they appear to be the most current and popular forms of research for L2 pronunciation (Derwing & Munro, 2015).

2.2 Methods for Teaching L2 Pronunciation

This subsection reviews current pronunciation teaching practices in the field and concludes by proposing three different types of instruction to be included the current comparative analysis. Addressing the literature for L2 pronunciation instruction, this section is divided into four parts: explicit pronunciation instruction (Section 2.2.1); visual feedback (Section 2.2.2); a combined explicit instruction and visual feedback approach (Section 2.2.3); production gaps (Section 2.2.4).

2.2.1 Explicit Pronunciation Instruction

Within the field of second language pronunciation instruction, many researchers have proposed that second language (L2) instruction should include some form of explicit pronunciation instruction (Derwing & Munro, 1997; Derwing & Munro, 2005; Derwing & Munro, 2009; Derwing et al., 1998; Elliott, 1997; Levis 2005; Levis & Grant, 2003; Miller, 2012; Munro & Derwing, 1999; Saito & Lyster, 2012; Simões, 1996; Sturm, 2013a; Yoshida & Fukada, 2014). Adapted from Ellis (2004), explicit pronunciation can be defined as the 'conscious awareness' of one's own pronunciation; therefore, explicit pronunciation instruction can be broadly defined as drawing L2 learners' attention to specific phonetic features in their L2, creating an environment in which learners must consciously focus on these features, as well as receiving direct explanations of how phonetic features in the L2 differ from the L1. More specifically, several common types of

explicit training have been employed: (1) incorporation of the International Phonetic Alphabet (IPA) (e.g., Miller, 2012); (2) explanation of articulatory features (e.g., place or manner) to produce segments in the L2 (e.g., Elliott, 1997); (3) repetition (e.g., Yoshida & Fukada, 2014); and (4) corrective feedback instruction for pronunciation teaching (e.g., Saito & Lyster, 2012).

Although many agree that there should be some form of explicit pronunciation teaching implemented, there are various means through which researchers have attempted to incorporate explicit training into their experimental designs. For many studies which focus on segmental features, phonetic training through a phonetics course has been one of the primary situations through which pronunciation has been taught (Lord, 2005; Lord, 2010; Rajab, 2013; Saalfeld, 2011; Sturm, 2013a). Within these phonetics courses, the most common type of training includes incorporating the IPA into the coursework. For example, instead of having students focus on graphemes, researchers have had students become familiar with the IPA symbols that represent sounds in the target language. As an example of this type of approach, in Lord (2005), participants not only learned the IPA, but they were also instructed how to associate sounds with place and manner of articulation by demonstration.

In other studies, researchers have primarily focused on instructing participants on the place and manner of articulation of problematic sounds in the target language for L1 speakers (Aliaga-García & Mora, 2009; Elliott, 1995; Elliott, 1997; Kissling, 2013; Saito, 2011). With this type of explicit pronunciation training, researchers have also had participants learn grapheme to phoneme correspondence; this is to say, researchers have attempted to aid L2 learners in disassociating graphemes with sounds in their L1, while instead associating these graphemes with the phonetic realizations in their L2, along with modeling the articulation and sounds for L2 learners.

In past methods of teaching pronunciation, such as the ALM, repetitions comprised a large part of pronunciation instruction prior to the development of the CTM (Saville-Troike, 1973). However, with the current practices principally incorporating CTM into L2 teaching, most methods of instruction involving pronunciation drills have been set aside, being regarded as a hindrance to L2 learning (Oxford et al., 1989; Wang & VanPatten, 2003). As a result, drills have tended to be frequently avoided in current practices in order to not regress back to an ALM-type of language teaching style. Although it was thought that drills and repetitions would be found not to be beneficial to L2 learners' language development (Trofimovich & Gatbonton, 2006), some researchers still claim that they have found repetitions and drills to aid L2 learners in improving

their pronunciation in more recent studies (Mizuno, 2007; Şenel, 2006; Trofimovich & Gatbonton, 2006; Yoshida & Fukada, 2014). For example, Trofimovich and Gatbonton (2006) found in one experiment that L2 learners of Spanish significantly improved in their productions of tokens after repeating NS models of these tokens. As such, repetitions appear to be valuable in aiding L2 learners in their pronunciation development (Trofimovich & Gatbonton, 2006).

In a more recent example, Yoshida & Fukada (2014) also argue that drills are beneficial to L2 learners in pronunciation development. In their study, L2 learners of Japanese were made to participate in an online platform in which they viewed videos of multiple productions recorded by NSs of Japanese. After viewing each video, L2 learners repeated and recorded their own productions, which allowed for L2 learners to have a type of modeling provided for them as they repeated each token. When results from pretest to posttest and pretest to delayed posttest were compared, it was found that L2 learners significantly improved in their pronunciation productions (Yoshida & Fukada, 2014; for similar findings in L2 learners of Spanish, see Hernández Morales, 2017).

One final method of explicitly teaching pronunciation to L2 learners is via corrective feedback and recasting. Corrective feedback is defined as explicitly revealing the error to the student, providing feedback in the form of modeling, recasting, and eliciting the correct form (Lyster & Ranta, 1997). Ellis et al. (2006) also explain "Corrective feedback takes the form of responses to learner utterances that contain an error. The responses can consist of (a) an indication that an error has been committed, (b) provision of the correct target language form, or (c) metalinguistic information about the nature of the error, or any combination of these" (p. 340). In terms of implementing this method for pronunciation instruction, corrective feedback can then be interpreted as acknowledging a mispronounced utterance or segment, correcting the mispronounced utterance or segment, as well as an explanation of the mispronunciation to accompany these.

In several studies, corrective feedback has taken the form of (1) acknowledging that L2 learners have not produced a segment in a target-like manner and (2) recasting these segments for L2 learners so that they are able to hear the target-like production again, as well as (3) the L2 learners having the opportunity to practice and reproduce the target phoneme or token, as these have been commonly used (Dlaska & Krekeler, 2013; Saito, 2013; Saito & Lyster, 2012; Yoshida & Fukada, 2014). Results from these studies that incorporated corrective feedback concluded that

corrective feedback aided L2 learners in improving their pronunciation of the focus phonemes. Yoshida & Fukada (2014) coupled repetitions with corrective feedback, while other research more specifically focuses on the sole use of corrective feedback (Dlaska & Krekeler, 2013; Saito, 2013; Saito & Lyster, 2012).

Dlaska and Krekeler (2013) implemented a form of individual corrective feedback (ICF) in their study involving L2 learners of German. This study included two groups of L2 learners: the first group received implicit feedback, in the form of instructor repetitions and recasts, while the second group received implicit feedback in addition to ICF. With this approach, ICF indicates that each learner receives feedback from an instructor. This included instructions and comments on how to produce vowels, consonants, word stress, and intonational features in a more native-like fashion. They concluded that L2 learners in the ICF group improved significantly in their productions from pretest to posttest in comparison to learners who did not receive ICF. This study also states that while their methods slightly differed, their results were similar to those of Saito & Lyster (2012), who also incorporated corrective feedback (Dlaska & Krekeler, 2013).

Combination of Explicit Instruction Methods

In several studies, instead of isolating one explicit instructional approach, researchers have combined several of the explicit methods. For example, Aliaga-García and Mora (2009) specifically used articulatory descriptions of different variations between Spanish and English stops, as well as teaching L2 learners to associate sounds with IPA symbols. Then the learners imitated the modeling of the researchers, followed by group work involving the production of isolated tokens and tokens embedded strings of speech. The learners then worked up to full dialogues and tongue twisters involving the target phonemes. As the learners were taking part in each of the activities, they were also receiving explicit feedback from NSs of the target language. In the study conducted by Kissling (2013), similar procedures were used to aid L2 learners of Spanish to reduce aspiration of word-initial voiceless plosives, as L2 learners participated in activities with diagrams of the vocal tract as well as grapheme to phoneme practice. The final activity consisted of learners indicating which sound was which through the use of IPA symbols.

With the review of the previous studies, a combination of several explicit methods appears to be the prevailing method and is thus adopted for the current study. With this combination consisting of IPA use, modeling, explanation of articulation, repetitions, and corrective feedback

(recasting and repetition), this description is proposed as being the most viable option for explicit pronunciation instruction.

2.2.2 Visual Feedback

Beginning in the 1980s with more accessible computer technology, researchers began to develop a pronunciation teaching technique called visual feedback (VF) (deBot, 1983; deBot & Mailfert, 1982; Weltens & deBot, 1984). In this type of feedback, learners were often shown productions of NSs on a computer screen, using this type of visual cue as a comparative basis for their own productions. Participants were then asked to attempt to match their pronunciation to that of the NS speaker's productions in order to facilitate production of native-like speech. This type of feedback relies on providing a second modality (visual) to help L2 learners perceive their own mispronunciations (Dlaska & Krekler, 2008).

Some of the first studies that included visual feedback focused primarily on suprasegmental features (Chun, 1989; Chun, 1998; deBot, 1983; deBot & Mailfert, 1982; Hardison, 2004; Weltens & deBot, 1984). In one of the earliest visual feedback studies, deBot and Mailfert (1982) found the L2 learners of English were able to significantly improve their production of intonation contours after receiving visual feedback involving different pitch ranges of NSs being displayed on a screen for NNSs. They concluded that L2 learners were able to improve significantly from the pretest to posttest comparisons after receiving the intonational visual feedback. Other earlier studies show improvement of intonational properties after receiving visual feedback training (Chun, 1989; deBot, 1983); however, these earlier studies did not include the treatment of individual segments for L2 learners.

Building on previous work at the suprasegmental level, along with acknowledging the role that segmental features play also in intelligibility, comprehensibility, and accentedness (Derwing & Munro, 1997; Derwing & Munro, 2009), visual feedback has more recently been employed for instruction at the segmental level. For example, Chun (2002) proposes visual feedback through observing the different segmental characteristics contained in a spectrogram, visually displaying differences concerning minimal pairs for vowels. For distinguishing between segments, this type of visual feedback could be most beneficial.

A growing number of studies continues this use of visual feedback in the classroom for a variety of segmental features (Motohashi-Saigo & Hardison, 2009; Offerman & Olson, 2016;

Olson, 2014b; Saito, 2007; Wang & Munro, 2004). Motohashi-Saigo and Hardison (2009), for example, found that students were able to significantly improve their productions of Japanese geminates after visual feedback training involving the use of sound waves to distinguish between the geminate vowel and singleton vowel productions. They conclude that "visual cues are a valuable source of input in L2 learning" (Motohashi-Saigo & Hardison, 2009, p. 42).

Praat (Boersma & Weenink, 2019) has been suggested as a tool for visual feedback instruction of segmentals, due to its display capabilities of vowels and consonants via sound waves and spectrograms (Olson, 2014a; Olson, 2014b; Saito, 2007). Saito (2007) found that participants made substantial improvement after visual feedback treatment that was provided via Praat for distinguishing between English and Japanese vowels, after looking at spectrograms of NS productions for the treatment. In addition, Olson (2014b) found that L2 learners of Spanish significantly improved in their productions of fricative variants of the voiced stops in intervocalic position after participating in visual feedback activities, which included comparison of learner and NS waveforms and spectrograms.

Visual feedback treatments have also been incorporated into an intermediate L2 Spanish course to demonstrate the difference between voiceless plosives [p,t,k], short-lag in Spanish, and the long-lag variants [p^h,t^h,k^h] in American English, which are found in word-initial position. Results revealed that the participants improved significantly after training was carried out (Offerman & Olson, 2016).

2.2.3 Combined Visual Feedback & Explicit Pronunciation Instruction

While visual feedback done by Offerman and Olson (2016) is more of an inductive paradigm, guiding L2 learners to notice differences without explicit explanations, there exist several studies that have incorporated more of a combined pronunciation instruction. However, these studies either involved L2 learners that had never had experience with the language before or higher level L2 learners that were often involved in a type of phonetics course for refinement of pronunciation skills.

As an example of this combined approach, Kartushina et al. (2015) conducted a study in which monolingual speakers of French received a combined training of corrective feedback coupled with visual feedback for Danish vowels. The use of monolingual speakers specifically calls attention to the idea of pronunciation instruction for novice learners, as these speakers had no

previous exposure to any variety of Scandinavian languages. For the monolingual speakers chosen for this study, they were organized into a control group and an experimental group. For the visual feedback portion, monolinguals were shown formants 1 (F1) and 2 (F2), which included an explanation of what these formants indicated in terms of accuracy of vowel production and tongue placement. The control group was also shown the same F1 and F2 values of their productions and those of NSs, but they were given no explanation nor any corrective feedback. It was found that the experimental group improved in their production of four of the five vowels under investigation after receiving the combined training, while the control group displayed no improvement after being shown formant values only (Kartushina et al., 2015). Although these monolinguals were not L2 learners of Dutch and were only tested on specific segments (vowels), it still supports the idea of implementation of a combined pronunciation instruction strategy.

In the study conducted by Lord (2005), visual aids in the form of spectrographs were incorporated into L2 pronunciation instruction for the purpose of showing L2 learners the differences in VOT values of the voiceless plosives in the onset position for NSs of English versus NSs speakers of Spanish. However, Lord (2005) mentions that the L2 learners that participated in this type of treatment were labeled as students enrolled in an "upper-division university Spanish phonetics class" (p. 561). While this may have aided these learners in the improvement of voiceless plosive production in Spanish, this was not the emphasis of the study, along with the course being specific to only higher-level learners where phonetics and pronunciation were the main focus for the entire semester.

For lower-level learners, as well as those who are not able to take part in a phonetics-specific course, types of empirically-tested pronunciation instruction should also be designed for such learners.

2.2.4 **Production Gaps**

To summarize, while there have been many attempts in the last several decades to begin incorporating L2 pronunciation instruction into learner curriculum, there is still a lack of empirically tested methods that involve various types of instruction. Additionally, not only is this issue absent from the current pronunciation literature, there is also evidence indicating that L2 learners want pronunciation instruction, and instructors need materials to be developed. In consideration of these issues, a study implementing various pronunciation instruction methods to

be empirically tested, as strongly suggested by Lee et al. (2015), is proposed at the conclusion of this chapter.

Further, there has yet to be a comparative analysis that focuses on L2 learners at a lower level, not involved in a phonetics course. While some studies have looked at the effects of pronunciation instruction on L2 learners at a lower level (e.g., Miller, 2012; Olson, 2019) or those without previous knowledge of a second language (Kartushina et al., 2015), there has yet to be a study conducted that observes which type of pronunciation instruction is most beneficial to L2 learners, and whether or not lower-level L2 learners benefit from various types of intervention.

As the first portion of this review has focused on production, there comes into question the role that perception of L2 sounds plays in L2 pronunciation learning and production. The following section details several views on perception and how it influences or is possibly influenced by L2 production.

2.3 L2 Perception

While the emphasis of pronunciation instruction has been squarely on the *production* of segmental (and suprasegmental) features, the role of *perception* in the L2 instructional process is less well-understood. Addressing the extant literature on the role of perception in L2 phonetic acquisition, this section is divided into three parts: Current L2 Perception Theories (Section 2.3.1); L2 Perception Methods (Section 2.3.2); Perception Gap (Section 2.3.3).

2.3.1 Current L2 Perception Theories

Within the field of L2 perception, there are multiple theories that have been proposed to explain how L2 learners attempt to perceive new or unfamiliar sounds, along with sounds that are very similar to their L1 phonological categories. Among these theories, four primary theories are referenced and utilized in a variety of research: the speech learning model (SLM) (Flege, 1987; 1991; 1993; 2003; Flege & MacKay, 2004; Flege et al., 1995; Flege et al., 1997); the perceptual assimilation model (PAM) (Best, 1994; Best, 1995; Best et al., 2001); the PAM-L2 (Tyler & Best, 2007) and the second language perception model (L2LP) (Escudero, 2005; 2007; 2009; Escudero & Boersma, 2002; Escudero et el., 2008; Escudero et al., 2009; Mayr & Escudero, 2010; van Leussen & Escudero, 2015).

The Speech Learning Model (SLM) (Flege, 1987; Flege, 1991; Flege, 1993; Flege, 2003; Flege et al., 1995), a production-oriented theory, was developed with the idea that an adult L2 learner relies heavily on their L1 phonological categories in order to perceive sounds in the L2, therefore causing the L1 to influence the L2, and vice versa over time (Flege & MacKay, 2004). SLM predicts that the longer a L2 learner is exposed to their L1, the more difficult category formation may become in the L2 (Flege, 2003). As such, any sounds that have similar features in the L1 and L2 may be subject to category assimilation, even though these sounds have at least one or more distinct differences. When this occurs, it is predicted that the L2 learner then develops a merged category of L1 and L2 sounds (Flege et al., 1995). In contrast, sounds in the L2 which are more distinct and salient, or differ significantly from L1 sounds across a number of features, are considered less difficult to approximate and acquire (Flege et el., 1995).

For the Perceptual Assimilation Model (PAM) (Best, 1994; 1995; Best et al., 2001), a perception-oriented model, this model hypothesizes that L2 learners also attempt to assimilate features of sounds in the L2 to the L1, as discussed in the SLM. However, one difference that the PAM proposes is that more unfamiliar sounds may be more difficult to acquire than similar sounds (Best, 1995). The PAM breaks down the differences in phonological categories between these three facets: single-category assimilation, two-category assimilation, and category goodness assimilation (Best, 1994; 1995; Best et al., 2001). For single-category assimilation, this is interpreted as two sounds in the L2 that consist of similar sounds to one sound in the L1, causing difficulty in production for the L2 learners. Two-category assimilation proposes that a sound in the L2 that contains features like that of two similar sounds in the L1, this will be more easily and accurately perceived by the L2 learner. Lastly, category goodness refers to the case when two different sounds in the L2 assimilate to a single sound in the L1 but are perceived as differing in their ‘fit’ or similarity to the L1 sound (Best, 1994; 1995).

An extension of the PAM is the PAM-L2 (Best & Tyler, 2007), which is a speech-oriented model for L2-learners as opposed to the perception-driven PAM. In the PAM-L2, how well L2 learners notice differences and acquire phonological categories is largely dependent on the L1. Similar to the PAM, it is thought that L2 learners develop one-category assimilations, two-category assimilations, along with category goodness assimilations, with the first and the latter not thought to be as frequent in a L2 classroom environment (Tyler, 2019). It is also posited that, for learning new lexicon in the L2, L2 learners should be provided with the opportunity to receive L2

perceptual training in L2 classrooms to account for assimilation issues in the L1 and L2 (Tyler, 2019).

The second language perception model (L2LP) (Escudero, 2005; Escudero, 2009) details a model of L2 perception that is more specific to individual learner differences according to their level of proficiency. In other words, this model was developed in order to describe perceptual acquisition of sounds to varying degrees dependent on level (Escudero, 2005). Within this model, the optimal perception hypothesis was developed, posing that L2 learners first perceive L2 sounds more similar to their L1, and as they develop in their L2, they are able to perceive more distinctions between phonemes and allophones in their L1 and L2 (Escudero, 2005). Typically, L2 learners create a new category, or a 'new scenario', or a 'similar scenario' in which they initially replicate their L1 category to imitate an L2 category, then eventually adjust to the different features of a sound in the L2 (Escudero, 2005). However, single category assimilation can still occur in which two sounds in the L2 are difficult to distinguish, and the learners assign the same category from the L1 to differing sounds in the L2 (Escudero & Boersma, 2002).

Lastly, the Noticing Hypothesis (Schmidt, 1990) brings forth the notion of *noticing* (Izumi & Bigelow, 2000; Schmidt, 1990; Schmidt, 1992; Swain & Lapkin, 1995) in the L2, and states that when L2 learners are made to direct their attention to a form, they are more likely to acquire this form, as this makes L2 learners aware or conscious of differences between the L1 and the L2. The Noticing Hypothesis goes in contrast to previous methods encouraging 'unconscious' learning (Robinson, 1995). Additionally, Tyler (2019), although never explicitly supporting the Noticing Hypothesis, urges that L2 instruction should include some type of explicit training for L2 learners, as this may assist them in better perceiving differences between the L1 and L2 if they are experiencing difficulty. While the Noticing Hypothesis was developed primarily for describing how explicit instruction could assist in making L2 learners aware of differences or novel grammatical structures, the idea of *noticing* can also be applied to L2 perception.

Among the various theories that have been detailed in this section, the following study will re-examine L2 perception theories to serve as possible explanations for the results (refer to Chapter 5 and Chapter 6). The following subsection will discuss more in detail a type of methodology that appears to be most appropriate for L2 learners when distinguishing between phonemes and allophones in the L1 and the L2, and finally pinpointing which version of the specific methodology chosen would best suit the interests of the study.

2.3.2 **Methods for L2 Perception Testing**

Considering perceptual testing methods for L2 learners, an AXB perceptual test, also sometimes implemented as ABX or XAB, has been among the most popular (Curtin et al., 1998; Escudero et al., 2009; Pater, 2003). This task involves having listeners choose if sound X is perceived as being more similar to sound A or sound B (Escudero et al., 2009). This has been carried out by having L2 learners listen to synthetic sounds (Escudero et al., 2009), as well with an accompanying picture task to have L2 learners associate certain sounds with pictures (Curtin et al., 1998). Other types of perceptual testing include discrimination tasks between two different sounds (Bond & Fokes, 1991; Kissling, 2015), as well as tasks where L2 learners were given the option to choose from a group of words in a closed-set response (Iverson & Levy, 2007). Other perceptual tasks have had L2 learners attempt change and no-change trials, where three tokens are presented, and the learners must decide if there is a difference in one of the three tokens presented (Flege & MacKay, 2004). Picture tasks have also been implemented (Brown, 1998) to test whether learners can perceive the difference in varying degrees of articulation; for instance, the difference between [b], [p] and [p^h] in Thai.

As the AXB is among the most common for discrimination tasks, and is the task included in the current study, it is worth considering several key aspects of the AXB paradigm. For this particular type of task, the principal goal deals with L2 learners' ability to differentiate between L1 and L2 phonemes or allophones, or among L2 phonemes, so it is worth reviewing a number of aspects about this particular paradigm. When choosing to implement this type of perceptual discrimination task, a primary concern is which macro-structure (AXB, XAB, ABX) is most appropriate for the goal of discrimination between sounds and for the participants.

In a review of these types of perceptual tests conducted, it has been mentioned that there is more of a time cost when X (the target sound on which the discrimination task lies) is positioned at the beginning (XAB) or at the end (ABX) of recordings that are provided for L2 learners (McGuire, 2010). With this observation, it can be concluded that the most ideal option for L2 learners would be a version of this test in which sound X is positioned between the target and non-target sounds, A and B; in other words, the most beneficial version of this discrimination task would be the AXB ordering.

Additionally, while these AXB tasks have focused on discrimination, there is a secondary component to perception in a L2; it is not sufficient to be able to discriminate between contrasts,

as L2 learners must also learn how to implement these contrasts, and how they relate to NS norms. That is, L2 learners must know how to accurately associate one sound with the L1 and another sound with the L2. For measuring how L2 learners rate sounds as either being native-like or non-native-like, a type of rating scale (e.g., Likert-scale) task is suggested, as this has been utilized in the past for NSs to rate the accentedness of NNSs (Derwing & Munro, 1997).

2.3.3 L2 Perception Gap

Although there are many propositions for explaining the process of L2 perception, the impact of pronunciation instruction, and particularly production-oriented pronunciation instruction, on L2 learner perception is less well-understood. Although some researchers have suggested employing some explicit methods to teach L2 perception and examine its outcomes on L2 learner perception and production (Tyler, 2019), there have been few studies that have specifically looked at the efficacy of production treatment on L2 perception. Among other open questions, there is an ongoing debate about whether improvement in production is predicated on a preceding improvement in perception.

In a study conducted by Levy and Law (2010), their results lead them to the conclusion that perception must precede production. However, in a study carried out by Flege et al. (1997) they conclude that the relationship between perception and production is still unclear, and it is not necessarily presumable that perception always precedes production. After conducting a study involving the Spanish and English word-initial, voiceless stops, Zampini (1998) states that "…it appears that some learners learn to make the phonetic category substitution for Spanish /p/ before they make corresponding changes in perception" (p. 98), indicating that, before perceptually differentiating between short-lag and long-lag /p/, some learners produce more target-like pronunciation before tightly grasping the differences between short-lag and long-lag /p/ perceptually. Along with this conclusion, as mentioned previously, the study conducted by Kartushina et al. (2015) also proposes that production training does not necessarily indicate a strong relationship between production improvement and perception improvement.

Kartushina et al. (2015) did not test formally instructed L2 learners of Dutch, as these novice learners were taught various segments in Dutch to examine the relationship between perception and production. Even though these monolinguals could reproduce more accurate productions of segments after treatment, it was not apparent that these phonological categories

became embedded within the L2 categories of these speakers; that is to say, it may be that the participants of this study merely memorized and practiced certain segments while not actually learning how to differentiate between specific segments. The assumption that perception tends to precede production (Levy & Law, 2010) has yet to be investigated more fully, as several authors maintain that this relationship is not clear (Flege et al., 1997; Zampini, 1998).

2.4 VOT Differences for Voiceless Plosives in English and Spanish

As the current study focuses on voice onset time (VOT) in L1 English-speaking L2 learners of Spanish, it is worth reviewing the inherent differences between the two languages. Broadly, VOT is defined as the onset of voicing that precedes or follows the release of a stop consonant (Abramson & Whalen, 2017). VOT is considered to be short-lag in Spanish (Hualde, 2005; Diaz-Campos, 2004), with very little aspiration (post-release) that precedes the onset of voicing. The phonemes /p,t,k/ are produced with short VOT values or 0ms in Spanish, whereas allophones of these phonemes produced by NSs of American English are typically produced with a greater VOT value in the onset position (Lisker & Abramson, 1964; Flege, 1991; Gonzalez-Bueno, 1997) and are represented as [p^h,t^h,k^h] (Hualde, 2005).

Lisker and Abramson's (1964) seminal study states that averages for VOT values in the onset position for Spanish are as follows: /p/ = 4ms; /t/ = 9ms; /k/ = 29ms. VOT values for American English, according to Lisker and Abramson (1964), were reported as: /p/ = 58ms; /t/ = 70ms; /k/ = 80ms. This production of Spanish voiceless stops can prove to be difficult for L2 learners whose native language is English (Diaz-Campos & Lazar, 2003). Within the framework of SLM theory (Flege, 2003), these difficulties can be explained by the idea that while L1 English speakers may have phonological inventories that contain the voiceless stop allophones (Lisker, 1957), this does not necessarily mean that they can perceive the difference between a short-lag and long-lag voiceless stop in a different phonological context than what they are accustomed to (word-initial vs. word-medial). Along with this theory, there may also be individual differences that can come into play that cause distinguishing these variations by lower-level L2 learners to be challenging, according to the L2LP theory (Escudero, 2005; Escudero, 2009).

2.5 Conclusions & Research Questions & Hypotheses

As the field of pronunciation instruction research has developed over the past decade, several key trends have begun to emerge. First, in response to the well-documented lack of pronunciation materials, two major approaches to pronunciation instruction have emerged: explicit instruction and visual feedback. Second, while empirical analysis of pronunciation instructional methods has begun to emerge, such research is likely to focus on a single treatment method. And finally, pronunciation instruction and research has focused primarily on L2 production and production-oriented methods, largely ignoring the development of L2 perception and its relationship to L2 production.

As such, three main gaps emerge in the current literature: First, there is a clear need to continue to empirically analyze L2 pronunciation instructional methods. Second, in order to maximize benefit for learners, there is a need for comparative research that tests the efficacy of different instructional approaches. And finally, further study is needed to examine the role that instruction plays on L2 perception, and the interaction between L2 production and L2 perception.

With respect to these conclusions and the literature reviewed in the previous subsections, the following research questions and their corresponding hypotheses have been proposed below:

(RQ1) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of L2 phonetic production? In this case, improvement of phonetic production is defined as a reduction in VOT by English-speaking L2 learners of Spanish in the onset position of a word.

> Hypothesis 1: Drawing on findings for explicit pronunciation studies (e.g., Aliaga-García & Mora, 2009; Elliott, 1997; Lord, 2010), visual feedback studies (e.g., Motohashi-Saigo & Hardison, 2009; Olson, 2014b), and a combination approach (Kartushina et al., 2015), it is hypothesized that all experimental groups will significantly reduce their VOT of the word-initial voiceless stops in Spanish after having received a form of production training.

(RQ1.1) In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback [VF] training or a combination training [CI]) on L2 Spanish production?

Hypothesis 1.1: Drawing on findings for explicit pronunciation studies (e.g., Aliaga-García & Mora, 2009; Elliott, 1997; Lord, 2010), visual feedback studies (e.g., Motohashi-Saigo & Hardison, 2009; Olson, 2014b), and a combination approach (Kartushina et al., 2015), it is hypothesized that L2 learners in both the EI and VF groups will improve significantly in their productions; however, there will be no statistically significant data showing that L2 learners improved more because of one, singular treatment over the other. They will instead improve more significantly with the combined training approach.

(RQ2) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of perception for discrimination skills? In this case, improvement of phonetic perception is defined as the ability to discriminate between long-lag and short-lag variants of /p,t,k/ in word-initial position.

Hypothesis 2: Carried out via an AXB task, findings will reveal that L2 learners in all experimental groups will display trends of improving their perception of the difference in long-lag and short-lag voiceless stops in word-initial position after receiving pronunciation training. It is hypothesized that L2 learners in all experimental groups will improve based on the Noticing Hypothesis (Schmidt, 1990), where learners are predicted to improve due to being directed to notice differences between the L1 and the L2 with a type of intervention.

(RQ 2.1) In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback [VF] training or a combination training [CI]) on L2 perception via discrimination between Spanish and English variants of /p,t,k/?

Hypothesis 2.1: Carried out via an AXB task, findings will reveal that L2 learners in the CI group will best improve their perception of the difference in long-lag and short-lag voiceless stops in word-initial position after receiving pronunciation training. The CI group will improve most due a combination of learner modalities, namely auditory and visual (Murphy, 1997), aiding learners in discrimination. This is thought to be because both explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as

helping learners notice differences between the L1 and the L2. Therefore, it is possible that a combination of both methods will cause maximal noticing.

(RQ3) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of perception for social categorization skills? In this case, improvement of phonetic perception is defined as the ability to perceive and categorize long-lag tokens as more non-native (English-accented) and short-lag variants as more native (Spanish).

Hypothesis 3: L2 learners in all experimental groups will improve in their ability to perceive long-lag variants of /p,t,k/ as more English-like (non-native) and short-lag variants as more Spanish-like (native) after participating in production training. Based on the Noticing Hypothesis (Schmidt, 1990), learners are predicted to improve due to being directed to notice differences between the L1 and the L2 via all three production treatments.

(RQ 3.1): In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback [VF] training or a combination training [CI]) on L2 perception via social categorization of Spanish and English variants of /p,t,k/?

Hypothesis 3.1: Findings will reveal that L2 learners in the CI group will best improve in their assignation of short-lag variants being native-like (Spanish) and long-lag variants being non-native (English-accented) after participating in production training. Based on the Noticing Hypothesis (Schmidt, 1990), learners in this group are predicted to improve due to being directed to notice differences between the L1 and the L2 via two different modalities (auditory and visual; Murphy, 1997), as explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as helping learners notice differences between the L1 and the L2. Therefore, it is possible that a combination of both methods will cause maximal noticing.

(RQ4) Does there exist a relationship between perception improvement and production improvement for all experimental groups? That is, does L2 perception improve after production training, indicating that perception does not always precede production?

Hypothesis 4: Drawing on research stating that it is not clear what the relationship is between L2 production and perception (Flege et al., 1997), it is hypothesized that all groups will improve in producing more target-like productions of /p,t,k/ in Spanish post-treatment, as well as improve in their discrimination between and assignation of short-lag and long-lag variants. Again, the Noticing Hypothesis (Schmidt, 1990) is proposed and an explanation for improvement, as explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as helping learners notice differences between the L1 and the L2. This would provide evidence that it is possible that L2 production and L2 perception may develop simultaneously for adult L2 learners after a form of production treatment.

CHAPTER 3. METHODOLOGY

To answer the above research questions, the current study evaluates the impact of three different types of training (explicit pronunciation instruction [EI], visual feedback training [VF], or a combination instruction [CI] of both explicit pronunciation instruction and visual feedback training) on both L2 production and L2 perception compared to a control group (CO) receiving no production training. This section details the methodology as follows: In Section 3.1, participants and backgrounds are described. To follow in Section 3.2, materials and tasks types are described, with Section 3.3 covering the procedures for the production and perception tasks, as well as a description of the production treatments for each group. Section 3.4 details the data measurements and statistical analyses used in order to interpret the results. Section 3.5 concludes this chapter, along with an introduction to Chapter 4, the results of the production portion of the study.

3.1 Participants

The participants were divided into four groups and were all enrolled in four separate intermediate-low level Spanish courses (level 3) at a midwestern university. Each group of participants consisted of 10-15[2] students (average size of intermediate-level classes at the university). Three of four instructors were NS of Spanish, with the VF group being the only group with a NNS instructor[3]. All participants completed a Language Background Questionnaire to assess L2 learner language dominance and proficiency, basing some of the content of the questionnaire on the Bilingual Language Profile (BLP) (Birdsong, 2014; Birdsong et al., 2012). Broadly, the Language Background Questionnaire focuses on a self-reported measure of language history, language attitudes, language proficiency, and language use to serve as a quantitative measure of language dominance; in other words, the questionnaire was utilized to determine that all participants were English-dominant. In short, all subjects involved in the data collection spoke only English from birth, had not studied any language besides Spanish, had not spent more than

[2] Originally, there were a number of participants for each group around 25; however, various participants were excluded for not meeting the descriptions of the language background questionnaire, not completing all required tasks, issues with the recordings, or not giving consent to participate in the study.

[3] The non-native instructor of Spanish is a NS of Georgian. They have extensive experience with the Spanish language (10+ year), use Spanish on a daily basis in both social and professional settings, along with pursuing a PhD in Spanish Literature.

6-weeks in a Spanish-speaking immersion experience (for similar criteria see Offerman & Olson, 2016), and had not had any experience with Spanish classes until after the age of six (Long, 1990). The final number of participants per group consisted of the following in Table 1:

Table 1 *Participants (n) by group for production and perception tasks*

Group	Production	Perception
CO	11	9
EI	11	14
VF	11	11
CI	14	11
Total (n)	47	45

The difference in the number of participants between the perception and production tasks is the result of participants not completing all necessary tasks for each experiment. For example, in the CO group, 11 participants completed all required tasks for the production portion of the study, but for the perception portion, they did not complete all required tasks (this was often due to being absent on the day when the perception pretests and/or posttests were distributed; perception pretests and posttests will be detailed later in Section 3.2).

3.2 Stimuli

Broadly, participants completed four different production tasks (a carrier phrase task; a novel sentences task; a controlled, continuous speech task; and a spontaneous speech task) and two different perceptual tasks (a discrimination task and a nativeness judgment task). All tasks were performed prior to treatment (i.e. pretest) and following the three treatment interventions for /p,t,k/ (i.e. posttest). While the procedure is described more in depth below in Section 3.3, Table 2 shows the general framework for tasks and interventions. Note, in Table 2, treatment-type differs by group (i.e. EI, VF, CI, CO):

Table 2 *Tasks & Phonemes of Focus*

Task	Focus Phoneme	Production Task	Perception Task
Pretest		1-4	1-2
Week 1	/p/ treatment		
Week 2	/t/ treatment		
Week 3	/k/ treatment		
Posttest		1-4	1-2

3.2.1 Production Stimuli/ Tasks

The primary goal for the production tasks is to evaluate whether participants improve not only on the focus phonemes in a carrier phrase, but also to see if generalized knowledge of isolated productions could also be incorporated into novel sentences, controlled, continuous speech and spontaneous speech. This section reviews each production task that was performed by each of the four groups, which include: a carrier phrase task; a novel sentences task; a controlled, continuous speech task; and a picture task. These tasks and stimuli have been reutilized from a previous study by Offerman and Olson (2016).

Carrier Phrase Task

Task 1 was comprised of list of 30 tokens within the carrier phrase *Di* _____ *de nuevo* ('Say _____ again'), like that of Fox et al.'s (1995) elicitation task *Digo ahora* _____ ('I now say _____'). For the 30 tokens contained within the carrier phrase *Di* _____ *de nuevo*, each voiceless stop was paired twice with each of the 5 Spanish vowels /a,e,i,o,u/ to create words containing an equal number of stop and vowel pairings (3 plosives /p,t,k/ × 5 vowels /a,e,i,o,u/ × 2 = 30 tokens). This structure served to control for any possible variance between the different vowel environments.

All five vowels were utilized due to possible VOT variances that could arise in the production due to the following vowel of each voiceless stop (Port & Rotunno, 1979). All tokens were all chosen as two-, three-, or four-syllable tokens, as well as being non-cognates for the language pair (Spanish and English) (Amengual, 2016). Tokens were also limited to an initial syllable structure of CV. For the vowel following each of the three stops, no diphthongs were included so as not to impact the vowel quality following the stops. Table 3 provides an example of the list for Task 1 (for a full list, see Appendix A):

Table 3 *Tokens Embedded in Carrier Phrase*

Plosive	Example 1	
/p/	Di *pesa* de nuevo. 'Say *weight* again.'	Di *poco* de nuevo. 'Say *little* again.'
/t/	Di *tela* de nuevo. 'Say *fabric* again.'	Di *toca* de nuevo. 'Say *touches* again.'
/k/	Di *quepo* de nuevo. 'Say *I fit* again.'	Di *copa* de nuevo. 'Say *wineglass* again.'

Novel Sentences Task

For Task 2, participants recorded 30 different novel sentences in order to elicit short speech segments containing the tokens (Elliott, 1997). The novel sentences, in contrast to the carrier phrases, provided limited focus on the target word, and represents a relatively more complicated speech task. Each novel sentence contains one token, taken from the list of 30 tokens embedded in the utterance *Di* ____ *de nuevo* for the phoneme of focus, seen below in Table 4 (for a full list, see Appendix B). Each stop was again paired with vowels /a,e,i,o,u/ twice (3 plosives /p,t,k/ × 5 vowels /a,e,i,o,u/ × 2 = 30 tokens).

Table 4 *Tokens within Novel Sentences*

Plosive	Example 1	Example 2
/p/	Quiero un *poco* de agua. 'I want a *little* water.'	No sé por qué *Paco* quiere irse. 'I don't know why *Paco* wants to leave.'
/t/	Hay un *testigo* con el juez. 'There is a *witness* with the judge.'	*Toca* la guitarra para mí. '*Play* the guitar for me.'
/k/	Esa *cosa* no sirve para nada. 'That *thing* is useless.'	Llévame a *casa* por favor. 'Take me *home* please.'

Controlled, Continuous Speech Task

Task 3, the controlled, continuous speech task, consisted of a short story reading. The short story contains 30 tokens for /p,t,k/ that are also controlled for the following vowel and syllable structure. Unlike production Tasks 1 and 2, the short story reading elicits continuous, connected, and contextually bound speech. This task also provides a total of 30 tokens in continuous speech (3 phonemes × 10 tokens per phoneme = 30 tokens). The story consisted of 479 words with the 30 tokens embedded, about 1.5 pages, double-spaced.

It could be argued that the short story creates a scenario in which participants may rely too heavily upon orthography to produce continuous speech; however, it does provide a control for the number of times each token is produced and distributed throughout the continuous speech (Diaz-Campos, 2004; Elliott, 1997; Lord, 2008). The purpose of the short story in the pretest and posttest is to observe how participants initially produced the target plosives within controlled, continuous speech, as well as to look at how well the participants transferred their later acquired knowledge of the voiceless stops within a longer speech context at the end of the study. Example 1 below (Offerman & Olson, 2016) provides a small excerpt from the short story (for the full task, see Appendix C):

Example 1

Me llamo Paco y quiero contarte sobre mi primera experiencia con mi compañero, Pedro. Había acabado de cumplir 18 años, y tuve que mudarme a Indiana para mi primer año de la universidad.

"My name is Paco, and I want to tell you about my first experience with my roommate, Pedro. I had just turned 18 years old, and I had to move to Indiana for my first year of undergrad."

Spontaneous Speech Task

For the spontaneous speech task, 30 tokens were chosen containing /p,t,k/ that are controlled for the following vowel and syllable structure, similar to other tasks (3 phonemes × 5 vowels × 2 = 30 tokens). More specifically, tokens were taken from the lists of tokens within the carrier phrase for consistency of CV ordered tokens controlled for stop and following vowel. In efforts to create an environment in which participants produced 'free speech', the spontaneous speech task (a picture task) was integrated into the set of tasks (Offerman & Olson, 2016). Although this method of eliciting free speech is still guided speech to an extent, it does not provide the learner with the orthography of the target token on the elicitation slide in which the participants were to produce a longer utterance containing the target token. The picture task (Bigelow et al., 2006; Elliott, 1997; Offerman & Olson, 2016; Willis, 2006) elicits the productions of the voiceless plosives in a spontaneous speech environment.

In the picture task, a PowerPoint presentation was created, which consists of 30 pairs of slides (30 tokens × 2 slides = 60 slides). For each pair of slides, an instructional slide with the

target token (Figure 1) preceded the spontaneous speech slide to elicit the target token (Figure 2). First, directions were distributed to the participants, having them study the picture associated with the token in the PowerPoint slide to remember it for the second PowerPoint slide. In the first slide either the name *Juan* or *Mariana* appeared written within the sentence. The second slide displayed either a picture of a boy or a girl with the names *Diego* and *Ana*, without these names contained in the sentence.

Participants were instructed to insert the new name into the sentence, along with remembering the target token associated with the picture. The purpose of having the students change the name of the person in the second slide is to serve as a distraction device, which does not permit participants to exclusively focus on the target token (Offerman & Olson, 2016).

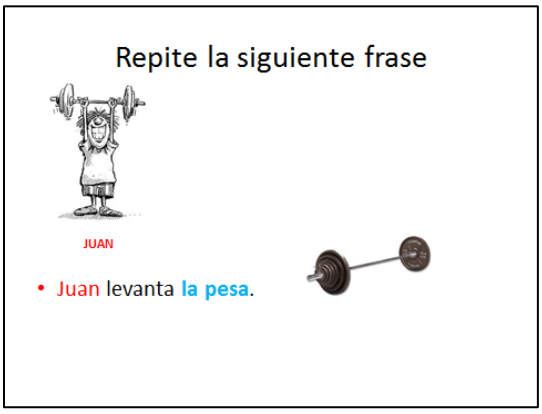


Figure 1 Picture task instructional slide

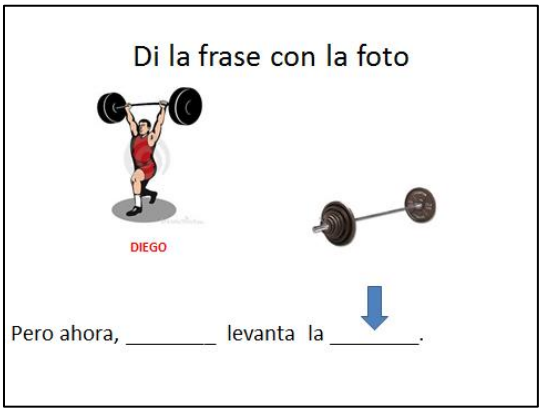


Figure 2 Picture task elicitation task

3.2.2 **Perception Tasks**

In previous studies, several authors claim that L2 production may not always be indicative L2 perceptual skills (Flege et al., 1997; Kartushina et al., 2015); therefore, this portion of the methods investigates the link between perception and production by implementing two types of perceptual tests: a discrimination task and a nativeness judgment task. The first seeks to discover whether the participants have made any improvement in discrimination after production training, as it has been suggested that L2 production training does not necessarily improve L2 perceptual skills (Kartushina et al., 2015).

The second perceptual test seeks to find whether participants are able to associate short-lag (Spanish) variants with NSs and long-lag (English-accented) variants with NNSs of Spanish[4]. This section of the methodology has been proposed in order to test the perception of L2 Spanish participants on their ability to distinguish between the variants [p,t,k,] in Spanish in word-initial position, and [p^h,t^h,k^h] in English in word-initial position. The same four groups of participants took part in the production tasks also took part in these perceptual tasks. As indicated in Table 2, both perceptual tests were administered once before (during the pretest) and once after (during the posttest) all forms of production training for each experimental group, and before and after the activities for the CO group (the experimental and CO group activities will be discussed in Section 3.3).

Discrimination Task

Materials for the first perceptual task consisted of pairs Spanish tokens, differing only in VOT values for word-initial stops of the same token for a discrimination task. All tokens were Spanish words containing a variation of short-lag (native-like) and long-lag (English-accented) productions of /p,t,k/ at the onset, and were balanced for syllable structure (all tokens were two-syllable and paraxytonic). A total of six pairs of tokens were used twice, balanced for six initial segments /p,t,k/ (three short-lag and three long-lag), as well as being balanced for the following vowel (/a,e,o/)[5] (6 segments × 3 vowels × 2 = 36 tokens). To create the audio version of each token,

[4] NNSs of Spanish refer to American English speakers who are second language learners of Spanish.
[5] The mid and low vowels /a,e,o/ were chosen due to time constraints for the perception portion; high vowels /i,u/ were not chosen since they are more susceptible to diphthongization (Whitley, 2002).

a male, NS of Spanish was recorded producing multiple iterations of each token, several times with shot-lag variants and several with long-lag variants.

To ensure a balanced design, both the short and long-lag versions of each token were created using a cross-splicing procedure. The short-lag version of each word was constructed by taking the target word from one native-speaker production and cross-splicing (cutting segments from other versions of the token produced and pasting them to a separate version of the token) the aspirated portion of a different production of the same token by the same speaker. The long-lag version was created by manipulating the aspiration contained in the short-lag token such that the duration in milliseconds equaled that produced by a NNS for a basis of comparison.[6] This was done by elongating the aspiration from a different short-lag production until the VOT value was equivalent to that of a NS of English, and splicing the new, longer aspiration, on to the same NS production used to create the short-lag token. Tokens were counterbalanced for order of the long-lag and short-lag iterations (A and B), as well as the nature (long-lag or short-lag) of the test token (X), resulting in four different orderings for each of the six pairs (4 orderings × 36 tokens = 144 total). Table 5 illustrates an example the counterbalanced stimuli pairings for the token *pato* ('duck'; /pa.to/).

Table 5 *AXB Pairings*

Sample	Order
[pa.to] [pa.to] [p^{h}a.to]	AXB
[pa.to] [p^{h}a.to] [p^{h}a.to]	AXB
[p^{h}a.to] [pa.to] [pa.to]	AXB
[p^{h}a.to] [p^{h}a.to] [pa.to]	AXB

Participants listened to 144 tokens in a randomized order (3 variations of a token per PowerPoint slide × 48 slides = 144 total tokens) and were asked to complete a response sheet to indicate whether the test token (X) was more similar to either version A or B of the token. Figure 3 shows a sample from the response sheet (also see Appendix D):

[6] NS tokens were produced by a male, native Spanish speaker of Puerto Rican Spanish. NNS tokens were produced by the same male, as this speaker has an extensive background with English.

Instrucciones: Escucha las siguientes palabras en el power point. Después, indica (circle) la palabra que es más parecida a X.

1)	A	B	12)	A	B	23)	A	B
2)	A	B	13)	A	B	24)	A	B
3)	A	B	14)	A	B	25)	A	B

Figure 3 AXB response sheet

In the perceptual test, these sounds were ordered in the form of AXB, which seems to be the most beneficial and easiest organization for accessing the more similar sound to X in comparison to orderings ABX and XAB (McGuire, 2010). One or two segments (Escudero et al., 2009) have been used with L2 learners of a particular language, as well as entire words with participants who are unfamiliar with the language (Pater, 2003). With this in mind, to ensure activation of the Spanish lexicon, this task employs word-length tokens. All presented recordings for this task were recorded, manipulated, and analyzed via Praat (Boersma & Weenink, 2019).

Nativeness Judgment Task

A second perceptual test was administered prior to and after production training to see on a more gradient scale if participants could identify how native (Spanish) or non-native (English-accented) a production sounded involving allophones of /p,t,k/. The same recordings were used from the first task in the pretest and posttest perception task, this time with only one token (short-lag or long-lag) presented at a time. Once again, all tokens were Spanish words containing /p,t,k/ at the onset, were balanced for syllable structure (all tokens were two-syllable and paraxytonic), and balanced for the following vowel. Following each auditory presentation, participants were asked to indicate on a 7-point Likert-scale how native-like or non-native-like the token sounded to them. Figure 4 illustrates a sample response sheet (also see Appendix E):

Instrucciones: Escucha las siguientes palabras en el power point. Después, indica (circle) un número para indicar cual se parece más a un hablante nativo de español o un hablante que no es nativo del español.

1)

1 2 3 4 5 6 7

Native Non-native

Figure 4 Likert-scale nativeness judgment rating

3.3 Procedures

Participants were divided into four separate groups, three experimental groups receiving different forms of pronunciation instruction and one control group (CO). The experimental groups included: (1) explicit pronunciation instruction (EI); (2) visual feedback instruction (VF); (3) a combination instruction (CI), consisting of visual feedback and explicit pronunciation instruction. For all groups, each of the required tasks was integrated in the course's syllabus as 'homework'. These assignments were graded for completion, not based on performance of any kind, to ensure that there was no compensation in the form of a grade for learner performance in relation to how well participants produced the tokens for each of the tasks. No compensation in any form was given to the participants, as the tasks were considered part of the curriculum for the level three course. Participants provided recordings at the pretest and posttest, along with recordings before and after each treatment session.[7]

3.3.1 Explicit Pronunciation Instruction Group

For the explicit instruction (EI) group, all participants took part in the perception pretest, as well as all recording tasks (pretest task, phoneme of focus treatments, posttest), ending with the perception posttest. For the production treatments, the first training session was comprised of activities surrounding /p/ for 15-20 minutes (20 minutes was allotted in case participants had extra questions, being that this was the first treatment session). The second and third trainings involving

[7] Intermittent recordings were not included into the final analysis due to inconsistent submission of these recordings by participants.

/t/ and /k/ lasted 15 minutes per training session. Both perceptual tasks, the discrimination task and the nativeness judgment task, for the EI group totaled 20 minutes for the pretest and about 15 minutes for the posttest.

Explicit Instruction Treatment

For the EI group, explicit pronunciation training and instruction consisted of presenting participants with IPA symbols (Lord, 2010; Miller, 2012; Rajab, 2013; Sturm, 2013a), articulatory gestures (Elliott, 1997), repetition and drills (Yoshida & Fukada, 2014) and corrective feedback (Saito & Lyster, 2012) for the respective day of instruction (i.e. [p] vs. [p^h], [t] vs. [t^h], [k] vs. [k^h]). This also included an explanation that the short-lag variants are typically produced by NSs of Spanish in word-initial position (Lisker & Abramson, 1964), while the long-lag variants are typically produced by NSs of English in word-initial position (Hualde, 2005).

The instructor then explained to the participants that NSs of English that are adult L2 learners of Spanish tend to produce the long-lag variants (Lord, 2005). In addition to the explanation through IPA symbols, participants were then explained the differences in articulation gestures (Aliaga-García & Mora, 2009; Elliott, 1997; Kissling, 2013) for the short-lag vs. long-lag variants on their respective days of instruction, specifically pointing out that there should be little to no air leaving the oral cavity during production for Spanish.

To follow these explanations, the instructor then modeled productions (Saito & Lyster, 2012) of the short-lag vs. long-lag variants, using tokens in Spanish for the short-lag variants and tokens in English for the long-lag variants to demonstrate the difference for participants in the case that they were unaware. The instructor then asked participants to repeat each of the tokens five times for a set of drills (Mizuno, 2007; Şenel, 2006) after the instructor modeled the tokens prior to their repetition (Trofimovich & Gatbonton, 2006; Yoshida & Fukada, 2014).

Participants were then given a worksheet with review questions, asking for them to identify differences about the production of the short-lag vs. the long-lag variants. Questions included on the worksheet asked participants which IPA symbol was used for which variant, how to identify whether or not they were producing large amounts of aspiration ("Put your hand in front of your mouth, then produce the word ____"), and if they believed a NS of Spanish noticed the difference in aspiration produced by a NNS in word-initial position. Examples from the in-class worksheet can be seen below in Table 6 (for the full worksheet, see Appendix F):

Table 6 *Sample Questions from EI In-Class Worksheet*

Sample	English Translation
¿Cómo es la producción de ´p´ diferente en español que el 'p' en inglés en posición inicial de una palabra?	*How is the production of 'p' in Spanish different from the ´p´ in English in word initial position?*
¿Cómo es la articulación de 'p' en español en posición inicial de una palabra?	*How is the articulation of ´p' in Spanish produced in word initial position?*
¿Qué es una prueba para ver si una persona aspira su ´p´?	*What is a test to see if someone aspirates their ´p´?*

At the end of the worksheet, participants found a list that contained five words beginning with the respective allophone being introduced that particular day. They were asked to repeat each word five times to their partner to get corrective feedback (Saito, 2011) from their peers and the instructor. At the close of the activity, participants were asked to create sentences using these five tokens and were then asked to read these sentences aloud to each other, attempting to produce the short-lag variant in a string of speech. As students were carrying out this final step, the instructor circled the room giving individual corrective feedback (Dlaska & Krekeler, 2013).

3.3.2 **Visual Feedback Group**

Prior to implementing any training, this group first completed the pretest tasks, including recording of production tasks 1-4 (detailed above), and perceptual tasks 1 and 2. After these initial tasks, participants in the visual feedback (VF) group took part in a guided VF treatment.

There were a total of 3 training sessions, one focused on each of the voiceless stops (/p,t,k/). Each training session consisted of: (a) pre-recording, (b) in-class analysis, and (c) re-recording. For the pre-recordings, participants were instructed to record tasks 1-2, send a digital copy of the audio files to their course instructor, and bring personal laptops or tablets to the class to observe the spectrogram and waveforms of four tokens with the target phoneme from the carrier phrase task. The recording of tasks 1 and 2 was expected to take approximately 5 minutes, and the recording of tasks 1-4 (for the pretest and posttest) was estimated to take about 15 minutes.

Visual Feedback Treatment

For the in-class analysis, participants were provided with sound waves and spectrograms (Kartushina et al., 2015; Motohashi-Saigo & Hardison, 2009; Offerman & Olson, 2016; Olson,

2014b; Saito, 2007; Wang & Munro, 2004) depicting NS productions of the Spanish short-lag variations [p,t,k] (Lisker & Abramson, 1964), as well as productions of NSs of American English and the long-lag variations [p^h, t^h, k^h] (Hualde, 2005) for a comparison of these short-lag and long-lag segments in word-initial position (Offerman & Olson, 2016; Olson, 2014b).

To direct participants toward the difference between the two languages, participants were provided with a worksheet that contained guided questions to aid participants in hypothesizing about the difference between these allophones produced by NSs of Spanish as well as NNSs of Spanish when producing utterances in Spanish (Offerman & Olson, 2016; Olson, 2014b; Olson, 2019). The handout is comprised of a series of questions (examples in Table 7; for the full worksheet, see Appendix G) leading the participants to conclusions about how their productions of [p^h], for example, differ from productions of [p] produced by a NS of Spanish. In other words, participants took part in a form of guided, implicit feedback in the VF group.

Figure 5 also provides an example of the word *Paco* (/pa.ko/), which was included on the worksheet without the boundaries marked for participants, as they were expected to find the boundaries for /p/ and /a/ on their own.

Table 7 *Sample Questions from VF In-Class Worksheet*

Sample	English Translation
¿Cómo puedes identificar tu vocal (vowel) 'a' de la consonante 'p'?	*How do you identify your vowel 'a' from the consonant 'p'?*
¿Es la 'a' más oscura o clara que la 'p'?	*Is the 'a' darker or lighter than the 'p'?*
¿Cómo es la 'p' del hablante nativo en el ejemplo?	*What is the 'p' of the native speaker like in the picture?*
Ahora marca los límites de tu 'p' y tu 'a.'	*Now, mark the boundaries for your 'p' and your 'a.'*

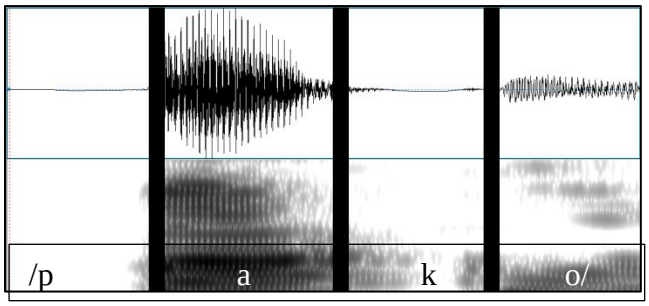


Figure 5 Paco produced by a NS of Spanish & segmented

After completing the worksheets and drawing their own conclusions, the instructor then provided participants with pictures of production samples of both NSs and NNSs in a PowerPoint slide, along with the audio recordings of both the NS and the NNS. The instructor then asked the participants to hypothesize about what was visually different between the sound waves, spectrograms, and sound clips of the NS and NNS. This was done by asking participants what they noticed about the beginning of a stop production (not specifically stating or identifying stop features) between a NS and a NNS. Then participants were directed to the spectrogram for the onset of a plosive, being asked to identify what they thought the difference was between the "dark shading closer to the top" and the "absence of dark shading at the top" of a production. Finally, sound clips were played again, along with the sound waves and spectrograms, to further demonstrate differences.

After participants collectively concluded that there was a difference in VOT values, the instructor then showed on a PowerPoint slide with these differences and had participants describe the differences. Descriptions of the sound waves and spectrograms were expressed in simpler terms such as "bigger" or "smaller", "longer" or "shorter", "no pattern" or "regular pattern", and "dark" or light". Then recordings of both the NS and NNS were played again for the participants that accompanied the PowerPoints slides. Participants were then asked to segment their own recordings, asked to indicate where the burst started and where the onset of the vowel began on their personal computers or tablets. After segmenting their own tokens, participants received guided instruction in which they compared their productions with similarly segmented NS productions.

For the first training session involving /p/, 15-20 minutes was allotted in case participants had extra questions, with training involving /t/ and /k/ being 15 minutes. For the pretest (before all trainings) and posttest only (after all trainings), participants then took part in the perception tasks (the discrimination and nativeness judgment tasks). The two perceptual tasks together totaled approximately 20 minutes for the pretest and about 15 minutes for the posttest.

For the re-recording component of each session, following the in-class analysis, participants were instructed to again record tasks 1-2, send a digital copy of the audio files to their course instructor.

3.3.3 Combination Instruction: Explicit Pronunciation + Visual Feedback Group

For the combination instruction (CI) group, all tasks are identical to those of the EI group and the VF group, with the exception of the treatment type. In the CI group, all participants took part in the perception pretest once, as well as all recording tasks (pretest tasks, phoneme of focus tasks, posttest tasks), ending with the perception posttest once. The CI group received a type of combination training and instruction, which consisted of procedures carried out in the above subsections.

Combination Instruction

For this type of instruction, participants were first explained the difference between the long-lag vs. short-lag variants with the respective IPA symbols for the designated phoneme of focus for that day, along with an explanation of articulatory differences. Participants were shown sounds waves and spectrograms of both NSs and NNSs while simultaneously being explicitly instructed on the difference in sound wave and spectrogram features, for [p] and [p^h], for example. The instructor additionally provided modeling of the sounds, along with showing NS and NNS productions with segmentations already completed to demonstrate these differences more directly. Such differences in sound waves and spectrograms were explained by also providing the appropriate IPA symbol of each allophone next to each production shown in a PowerPoint slide.

This group was also asked to bring laptops or tablets in order to have access to their productions of the first four tokens from the carrier phrase task to begin the treatment for /p/, much like the VF group. Unlike the VF group, participants in the CI group were asked to segment their own productions before taking part in the in-class worksheet, due having been explicitly instructed on how to do so, and were allowed to ask for and receive explicit clarification from the instructor.

After this explanation, the instructor distributed a worksheet, which was comprised of a combination of the EI and VF worksheets. The worksheet for the CI group consisted of questions asking participant to compare NS productions to their own productions. Samples from the CI group's worksheet can be seen below in Table 8 (for the full worksheet, see Appendix H):

Table 8 *Sample Questions from CI In-Class Worksheet*

Sample	English Translation
¿Cómo puedes identificar tu vocal (vowel) 'a' de la consonante 'p'?	*How do you identify your vowel 'a' from the consonant 'p'?*
¿Cómo es la articulación de 'p' en español en posición inicial de una palabra?	*How is the articulation of 'p' in Spanish produced in word initial position?*
¿Cómo puedes diferenciar entre la 'p' en inglés y la 'p' en español? ¿Cuál manera es más fácil para distinguirlos?	*How can you differentiate between the 'p' in English and the 'p' in Spanish? What is the easiest way to distinguish these?*

To conclude this activity, participants in this group were also asked to replicate the model the instructor provided and to repeat the five tokens, which was also contained in the EI worksheet. They were also asked, like the EI group, to create utterances containing the five tokens, to give peer feedback on pronunciation, while the instructor also circulated the room to provide individual feedback. For the CI treatment, this was matched for time to coincide with the time allotted for the EI and VF treatments; 20 minutes for the first treatment (/p/), and 15 minutes for the following two treatments (/t/ and /k/).

3.3.4 **Control Group**

For the control group (CO), participants took part in all perception and production tasks for the pretest, phoneme of focus tasks, along with the posttest tasks. The only major change for this group was that they did not participate in any sort of production treatment. To provide L2 learners with another type of educational material, this group took part in a cultural reading activity coupled with comprehension questions.

Cultural Activity

Participants in the CO group were given in-class activities with an equal duration to those that were in the EI, VF, and CI groups. The activities consisted of three different class sessions in which participants took turns reading aloud a cultural excerpt from the textbook to each other, along with answering a set of comprehension questions about the reading. The cultural topic was related to the class, and participants were told that the goal of this read-aloud task was for them to practice reading and comprehension; no explicit instruction nor any form of corrective feedback was provided by the instructor.

This group participated in the first cultural activity for 15-20 minutes, and the second and third cultural activities consisted of 15 minutes each (see Appendix I for an example). As completed by the three experimental groups, participants from the CO group re-recorded the list of tokens within the carrier phrase and sentence environment after each treatment for /p,t,k/ to be sent to the instructor. This group received no in-class instruction or guidance before submitting their recordings of tokens within any of the tasks nor after the cultural readings and comprehension questions; they received normal, daily classroom instruction, and only took part in a read-aloud task coupled with contexting questions.

As with all the experimental groups, also mentioned in the previous paragraph, the CO group took part in the first cultural reading activity for 15-20 minutes. This was done to mirror the amount of time spent by experimental groups that took part in one of the three treatments. Therefore, the second and third cultural readings were instructed to last 15 minutes. As with the perceptual tasks for the experimental groups, both perception tests for the CO group totaled 20 minutes for the pretest and 15 minutes for the posttest, with the pretest allowing more time in case participants had questions.

3.3.5 Training for Instructors

It should be noted that each instructor that was asked to conduct one of the three pronunciation treatments or the cultural activity was provided with an hour-long training session for the production tasks and production activities (in-class), as well as a 30-minute training session for the perceptual tasks. All trainings for instructors were conducted by the researcher, and after explanation, all instructors were asked to practice the tasks with the researcher, as well as encouraged to ask the researcher any clarification questions.

In addition, the researcher designed all in-class activity worksheets, as well as the all PowerPoint slides for the three pronunciation instructional groups. For the PowerPoint slides, the researcher provided a detailed step by step process on how to deliver the specific pronunciation instruction for each experimental group (EI, VF, or CI), as well as modifying the PowerPoints with the instructors if there appeared to be any confusion on how to execute the pronunciation treatment. The researcher went over the trainings with each instructor a second time for the focus phoneme of /t/, and then asked each instructor if they would like a refresher for the last focus phoneme /k/. For the CO group, the researcher also met with this instructor for the when the phoneme of focus

was /p/ and /t/, then leaving it up to the instructor for additional instruction once the focus phoneme became /k/.[8]

3.3.6 Conclusion of Procedures

To conclude this section for the production methods, there were three experimental groups that received pronunciation instruction; the first explicit pronunciation instruction (EI) only, the second received visual feedback (VF) only, and the third received combination instruction that consisted of explicit pronunciation plus visual feedback (CI). For the fourth group, the CO group, the participants took part in a read-aloud task, which consisted of a cultural reading from the university-assigned textbook, and they received no pronunciation training or feedback. All groups took part in both the pretest and posttest for the perceptual tests, as well as completing the recordings of all of the same tasks for the appropriate activity (pretest, phoneme of focus activities, posttest). In addition, all groups were estimated to have spent an equal amount of time for each of the tasks and training or cultural reading.

3.4 Data Analysis

3.4.1 Production Data Analyses

Voice onset time (VOT) values for each target token were analyzed and measured in Praat (Boersma & Weenink, 2019) (a more detailed description can be found in Chapter 4). Statistical analyses were carried out using RStudio (Allaire, 2012) using the lme4 package (Bates et al., 2014). For all mixed-effects models, the significance criterion was set at $|t| > 2.00$. Statistical analyses will be detailed in Chapter 4 for the production tasks.

3.4.2 Perception Data Analyses

For the discrimination task (AXB task), answers indicated by the participants were counted as either correct or incorrect (0 or 1 point), and a total percentage correct was calculated for each participant. For nativeness judgment task (the 7-point, Likert-scale task), ratings by L2 learners were totaled and averaged for each recording (24 total) presented to the participants (each

[8] None of the four instructors asked for a third training session for the phoneme of focus /k/.

production was to be rated as more native-like or more non-native-like). Statistical analyses were carried out using RStudio (Allaire, 2012) and the lme4 package (Bates et al., 2014). For all mixed-effects models, the significance criterion was set at $|t| > 2.00$. Statistical analyses for the perception tasks will be detailed in Chapter 5.

3.5 Conclusion of Methods

In sum, this study investigates the outcomes of three different approaches to pronunciation training, EI, VF, and CI in relation to both L2 production and L2 perception domains. Production-oriented tasks range from more constrained (carrier phrase) to more spontaneous (picture task) tasks in order to observe improvements (if any) in more controlled as well as more spontaneous speech. The results of the production tasks are presented in Chapter 4.

Perceptual tasks seek to determine if participants can discriminate between the two allophonic variations of the stops /p,t,k/, short-lag and long-lag variants, for the purpose of identifying to what extent participants improved in perception of these allophones after production training. It was also sought to determine what social value they perceive and ascribe to such productions of these allophones in the nativeness judgment task via a Likert-scale rating. The results of the perception tasks are presented in Chapter 5.

CHAPTER 4. PRODUCTION RESULTS & DISCUSSION

The current chapter focuses on the data coding, statistical analyses, and interpretation of the data analyses for the four tasks mentioned previously in Chapter 3: tokens in a carrier phrase (Task 1), embedded in novel sentences (Task 2), in a controlled, continuous speech task (Task 3), and in a spontaneous speech task (Task 4). Again, the four groups consisted of: explicit instruction (EI), visual feedback (VF), combination instruction (CI), and a control group (CO). Data from all four groups were analyzed for each of the four tasks at the pretest and posttest and are examined in detail in this chapter.

Each type of training is examined with regard to being matched at the pretest, individual group improvement, as well as between-group improvement; that is to say, if all groups perform equally at the pretest, if there were any significant interactions at the posttest, along with looking at each group's performance individually from pretest to posttest. Additionally, the discussion of these results considers the effectiveness of the treatments relative to the CO group and offers a comparative analysis of each of the experimental treatments. Implications are drawn in relation to theory of L2 phonetic production.

4.1 Data Analysis

A total of 47 participants took part in four tasks of 30 tokens at two times for the pretest and the posttest, resulting in a total of 11,280 tokens (47 participants × 4 tasks × 30 tokens × 2 sessions = 11,280 tokens). Of these 11,280 tokens, 442 (3.9%) were eliminated from the final analysis for recording errors or issues (e.g. background noise, producing the incorrect word[9], laughing, interruptions, poor recording quality, or not completing the entirety of a task). An additional 286 (2.6%) were identified as outliers, defined as three standard deviations above or below the mean. For calculating outliers, the mean was taken for each group at each time (pretest and posttest) for each task. For example, the mean was taken for the CO group once at the pretest and once at the posttest for the carrier phrase task, as well as at the pretest and posttest for the

[9] In multiple instances, participants confused target tokens with other words in Spanish (e.g. producing 'poco' in place of 'paco'), and these were eliminated.

remaining tasks. Again, this was also done for the EI, VF, and CI groups. The final total of tokens eliminated was 728 (6.5%). For the final analysis, a total of 10,552 were used.

For each of the following analyses, a mixed-effects model was run evaluating normalized VOT values (i.e. normalized for voice onset time [VOT], see below). The statistical program RStudio (Allaire, 2012) was used to carry out the analyses. Further, the lme4 package (Bates et al., 2014) was used to conduct all mixed-effect models. For the mixed-effects model, *time* (pretest or posttest) as well as *group* (EI, VF, CI, CO) were included as fixed effects, while *participant* (subject) and *token* (item) were set as random effects with random intercepts[10]. Effect sizes were also calculated for all groups, comparing results from the CO group at the pretest to the CO group at the posttest and to results from each experimental group at the pretest and the posttest. This was done to assess the magnitude of effect utilizing Cohen's *d*. Incorporating the findings of Plonsky and Oswald (2014), interpretation criteria for the value of *d* is set at: small = 0.40, medium = 0.70, and large = 1.00.

All data points were measured by hand via Praat (Boersma & Weenink, 2019) to record VOT values. The onset of voicing was determined as initial vocal fold vibration relative to the release of a stop (Abramson & Whalen, 2017). Additionally, given the differences in VOT by place of articulation (Cho & Ladefoged, 1999), a normalized value was calculated for all productions of /p,t,k/ (Olson, 2019). Specifically, voiceless stop consonants with different places of articulation in word-initial position (in this case: bilabial, dental, velar) differ in VOT values. For /p/, in Spanish, the mean is 4ms, with English being 58ms; the average for /t/ is 9ms in Spanish and 70ms in English, with /k/ being averaged at 29ms in Spanish and 80ms in English (Lisker & Abramson, 1964). Following Olson's (2019) normalization procedure, each participant production was normalized by subtracting the mean reported Spanish VOT value for a given phoneme (Lisker & Abramson, 1964) from the actual participant-produced VOT value and dividing by the difference between reported English and Spanish values (Lisker & Abramson, 1964). As such, a value of 0 was assigned to participant productions that were equal to previously reported mean Spanish VOT

[10] Initially, a maximal random effects structure was included in the model, with random intercepts and slopes by time by group (Barr et al., 2013). Singularity or convergence issues indicated that the random effects structure needed to be simplified. This was done by stepwise simplification of the random effects structure. For each of the four tasks, the initial model was run; however, the maximal random effects structure that permitted convergence was random intercepts only for each of the four tasks. This model was used for all four tasks, as it was the maximal model that could produce the random slopes and intercepts without any convergence errors (Barr et al., 2013).

values by place of articulation (e.g., 4ms for /p/ = 0; 9ms for /t/ = 0; 29ms for /k/ = 0). Similarly, A value of 1 is assigned to mean English VOT values (e.g., 58ms for /p/ = 1; 70ms for /t/ = 1; 80ms for /k/ = 1). Any value found in between the average Spanish and English values in milliseconds is assigned a value between 0 and 1. Any value above the English averages is assigned a value greater than 1, along with a value less than the average values for Spanish is assigned a negative value. For example, if /p/ is produced with a raw VOT of 45ms, this corresponds to a normalized value of 0.76. If /k/ is produced with a raw VOT value of 45ms, corresponds to a normalized value of 0.31. The normalization procedure allowed for data to be collapsed across all three phonemes with different places of articulation.

4.1.1 Carrier Phrase Task

For the carrier phrase task, a mixed-effects model was conducted to determine (1) if groups were well-matched at the pretest, (2) if there were any significant interactions at the posttest, and (3) if any group(s) individually made significant improvement from pretest to posttest. In assessing these, observance of the *t* value is crucial; the significance criterion was set at $|t| > 2.00$.

Results of the initial model (Table 9), specifically comparisons between the intercept (CO group, pretest) and each of the other groups at the pretest, indicated that there were no significant differences in the groups at the pretest. As can also be seen in Table 9 when observing performance at the posttest, all experimental groups (EI, VF, CI) outperform the CO group. With these values, each of these groups is observed to have improved significantly for the carrier phrase task in reducing the VOT values when compared to the CO group at the posttest. Considering the effect of time, the results demonstrated that there was no significant difference between the intercept (CO, pretest) and the CO group at the posttest, suggesting that the CO group did not improve in the production of VOT over time. However, there was a significant interaction for each of the experimental groups at the posttest, such that there was a significant difference between the intercept (CO, pretest) and each of the experimental groups at the posttest.[11]

[11] For within-group analyses, not pictured in Table 9, the EI group (β = -0.175, SE = 0.028, t = -6.175), VF group (β = -0.077, SE = 0.029, t = -2.686) and the CI group (β = -0.337, SE = 0.024, t = -13.977) all improved significantly from the pretest to the posttest.

Table 9 *Results for the Carrier Phrase Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	0.672	0.085	[0.502, 0.842]	7.900	-
EI	-0.071	0.115	[-0.301, 0.159]	-0.616	0.185
VF	0.151	0.118	[-0.085, 0.387]	1.276	0.296
CI	0.156	0.106	[-0.056, 0.368]	1.475	0.390
Posttest: CO	0.022	0.029	[-0.036, 0.080]	0.755	0.020
Posttest: EI	-0.197	0.040	[-0.277, -0.117]	-4.896	0.528
Posttest: VF	-0.099	0.041	[-0.181, -0.017]	-2.440	0.131
Posttest: CI	-0.359	0.037	[-0.433, -0.285]	-9.605	0.386

While the initial model set the intercept as the CO group at the pretest, it is further worth comparing the performance of the experimental groups to each other. As such, a subsequent model was run with identical parameters, but with the CI group at the pretest as the intercept (not pictured in Table 9). An analysis of the interaction between group and time revealed that the CI group outperformed both the EI (β = 0.162, *SE* = 0.037, *t* = 4.332) and VF (β = 0.260, *SE* = 0.036, *t* = 6.919) groups. With regard to the EI and VF groups at the pretest being set as the intercept, results indicate that the EI group outperformed the VF group (β = 0.098, *SE* = 0.040, *t* = 2.429) at the posttest, and the VF group did not outperform either experimental group. In looking at the effect sizes, we see that all treatments had a small to medium effect, ranging from 0.10 to 0.53.

Figure 6 displays each group's normalized VOTs at the pretest and posttest for the carrier phrase task. In observing the results, it is concluded that while all forms of pronunciation instruction resulted in improvement (i.e., reduction of VOT) in the carrier phrase task, this effect was most pronounced for the CI group.

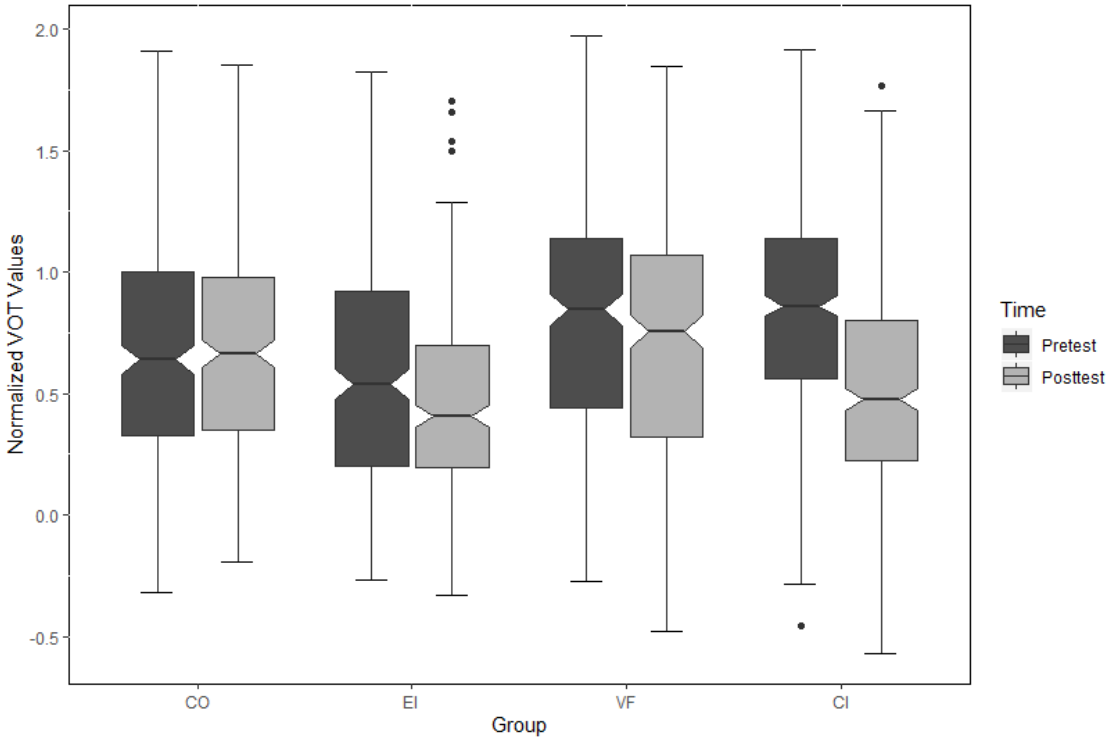


Figure 6 Normalized voice onset time by group and time for carrier phrase task.

4.1.2 **Novel Sentences Task**

For the tokens embedded in the utterance task, a mixed-effects model was run again to analyze: (1) if groups were well-matched at the pretest, (2) if there were significant interactions at the posttest, and (3) if groups made any significant improvement from pretest to posttest. In looking at Table 10, we observe that there are no significant differences between the intercept (CO, pretest) and any of the other groups at the pretest, illustrating that all groups are matched at the pretest. In other words, all groups produced similar VOT values prior to any form of intervention. Considering the interaction between group and time, Table 10 illustrates that the EI, VF, and CI groups again outperformed the CO group at the posttest.[12]

[12] As seen in the previous task, the EI (β = -0.162, *SE* = 0.028, *t* = -5.691), the VF (β = -0.089, *SE* = 0.028, *t* = -3.113) and CI (β = -0.234, *SE* = 0.023, *t* = -9.988) groups made a significant improvement for the within-group analyses.

Table 10 *Results for the Novel Sentences Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	0.631	0.090	[0.451, 0.811]	7.026	-
EI	-0.064	0.115	[-0.294, 0.166]	-0.558	0.147
VF	0.097	0.118	[-0.139, 0.333]	0.824	0.197
CI	0.146	0.107	[-0.068, 0.360]	1.365	0.315
Posttest	0.013	0.027	[-0.041, 0.067]	0.471	0.022
Posttest: EI	-0.174	0.039	[-0.252, -0.096]	-4.429	0.434
Posttest: VF	-0.101	0.039	[-0.179, -0.023]	-2.574	0.008
Posttest: CI	-0.247	0.036	[-0.319, -0.175]	-6.860	0.219

Considering the interaction, in assessing which, if any, group improved the most, we begin by placing the CI group at the pretest as the intercept to observe interactions at the posttest. It was found that the CI group at the posttest (β = 0.234, *SE* = 0.028, *t* = -9.988) outperformed the VF group (β = 0.145, *SE* = 0.037, *t* = 3.937). However, the CI group for this particular task did not outperform the EI group (β = 0.072, *SE* = 0.037, *t* = 1.962), unlike in the previous task. As for the EI and VF groups at the pretest being set as the intercept, the EI group (β = -0.162, *SE* = 0.028, *t* = -5.691) outperformed the VF group (β = 0.073, *SE* = 0.040, *t* = 1.813) at the posttest. Although we saw initially that the VF group significantly improved in Table 10 with the CO group as the intercept, they did not outperform either of the other two experimental groups. In relation to effect sizes, there was a small effect for treatment for all experimental groups.

A representation of the results can be found in Figure 7 for the novel sentences task, with each group's performance at the pretest and posttest. Taken as a whole, these results show that, again, while all forms of pronunciation instruction resulted in improvement (i.e., reduction of VOT) in the novel sentences task, this effect was most pronounced for the CI group.

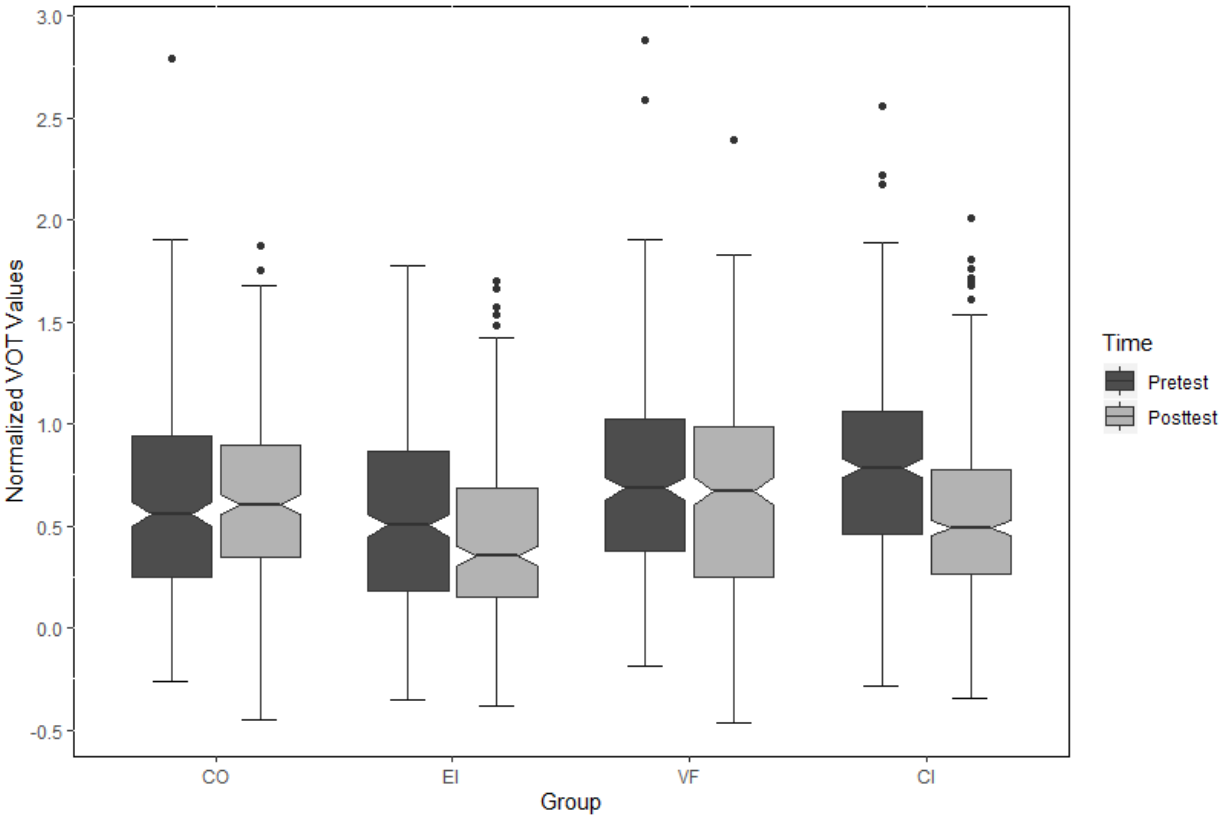


Figure 7 Normalized voice onset time by group and time for novel sentences task.

4.1.3 **Controlled, Continuous Speech Task**

Once again, a mixed-effects model was carried out for the controlled, continuous speech task in the form of a short story. Again, this was run to analyze: (1) if groups were well-matched at the pretest, (2) if there were significant interactions at the posttest, and (3) if individual groups made any significant improvement from pretest to posttest. As seen in Table 11, all groups are matched at the pretest when compared to the intercept (CO, pretest). While there is no significant interaction between groups at the pretest, we do see significant interactions between group and time, such that each experimental group outperformed the CO group at the posttest. [13]

[13] For within-group analyses, we see the same patterns in the statistics that are mentioned in 4.1 and 4.2. As seen in Table 11, the EI (β = -0.147, *SE* = 0.028, *t* = -5.217), VF (β = 0.100, *SE* = 0.028, *t* = -3.534), and CI groups (β = -0.218, *SE* = 0.024, *t* = -9.081) improved significantly from the pretest to the posttest, while this was not the case for the CO (β = 0.011, SE = 0.027, t = 0.417) group.

Table 11 *Results for the Controlled, Continuous Speech Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	0.659	0.077	[0.505, 0.813]	8.593	-
EI	-0.053	0.108	[-0.269, 0.163]	-0.486	0.144
VF	0.090	0.108	[-0.126, 0.306]	0.826	0.207
CI	0.145	0.098	[-0.051, 0.341]	1.472	0.332
Posttest: CO	0.011	0.027	[-0.043, 0.065]	0.417	0.009
Posttest: EI	-0.158	0.039	[-0.236, -0.080]	-4.047	0.519
Posttest: VF	-0.111	0.039	[-0.189, -0.033]	-2.837	0.034
Posttest: CI	-0.229	0.036	[-0.301, -0.157]	-6.323	0.168

Again, considering the possible differences between each of the experimental groups, a separate model was conducted with the CI group as the intercept. Results of the interactions between time and group suggest that the CI group outperformed the VF group (β = 0.118, *SE* = 0.037, *t* = 3.181). A comparison of the CI group and the EI group showed a similar, but non-significant trend, such that the CI group showed a greater reduction in VOT values than the EI group (β = 0.071, *SE* = 0.037, *t* = 1.907). With regard to effect sizes for the initial model, there is a small to medium effect for the EI group, and a small effect for both the VF and CI groups.

The graph in Figure 8 below illustrates the normalized VOT values at the pretest and posttest for all groups. As seen in the previous tasks, all types of intervention resulted in improvement (i.e., reduction of VOT) for the controlled, continuous speech task. Also, this effect was again most pronounced for the CI group, followed by the EI and VF groups.

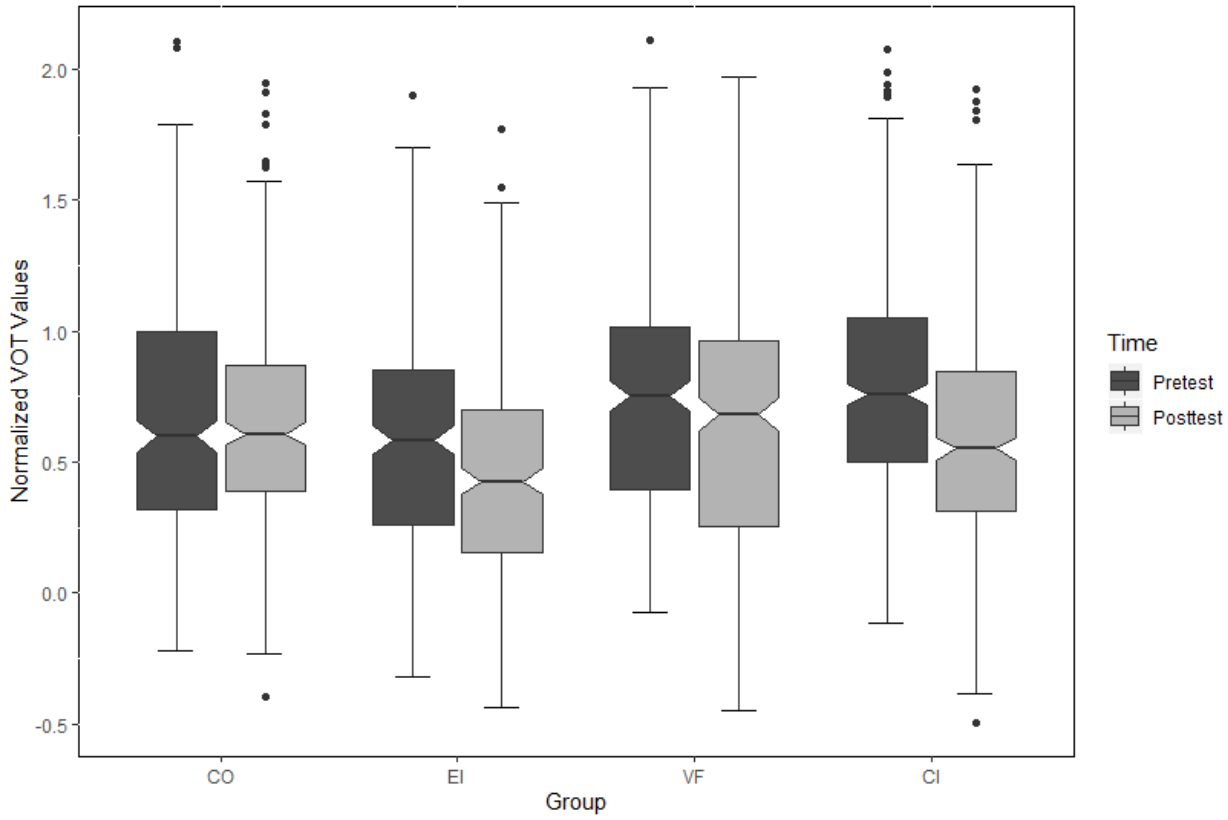


Figure 8 Normalized voice onset time by group and time for controlled, continuous speech task.

4.1.4 **Spontaneous Speech Task**

Finally, a mixed-effects model was conducted for the spontaneous speech task in the form of a picture task to, again, analyze group performance in looking at: (1) if groups were well-matched at the pretest, (2) if there were significant interactions at the posttest, and (3) if individual groups made any significant improvement from pretest to posttest. Illustrated in Table 12, the three experimental groups when compared to the CO group are matched, meaning that all groups performed equally at the pretest. In reference to the posttest performance for all groups, there exist significant findings for all three experimental groups (EI, VF, CI), with the CO group showing a significant decline in performance at the posttest. [14]

[14] The statistical analyses from the mixed-effects model from each of the experimental groups at the pretest as the Intercept show that the EI (β = -0.095, *SE* = 0.027, *t* = -3.551), VF (β = -0.146, *SE* = 0.028, *t* = -5.158), and CI (β = -0.195, *SE* = 0.024, *t* = -8.061) groups significantly improved from the pretest to the posttest, with no indicated significant improvement for the CO (β = 0.093, *SE* = 0.030, *t* = 3.135) group.

Table 12 *Results for Spontaneous Speech Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	0.626	0.079	[0.468, 0.784]	7.919	-
EI	0.007	0.107	[-0.207, 0.221]	0.068	0.057
VF	0.148	0.107	[-0.066, 0.362]	1.387	0.289
CI	0.078	0.097	[-0.116, 0.272]	0.799	0.120
Posttest	0.093	0.030	[0.033, 0.153]	3.135	0.148
Posttest: EI	-0.188	0.040	[-0.268, -0.108]	-4.708	0.289
Posttest: VF	-0.239	0.041	[-0.321, -0.157]	-5.830	0.156
Posttest: CI	-0.288	0.038	[-0.364, -0.212]	-7.524	0.357

For this task, it is worth noting that the CI group (β = -0.195, *SE* = 0.024, *t* = -8.061) outperformed the EI group (β = 0.099, *SE* = 0.036, *t* = 2.752), with there being no significant interaction with respect to the VF group (β = 0.049, *SE* = 0.037, *t* = 1.309) at the posttest. Further, with the EI group at the pretest set as the intercept, they did not outperform the VF, as seen in other tasks. Likewise, with the VF group at the pretest set as the intercept, they did not outperform either of the other two experimental groups. Referencing the effect sizes, it appears that all treatments had a small effect.

In line with the other production tasks, all treatments resulted in improvement (i.e., reduction of VOT) for the spontaneous speech task. Further, this effect was, again, most pronounced for the CI group. For a visual comparison of each group at both the pretest and the posttest, refer to Figure 9 below.

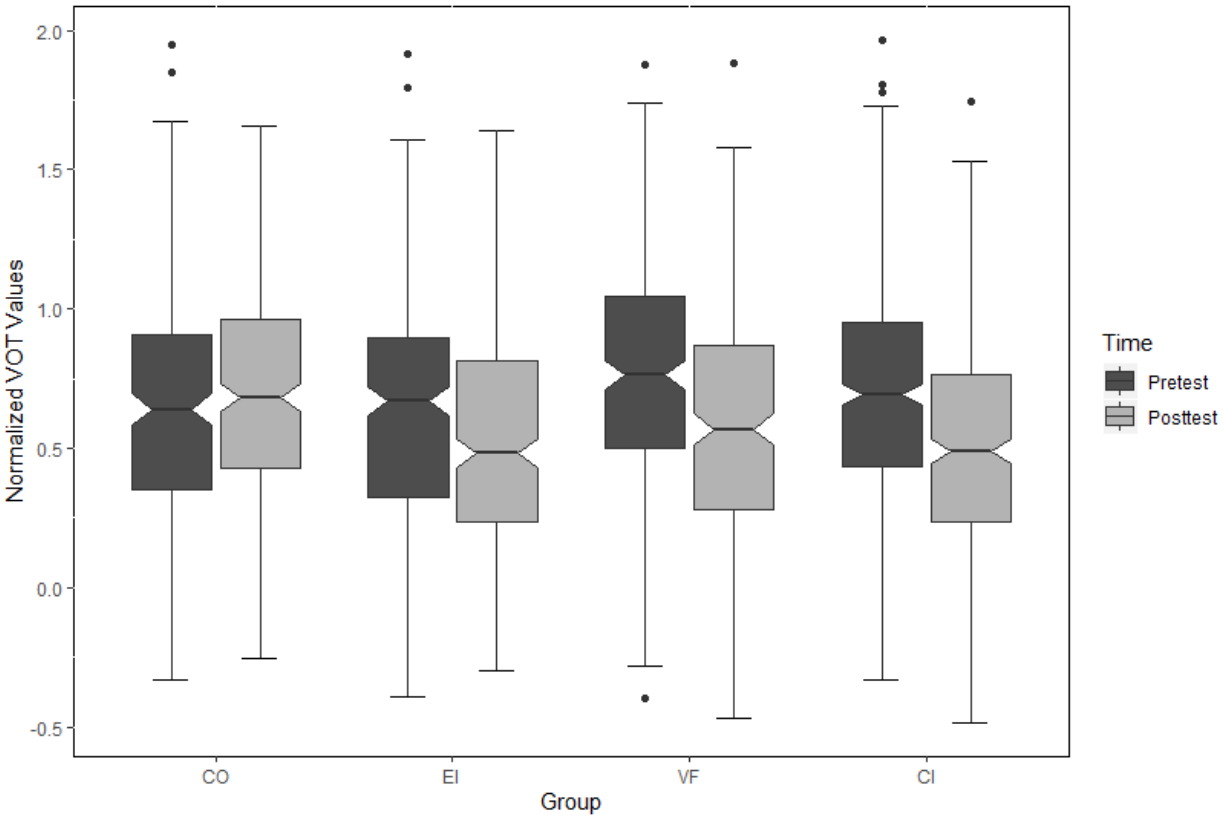


Figure 9 Normalized voice onset time by group and time for the spontaneous speech task.

4.1.5 Conclusion of Results

The current section, 4.1 Analysis & Results, covers the statistical analyses run via mixed-effect models to elaborate on (1) whether groups were well-matched at the pretest, (2) if any significant interactions existed at the posttest, and (3) the effect of time on all experimental groups from pretest to posttest. The interpretation of the results as well as how they relate to the research questions will be discussed in detail in section 4.2 below.

4.2 Discussion of Production Results

Two research questions addressed the production-oriented portion of this study, examining both the efficacy of each treatment type individually (RQ1) and the comparison between each of the experimental groups (RQ1.1). These research questions are discussed in both 4.2.1 (a within-group comparison) and 4.2.2 (a between-group comparison). A discussion of the treatment effectiveness is assessed via group performance, with improved, more Spanish-like VOT values

indicated via a reduction in VOT values for /p,t,k/. Finally, this section will provide a more detailed examination of the theoretical implications of the production portion of this study, while addressing the gaps in the literature previously outlined in Chapter 2. Literature Review.

4.2.1 Treatment Effectiveness – Comparison within-group

The first research question addresses the effectiveness of each treatment on the experimental groups; in other words, if each treatment-type was successful in aiding participants in reducing their VOT values after intervention.

(RQ1) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of L2 phonetic production? In this case, improvement of phonetic production is defined as a reduction in VOT by English speaking L2-learners of Spanish in the onset position of a word.

> Hypothesis 1: Drawing on findings for explicit pronunciation studies (e.g., Aliaga-García & Mora, 2009; Elliott, 1997; Lord, 2010), visual feedback studies (e.g., Motohashi-Saigo & Hardison, 2009; Olson, 2014b), and a combination approach (Kartushina et al., 2015), it is hypothesized that all experimental groups will significantly reduce their VOT of the word-initial voiceless plosives in Spanish after having received a form of pronunciation training.

In addressing RQ1, it can be seen that the three experimental groups (EI, VF, CI) and the CO group were matched at the pretest for each task, suggesting that all groups performed similarly in terms of their productions of /p,t,k/ at the pretest. At the posttest, statistical analyses reveal that each experimental group significantly outperformed the CO group in every task. What is noteworthy here is that each experimental group was able to outperform the CO group significantly in multiple types of tasks, and not merely in a controlled task; each experimental group showed gains when compared to the CO group at the posttest both in more controlled tasks (e.g., carrier phrase task) and in less controlled tasks (e.g., spontaneous speech task).

Based on the findings for L2 studies, effect size ranges for small, medium, and large have been determined according to Plonsky and Oswald (2014).[15] Effect sizes for each group at the posttest of each task ranged from a small to medium effect, without reaching an effect size of 0.70 (the estimated median medium effect size) (Plonsky & Oswald, 2014). Although a small effect may imply that there might have been less of an impact with respect to treatment and task, it is crucial to consider the constraints of time and frequency when observing effect sizes. For all experimental groups, treatment occurred three times, once every two weeks for 15 to 20 minutes (20 minutes allotted for the first treatment), in addition to the eight, at-home recordings that took roughly 10 minutes. This totals to 50-60 minutes total for in-class trainings, with 80 minutes of at-home recordings, totaling to 130-140 minutes of the entirety of treatment and at-home work for each instructional-type. It has been suggested that if participants are exposed to a treatment for longer increments of time and/ or more frequent treatments over an extended period time, this may result in larger effect sizes (Plonsky & Oswald, 2014). Keeping in mind the brevity of each treatment along with limited exposure, small effect sizes in this particular case still demonstrate a significant impact of treatment on group performance.

Explicit Instruction

Previously detailed in Chapter 3, explicit instruction (EI) defined here incorporates the use of IPA symbols and an explanation of place and manner of articulation (Elliott, 1997; Lord, 2010; Miller, 2012; Rajab, 2013; Sturm, 2013a), modeling and repetitions (Trofimovich & Gatbonton, 2006; Yoshida & Fukada, 2014), and the incorporation of explicit corrective feedback (Dlaska & Krekeler, 2013; Saito & Lyster, 2012). It is important to reiterate that in many of these previous explicit pronunciation studies, participants achieved significant improvement post-intervention. For the EI group in this study that received the treatment delineated above for /p,t,k/, participants improved significantly from pretest to posttest for all four of the production tasks. As has been previously seen in other types of explicit instruction studies (e.g., Lord, 2010; Saito & Lyster, 2012; Yoshida & Fukada, 2014), it is concluded that this type of treatment is beneficial in aiding learners' in their pronunciation learning, more specifically involving segments such as /p,t,k/, and further supports explicit instructional methods. The results also demonstrate that participants are

[15] The study by Plonsky & Oswald (2014) determined appropriate effect sizes for broad L2 research and SLA studies; to date, there is no study that specifically addresses effect sizes for L2 pronunciation studies.

capable of improving in their productions of tokens in more controlled and less controlled scenarios.

Visual Feedback

Also reviewed, in Chapter 3, visual feedback consists of having participants compare their NNS soundwaves and spectrograms to those of NSs (Motohashi-Saigo & Hardison, 2009; Offerman & Olson, 2016; Olson, 2014b; Saito, 2007), while making hypotheses for a more implicit and inductive approach (Offerman & Olson, 2016; Olson, 2019). As noted with the EI group, the VF group showed significant improvement for all four tasks from pretest to posttest, indicating the ability to improve in more controlled and less controlled tasks. Multiple studies conducted on visual feedback paradigms in the past have also displayed learner improvement from pretest to posttest, with several studies implementing this implicit and inductive approach (Offerman & Olson, 2016; Olson, 2019) such as the one conducted in this study. Therefore, this study not only provides further evidence for significant improvement after treatment, but this method is further supported in directing L2 learners towards noticing differences in the L1 and the L2 (Olson, 2014b).

Combination Instruction

With respect to the literature and the current study, two previously tested treatments have been effective in aiding L2 learners in improving their L2 pronunciation, namely explicit instruction (e.g., Elliott, 1997; Kissling, 2013; Lord, 2010; Miller, 2012) and visual feedback (e.g., deBot, 1983; Offerman & Olson, 2016; Olson, 2014b). Moreover, it is also confirmed that the combination instruction approach (e.g., Kartushina et al., 2015; Lord, 2005) is an effective treatment type for learners, although few studies have incorporated this method of pronunciation teaching.

In observing the within-group results for the CI group, this group was found to improve significantly from pretest to posttest in all four tasks, as also illustrated by both the EI and VF groups in this study. This study supports the notion that a combined instructional approach such as this is effective in improving segmental productions of specific features in a L2.

Control Group

The CO group, however, showed no significant improvement from pretest to posttest for any task, even showing some significant decline in performance for the spontaneous speech task (picture task) at the posttest. With the results of the each experimental group compared to the results of the CO group at the pretest and the posttest, we can propose that, since all groups were well-matched at the pretest, all three types of intervention result in improvement at the posttest. This is in contrast to having L2 learners rely on input-only strategies to improve pronunciation skills (i.e. no type of instruction that specifically targets pronunciation, as often done in many common frameworks) (Derwing & Munro, 2005) since there was no significant improvement evidenced for the CO group over time.

4.2.2 Treatment Effectiveness: Between-group Comparison

Research question 1.1 addresses which experimental group, if any, outperformed the other experimental groups. The research question and accompanying hypothesis are reviewed below:

(RQ1.1) In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback [VF] training or a combination training [CI]) on L2 Spanish production?

> Hypothesis 1.1: Drawing on findings for explicit pronunciation studies (e.g., Aliaga-García & Mora, 2009; Elliott, 1997; Lord, 2010), visual feedback studies (e.g., Motohashi-Saigo & Hardison, 2009; Olson, 2014b), and a combination approach (Kartushina et al., 2015), it is hypothesized that L2 learners in both the EI and VF groups will improve significantly in their productions; however, there will be no statistically significant data showing that L2 learners improved more because of one, singular treatment over the other. They will instead improve more significantly with the combined training approach.

Explicit Instruction

Regarding the EI performance in comparison to the other two experimental groups (VF and CI), the EI group outperformed the VF group in only the carrier phrase task and outperformed the CO group in all tasks, with no evidence of outperforming the CI group in any task. The EI

group performed similarly to all other experimental groups for the other three remaining tasks (novel sentences task; controlled, continuous speech task; spontaneous speech task). Table 13 below summarizes the performance between groups at the posttest, with check marks indicating tasks in which the EI group outperformed the other groups:

Table 13 *EI Group Outperformance*

Outperformance of EI Group at Posttest	Carrier Phrase Task	Novel Sentences Task	Controlled, Continuous Speech Task	Spontaneous Speech Task
VF	√	-	-	-
CI	-	-	-	-
CO	√	√	√	√

Visual Feedback

Concerning between-group performance, the VF group only outperformed the CO group for all tasks. The VF group did not outperform the other two experimental groups for any of the four tasks. Table 14 (below) summarizes the between-group results when compared to the VF group at the posttest. Again, check marks demonstrate tasks in which the VF group outperformed other groups.

Table 14 *VF Group Outperformance*

Outperformance of VF Group at Posttest	Carrier Phrase Task	Novel Sentences Task	Controlled, Continuous Speech Task	Spontaneous Speech Task
EI	-	-	-	-
CI	-	-	-	-
CO	√	√	√	√

Combination Instruction

One striking difference with the CI group is the significant interaction at the posttest when compared to the EI and VF groups. In addition to outperforming the CO group for all tasks, the CI group outperformed the EI group at the posttest for the carrier phrase task and the picture task, along with outperforming the VF group for the carrier phrase task, the novel sentences task, and the controlled, continuous speech task. This difference is notable in that the EI group only

outperformed the VF group once, and the VF did not outperform any experimental group; however, the CI group outperformed both the EI and VF groups both in multiple tasks. It should also be noted that the other two experimental groups never outperform the CI group in any task. Table 15 is provided below to better illustrate the CI group's performance; a check mark indicates instances in which the CI group outperformed the remaining groups:

Table 15 *CI Group Outperformance*

Outperformance of CI Group at Posttest	Carrier Phrase Task	Novel Sentences Task	Controlled, Continuous Speech Task	Spontaneous Speech Task
EI	√	-	-	√
VF	√	√	√	-
CO	√	√	√	√

Comparison of Treatments

Relating back to the literature, a call in previous research asserted the need for a comparative analysis of instructional-types to investigate which method, if any, is most effective and beneficial to L2 learners (Derwing & Munro, 2015; Lee et al., 2015). Additionally, studies in recent years have not incorporated a comparative analysis of various methodologies like that of the current study (e.g., Aliaga- García & Mora, 2009, Lord, 2010; Miller, 2012; Offerman & Olson, 2016; Olson, 2014b; Ruellot, 2011). Additionally, noted by Lee et al. (2015) in their meta-analysis of different L2 pronunciation studies, many studies have also not included a control group for a basis of comparison to different experimental groups. This study sought to compare multiple methods to satisfy the call for a comparative analysis, as well as include a control group within the comparative analysis to serve as the basis of comparison for improvement when observing intervention efficacy as opposed to no intervention.

For this reason RQ1.1 was proposed, investigating whether one instructional type had more effect on L2 learner productions of /p,t,k/ than the two other methods and a control group. Although the CI group did not outperform the EI and VF groups in every task, it is evident that they outperformed the EI and VF groups in the majority of tasks, while performing similarly to the EI and VF groups in the remaining tasks (EI: novel sentences and controlled, continuous speech task; VF: spontaneous speech task). In sum, the findings for the CI group support the notion

that this instructional-type is most effective for L2 learners of Spanish in improving productions of segments such as word-initial /p,t,k/.

4.2.3 Theoretical Implications

The current section references one particular theory relating to L2 production and how this theory applies to the current study. The specific theory that is extrapolated on is the Noticing Hypothesis, which provides an explanation of how L2 learners are able to improve productions after being made aware of differences. Pedagogical implications for the production portions are discussed at the close of the study in Chapter 7.

The idea of *noticing* (Izumi & Bigelow, 2000; Schmidt, 1990; Schmidt, 1992; Swain & Lapkin, 1995) is considered as a potential explanation as to why participants in all experimental groups significantly improved in reducing their VOT values to more Spanish-like values. To elaborate, *noticing* refers to L2 learners being directed to focus their attention on a specific feature of the L2 in order to establish differences between the L1 and the L2 and modify their productions thereafter (Schmidt, 1992; Swain & Lapkin, 1995). Schmidt (1990) proposed the Noticing Hypothesis, stating that L2 learners need to direct their attention to specific aspects of the target language in order to better acquire these components of the L2. Derwing and Munro (2005) mention noticing in the broader sense, positing that drawing attention to other aspects of a L2, such as syntax, is equally important for drawing learner attention to other aspects of a language, such as pronunciation. Olson (2014) specifically discusses noticing as it relates to visual feedback, stating that a visual feedback paradigm may enable noticing for L2 learners that have limited auditory perceptual skills in the L2. While Tyler (2019) does not specifically highlight the Noticing Hypothesis, it was suggested that a form of explicit instruction could help L2 learners better perceive differences between the L1 and the L2, which would result in more target-like productions in the L2. Therefore, it is posited that perhaps L2 learners from the three experimental groups in this study improved in their productions via having their attention directed to the differences between English and Spanish for word-initial /p,t,k/ through the three different types of instruction.

Moreover, it is important to consider that the CI group consistently outperformed the EI and VF groups in the production portion of this study. In considering once again this notion of *noticing* through having L2 learners focus on specific aspects of the target language, it could be that, in place of one type of pronunciation instruction to have learners focus and notice different

language features (here specifically having L2 learners focus on differences between Spanish and English productions of /p,t,k/), a combined training aids learners in further enhancing their noticing skills. In previous explicit studies, participants have been directed to observe differences in sounds through learning place and manner of articulation via IPA symbols (Lord, 2005), along with modeling and repetitions (Yoshida & Fukada, 2014), and corrective feedback (Saito & Lyster, 2012). These explicit methods tend to rely more on the L2 learners' capacity to hear and reproduce the sounds that are presented to them, relying in part on the auditory modality. In visual feedback studies, learners have often been guided to look at visual representations of sounds via spectrograms and sounds waves (Olson, 2014b), which relies more on the visual modality.

For the combination approach, it could be that learners are not relying on only one type of modality (auditory vs. visual) (Baran-Łucarz, 2012; Murphy, 1997) to draw their attention to differences, but they are now relying on two types of modalities in order to better notice the differences between the L1 and the L2. Knowing that both types of modalities have been used for directing L2 learners to differences in the past (Baran-Łucarz, 2012; Murphy, 1997), it could be that this combination of both auditory and visual modalities provides reinforcement and is a more beneficial tool for adult L2 learners to notice different aspects in their L2 that are novel or different from their L1. Another possibility could be that different learners benefit more from one type of modality over the other (auditory vs. visual), and as a group, the CI group outperformed other groups simply because both options were made available to them (see Chapter 7 for more discussion). However, these possibilities would need to be further tested, involving more phonemic features of Spanish and English, as well as other language pairings for different L2 learners.

4.2.4 **Conclusion of Discussion**

As has been reviewed in this section, each experimental group made significant gains from pretest to posttest in all production tasks, and all experimental groups outperformed the CO group for all tasks. This further supports the notion for empirically tested pronunciation interventions to be integrated into L2 teaching materials, along with providing further evidence that each of the three treatments are beneficial to L2 learners.

Additionally, in recognizing that the CI group outperformed both the EI and VF groups in multiple tasks of the production study, it is also practically suggested that this method may be a more beneficial method to teaching pronunciation in the L2 Spanish classroom. However, there is

more research to be conducted in order to uncover whether or not all methods presented here are beneficial in teaching other features in Spanish, as well as other languages, and whether one method should be considered for best practices in L2 instruction. Moreover, more comparative analyses should be conducted for different features of various languages to assess best practices, as the combination approach may not be the most beneficial for all languages and their specific features.

CHAPTER 5. PERCEPTION RESULTS & DISCUSSION

The current chapter focuses on the data coding, statistical analyses, and discussion of the data analyses for the two perceptual tasks: a discrimination task (AXB task) and a nativeness judgment task (Likert-scale task). While the discrimination task was employed to assess whether participants were able to distinguish between short-lag and long-lag variants in word-initial position, the nativeness judgment task assessed whether participants ascribed the correct social judgment to Spanish productions and English-accented productions. The purpose of these tasks was to determine (1) whether production training impacts L2 perceptual skills, and (2) the relationship between L2 perception and L2 production. Data from all four groups were analyzed for each of the two tasks at the pretest and posttest and will be discussed in detail in the following sections of this chapter. As described in Chapter 3, the four groups are: explicit instruction (EI), visual feedback (VF), combination instruction (CI), and the control group (CO).

In the final section of this chapter, 5.2, the analyses are discussed in detail with respect to interpretation of results, as well as the theoretical implications of this portion of the study. Results, once again, are assessed in relation to within-group improvement and between-group improvement.

5.1 Data Analysis & Results

For each of the following analyses, a mixed-effects model was run to evaluate the percent correct for the discrimination task (Section 5.1.1) and Likert-scale ratings for the nativeness judgment task (Section 5.1.2). The statistical program RStudio (Allaire, 2012) was used to carry out the analyses. The lme4 package (Bates et al., 2014) was used to conduct mixed-effect models for both experiments. Effect sizes (Cohen's *d*) were also calculated for all groups. Effect sizes are discussed with reference to L2 research benchmarks outlined in Plonsky and Oswald's (2014) meta-analysis: small = 0.40, medium = 0.70, and large = 1.00.

5.1.1 Discrimination Task

A total of 45 participants[16] took part in the discrimination task (AXB task): EI = 14; VF = 11; CI = 11; CO = 9. This task contained 48 tokens at the pretest and at the posttest, resulting in a total of 4,320 tokens (45 participants × 48 tokens × 2 sessions = 4,320 tokens). For this task, percent correct was calculated (Best, 1991; Best, 1995; Best et al., 2001) for each of the 45 participants at the pretest and posttest, resulting in 90 total calculations. With respect to percent correct calculations for each participant, outliers were defined as two standard deviations above or below the mean for each group at each point in time (pretest vs. posttest). For calculating outliers, the mean was again taken for each group at each time (pretest and posttest) for each task. Of these 90 evaluations and their percentages, two outliers (two participants' data) were identified and removed, eliminating 96 tokens (2 × 48 = 96; 2% of data) from the total. This then resulted in a total token count of 4,224 (4,320 − 96) and 88 evaluations included in the analysis.

For the mixed-effects model, *time* (pretest or posttest) as well as *group* (CO, EI, VF, CI) were included as fixed effects, along with *participant* (subject) set as a random effect with random intercepts[17]. The goal of running the mixed-effect model was to determine if (1) all groups were matched at the pretest, (2) if there existed any significant interactions at the posttest, and (3) if any group(s) made improvement from pretest to posttest. In assessing these, the significance criterion was set at $|t| > 2.00$.

Results of the mixed-effects model (Table 16 below), specifically comparisons between the intercept (CO group) and each of the experimental groups at the pretest, indicated that there were no significant differences for group performance at the pretest, except between the CO and VF groups. This suggests that the different groups were mostly well-matched at the pretest, with the VF group performing better than only the CO group. However, as can be seen in Table 16 when observing the interaction between time and group, something unexpected occurs; while none

[16] Although 47 participants were included in the analyses for the production tasks, several participants did not complete all required tasks for either the production portion or the perception portion, resulting in different groups sizes and total participants.

[17] Initially, a maximal random effects structure was included in the model, with random intercepts and slopes by time by group (Barr et al., 2013). Singularity or convergence issues indicated that the random effects structure needed to be simplified. This was done by stepwise simplification of the random effects structure, and the maximal random effects structure that permitted convergence was random intercepts only.

of the experimental groups outperform the CO group, nor does the CO group significantly improve, the VF group significantly declines in performance.

Table 16 *Results for the Discrimination Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	72.454	3.277	[65.900, 79.008]	22.109	-
EI	8.350	4.201	[-0.052, 16.752]	1.988	0.864
VF	10.121	4.419	[1.283, 18.959]	2.290	1.519
CI	1.598	4.419	[-7.240, 10.436]	0.362	0.163
Posttest: CO	7.311	4.331	[-1.351, 15.973]	1.688	0.897
Posttest: EI	-2.697	5.475	[-13.647, 8.253]	-0.493	1.464
Posttest: VF	-22.462	5.748	[-33.958, -10.966]	-3.908	0.349
Posttest: CI	-0.870	5.748	[-12.366, 10.626]	-0.151	1.007

Based on a reordered model, conducted with the same parameters as listed above, within-group differences between the pretest and posttest were examined (not pictured in Table 16). By the same token, none of the experimental groups significantly improved from pretest to posttest (EI: β = 4.614, *SE* = 3.350, *t* = 1.377; VF: β = -15.151, *SE* = 3.779, *t* = -4.009; CI: β = 6.440, *SE* = 3.780, *t* = 1.704). Figure 10 displays each group's performance at the pretest and posttest for the discrimination task and the respective group performance in relation to percent correct. A total of all averages (all groups included) at the pretest was 77.92%, with an overall average (all groups included) of 78.69% at the posttest.

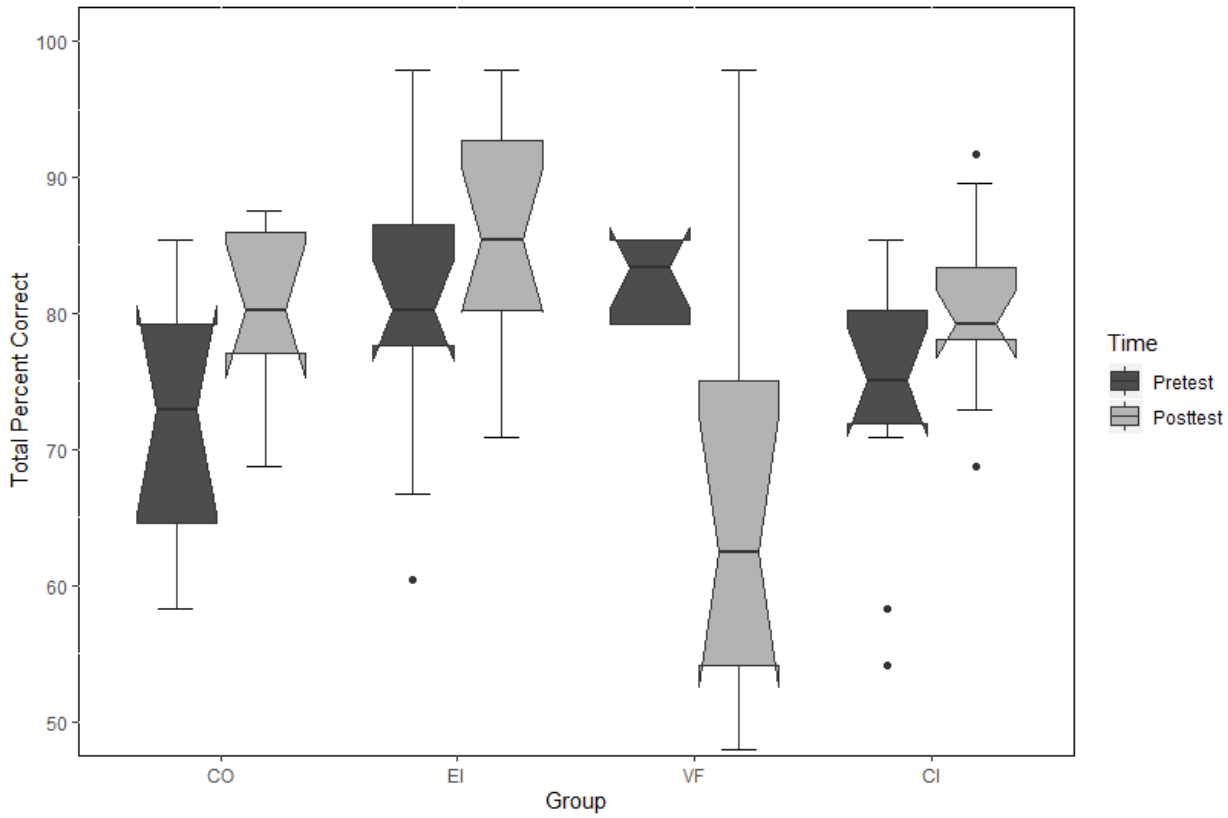


Figure 10 Total percent correct by group and time for the discrimination task.

5.1.2 **Nativeness Judgment Task**

Regarding the nativeness judgment task (Likert-scale task), a total of 24 tokens were rated by the 45 participants both at the pretest and the posttest (45 participants × 24 tokens × 2 sessions = 2,160). Since all responses were limited to a range of 1-7 for ratings, outliers were not calculated for this task due to a ceiling effect (Treiblmaier & Filzmoser, 2009). As such, all responses at the pretest and posttest for all participants were included. All contrast coded (see below for details) nativeness ratings were averaged for each group at the pretest and posttest.

For the mixed-effects model, *time* (pretest or posttest) as well as *group* (CI, EI, VF, CO) were included as fixed effects, while *participant* (subject) and *token* (item) were set as random effects with random intercepts.[18] For the Likert-scale task, a mixed-effects model was conducted

[18] Initially, a maximal random effects structure was included in the model, with random intercepts and slopes by time by group (Barr et al., 2013). Singularity or convergence issues indicated that the random effects structure needed to be simplified. This was done by stepwise simplification of the random effects structure, and the maximal random effects structure that permitted convergence was random intercepts only.

to determine (1) if groups were well-matched at the pretest, (2) if there were significant interactions at the posttest, and (3) if individual groups made significant improvement in assignation of tokens native-like (Spanish) as more non-native-like (English-accented) from pretest to posttest.

It is worth noting that the participants were instructed to indicate what they perceived to be a native-like production (Spanish) with a lower numerical rating (Native = 1), and what they perceived to be a non-native-like production (English-accented) with a higher numerical rating (Non-native = 7). In this analysis, the ratings that participants provided for the Spanish variants (short-lag variants) were then contrast coded (e.g., a rating of 1 was contrast coded as 7, a rating of 2 was contrast coded as 6, etc.) in order to collapse the data for a full analysis. With all the data collapsed, it was possible to perform an analysis that incorporated all data points for each group at the pretest and the posttest. Ratings of 7 were then associated with the ability to accurately assign native or non-native labels to short-lag and long-lag productions of /p,t,k/. Therefore, in the new categorization of the data points, a rating of 7 = accurate, a rating of 4 = neither accurate nor inaccurate, and a rating of 1 = inaccurate. In this instance, if a group's total average is higher at the posttest than at the pretest, this is then was interpreted as a trend towards improvement. In assessing these data, the significance criterion was set at $|t| \geq 2.00$. Table 17 details the results of the mixed-effects model.

Table 17 *Results for the Nativeness Judgment Task*

Parameters	*β*	*SE*	*95% CI*	*t*	*d*
Intercept (Pretest: CO)	4.519	0.250	[4.019, 5.019]	18.097	-
EI	-0.224	0.257	[-0.738, 0.290]	-0.870	0.055
VF	-0.470	0.261	[-0.992, 0.052]	-1.804	0.304
CI	-0.032	0.269	[-0.570, 0.506]	-0.120	0.016
Posttest: CO	-0.620	0.252	[-1.124, -0.116]	-2.464	0.322
Posttest: EI	0.887	0.318	[0.251, 1.523]	2.786	0.032
Posttest: VF	0.680	0.323	[0.034, 1.326]	2.107	0.289
Posttest: CI	1.280	0.333	[0.614, 1.946]	3.844	0.211

At the pretest, all groups were matched in their initial performance. When looking at the posttest, it is evident that all three experimental groups also significantly outperform the CO group as well. However, upon examining the performance of the CO group at the posttest, it can be observed that this group performed significantly worse from pretest to posttest. With this significant decline in performance by the CO group, it is possible that these results have influenced

the results of the experimental groups by inflating the results (i.e. making these groups [EI, VF, CI] appear to have performed significantly better at the posttest). As such, a second analysis was run to observe within-group performance from pretest to posttest; this analysis was the same analysis used with all four groups (a mixed-effects model incorporating the same fixed effects and random effects with random intercepts). With respect to the EI group, there was no significant improvement from pretest to posttest (β = 0.136, *SE* = 0.140, *t* = 0.970). This is also the case for the VF group (β = 0.051, *SE* = 0.145, *t* = 0.348). In contrast, the CI group did achieve significant improvement from pretest to posttest (β = 0.490, *SE* = 0.157, *t* = 3.120).

With respect to all groups, means appear to be close to the midpoint of the Likert-scale continuum at both the pretest and posttest, revealing little to no progress (with the midpoint being a value of four, and 4 = neither accurate nor inaccurate). Regarding effect size for the CI group, as they were the only group with significant results, there appears to be a small effect for the CI treatment.

Results can also be seen for all groups below in Figure 11, with each group's pretest and posttest responses (contrast coded).

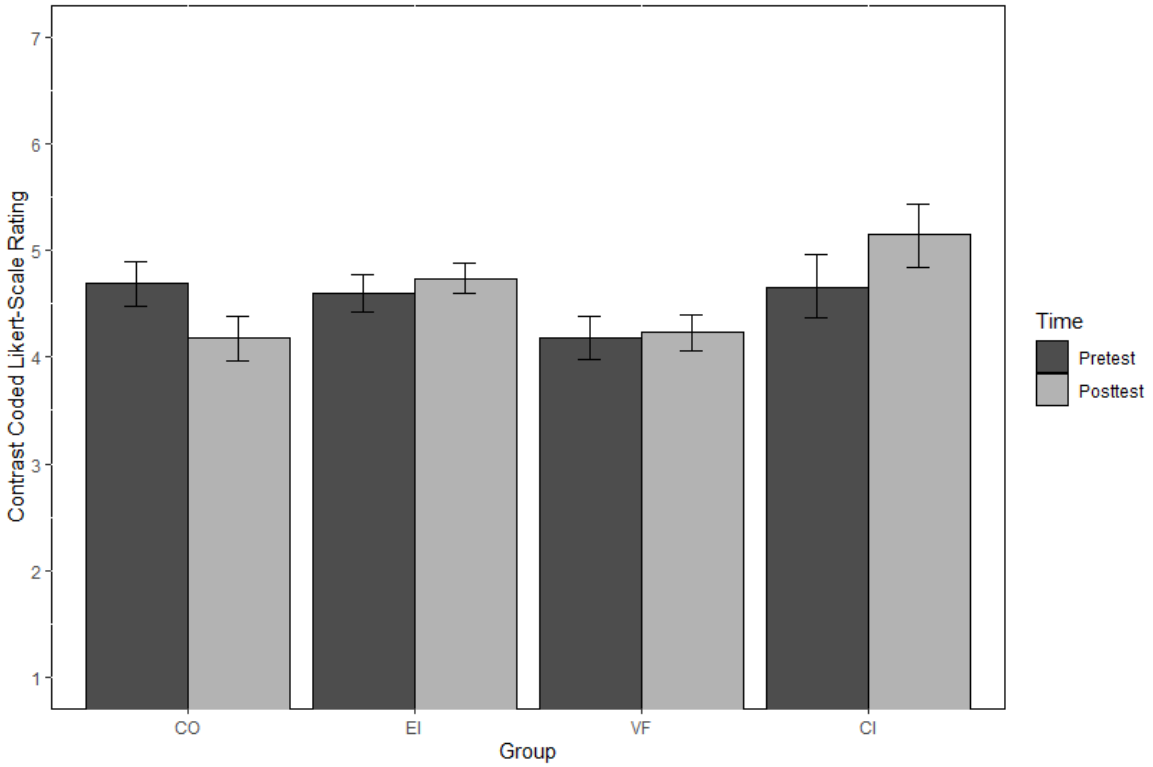


Figure 11 Contrast coded mean ratings by group and time for the nativeness judgment task. Error bars represent ±1 SE.

5.1.3 Conclusion of Analysis & Results

The current section, 5.1 Data Analysis & Results, covers the statistical analyses run via mixed-effect models to elaborate on (1) whether all groups were matched at the pretest, (2) the effect of time on all experimental groups and the CO group, and (3) if any group(s) made significant improvement in the assignation of tokens. As seen in section 5.1.1, there is no significant improvement in discrimination skills from pretest to posttest for any of the groups. As see in section 5.1.2, while there was no improvement for CO, EI, and VF groups for the nativeness judgment task, the CI significantly improved from pretest to posttest, once an additional analysis was conducted excluding the CO group. A more thorough assessment of these results is considered and examined in section 5.2. Again, discussion of the links between L2 production and L2 perception is presented in Chapter 6.

5.2 Discussion of Perception Results

The current section of this chapter re-examines the research questions and respective hypotheses, as well as consider the effectiveness of each type of production treatment with respect to how learners in each experimental group and the CO group performed in both perceptual tasks. The discussion of each task of the perception portion of this study is assessed in relation to how groups performed in terms of their ability to discriminate between long-lag (English) and short-lag (Spanish) variants of /p,t,k/ (Section 5.2.1), along with looking at how accurately groups were able to ascribe a variant of /p,t,k/ as being native-like (Spanish) or non-native-like (English) (Section 5.2.2).

5.2.1 Treatment Effectiveness on the Discrimination Task

Beginning with research questions 2 and 2.1 and their accompanying hypotheses, this subsection will discuss the results of the discrimination task.

(RQ2) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of perception for discrimination skills? In this case, improvement of phonetic perception is defined as the ability to discriminate between long-lag and short-lag variants of /p,t,k/ in word-initial position.

Hypothesis 2: Carried out via an AXB task, findings will reveal that L2 learners in all experimental groups will display trends of improving their perception of the difference in long-lag and short-lag voiceless stops in word-initial position after receiving pronunciation training. It is hypothesized that L2 learners in all experimental groups will improve based on the Noticing Hypothesis (Schmidt, 1990), where learners are predicted to improve due to being directed to notice differences between the L1 and the L2 with a type intervention.

(RQ 2.1) In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback training [VF], or a combination training [CI]) on L2 perception via discrimination between Spanish and English variants of /p,t,k/?

Hypothesis 2.1: Carried out via an AXB task, findings will reveal that L2 learners in the CI group will best improve their perception of the difference in long-lag and short-lag voiceless stops in word-initial position after receiving pronunciation training. The CI group will improve most due a combination of learner modalities, namely auditory and visual (Murphy, 1997), aiding learners in discrimination. This is thought to be because both explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as helping learners notice differences between the L1 and the L2. Therefore, it is possible that a combination of both methods will cause maximal noticing.

In addressing RQ2 and RQ2.1, the three experimental groups (EI, VF, CI) all performed equally to the CO group at the pretest, except for the VF group; the VF performed significantly better than the CO group at the pretest. However, as seen in the above analysis, the VF group performs significantly worse at the posttest. Regarding the CO, EI, and CI groups, none of these groups performed significantly better at the posttest. While there was no significant improvement at the posttest, it was evident that none of the groups had difficulties in discriminating between short-lag and long-lag variants of /p,t,k/ at the pretest. Average percent correct scores at the pretest for all groups ranged from 72%-83% (EI, M = 80.80% [SD = 9.81]; VF, M = 82.58% [SD = 5.26]; CI, M = 74.05% [SD = 10.09]; CO, M = 72.45% [SD = 9.42]), likely revealing that these learners already possessed an ability to discriminate between the two variants.

One plausible reason for why learners did not display difficulty in differentiating between the long-lag and short-lag variants at the pretest is due to the fact that voiceless stops in a post-vocalic, unstressed, word-medial position in English are often produced with short-lag VOTs or sometimes VOT values near 0ms (Antoniou et al., 2011; Cox & Palethorpe, 2007; Lisker, 1957), similar to the variants produced in Spanish in word-initial position.

Taking into consideration these patterns in English regarding VOT values for /p,t,k/ in the onset and medial positions, no improvement from pretest to posttest does not necessarily indicate that participants are not able to distinguish between sounds; again, it appears that they already possess the ability to discriminate between these variants due to features in their L1, and their performance at the posttest merely demonstrates that they performed as expected if these phonological categories had already been established, similar to a ceiling effect.

One finding that was unexpected at the posttest was the VF group's performance. While they collectively performed the best at the pretest in comparison to the other groups their performance significantly declined at the posttest, although this was not the case for any of the other groups. It is not readily apparent why this group's performance would decline following a type of intervention and is worth noting that each group had a different instructor, which could represent a potential confound. Although this could be the case, it is more likely that there was an error produced in the task execution. Nevertheless, more testing involving instructor-type would need to be conducted to test this as cause for decline in performance.

5.2.2 Treatment Effectiveness on the Nativeness Judgment Task

Again, beginning with the research questions 3 and 3.1 and their accompanying hypotheses, this subsection examines the results found for the nativeness judgment task:

(RQ3) Does pronunciation instruction, implemented here as either explicit instruction (EI), visual feedback (VF), or a combination of the two (CI), result in an improvement of perception for social categorization skills? In this case, improvement of phonetic perception is defined as the ability to perceive and categorize long-lag tokens as more non-native (English-accented) and short-lag variants as more native (Spanish).

Hypothesis 3: L2 learners in all experimental groups will improve in their ability to perceive long-lag variants of /p,t,k/ as more English-like (non-native) and short-lag variants

as more Spanish-like (native) after participating in production training. Based on the Noticing Hypothesis (Schmidt, 1990), learners are predicted to improve due to being directed to notice differences between the L1 and the L2 via all three production treatments.

(RQ 3.1): In a comparative analysis, is there a difference in the effectiveness of the three different pronunciation approaches (explicit pronunciation instruction [EI], visual feedback training [VF], or a combination training [CI]) on L2 perception via social categorization of Spanish and English variants of /p,t,k/?

Hypothesis 3.1: Findings will reveal that L2 learners in the CI group will best improve in their assignation of short-lag variants being native-like (Spanish) and long-lag variants being non-native (English-accented) after participating in production training. Based on the Noticing Hypothesis (Schmidt, 1990), learners in this group are predicted to improve due to being directed to notice differences between the L1 and the L2 via two different modalities (auditory and visual; Murphy, 1997), as explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as helping learners notice differences between the L1 and the L2. Therefore, it is possible that a combination of both methods will cause maximal noticing.

For the nativeness judgment task, all groups at the pretest performed equally in their ability to identify productions as more non-native-like (English-accented) or more native-like (Spanish). At the posttest, statistics reveal a significant decline in performance by the CO group, and an analysis of the interactions implies that all experimental groups outperform the CO group. Although performance by all experimental groups was originally interpreted as improvement when compared to the CO group in the initial analysis, concerns relating to the CO group's decline in performance and conceivably inflating the results for the EI, VF, and CI groups at the posttest was determined as valid reasoning for a subsequent analysis. A within-group analysis was done for the remaining experimental groups, and it was found that only the CI group significantly improved in their ability to assign short-lag variants as more native-like and long-lag variants as more non-native-like. Additionally, while the EI and VF groups did not improve significantly from pretest to posttest, it was seen that they did make slight trends towards improvement in their assignation of variants at the posttest.

Further, for the CI treatment, there was a small effect size, indicating that there was indeed an impact on perception as it relates to social categorizations of native-like and non-native-like productions. As stated in Chapter 4, although the effect size is small in magnitude for the CI group, this treatment had an impact on participants. Again, it could be that if participants in this group were exposed to more frequent treatments over time, it is possible that the effect size could have increased with longer treatment sessions and/ or more exposure over time (Plonsky & Oswald, 2014). All groups were exposed to 50-60 minutes of treatment over the course of six weeks, with a total of 15-20 minutes for both perceptual tasks performed twice (30-40 minutes total), for a final total of 80-100 minutes. If perhaps the CI participants took part in treatments that were longer in duration, more frequent, and over the course of a longer period, it is possible that larger effects could be seen when analyzing the CI group's effect size.

Further, a possible interpretation of the results for this analysis could be that while all groups do show trends for improvement, the CI treatments were more beneficial in aiding learners to perceive differences in VOT as either native-like or non-native-like. Notwithstanding, there is still more research needed to assess whether a combination instructional approach is best for L2 learners in developing in their L2 perception. In addition, it is worth assessing whether explicit instructional methods and visual feedback paradigms truly do not benefit L2 learners for this specific type of perceptual skill.

5.2.3 Theoretical Implications

This section addresses how differing L2 perception theories explain the results found in the discrimination and the nativeness judgment tasks. Finally, the Noticing Hypothesis will be revisited to explain group performance for the nativeness judgment task, since only one group (CI) improved.

L2 Perceptual Theories

The results for the perceptual portion of this study can be explained within either the Speech Learning Model (SLM) (Flege, 1987), through the two-category assimilation facet of the PAM (Best, 1994), or through the PAM-L2 (Best & Tyler, 2007). As explained in Chapter 2 and in the previous section, SLM predicts that learners will assimilate similar sounds in the L1 and L2

to the L1, especially if these sounds are less salient. Considering that [p,t,k] and [p^h,t^h,k^h] are allophones of /p,t,k/ (Hammond, 2001; Hualde, 2005), it was predicted that since these allophones are similar in production, only differing in VOT values (Lisker & Abramson, 1964), learners may experience some difficulty in perceiving this distinction in word-initial position. This prediction was, in part, driven by the fact that previous research has shown that L1 English-speaking L2 learners of Spanish tend to produce long-lag [p^h,t^h,k^h] in word-initial position in Spanish (Lord, 2005). However, the discrimination task showed that these L2 learners successfully perceive differences between the long-lag and short-lag productions in word-initial position.

Another possible explanation for learner performance regarding the discrimination task is the two-category assimilation notion of the PAM (Best, 1995). With two-category assimilation, it could be that participants were able to perceive differences in the two variants (long-lag vs. short-lag), relying in part on extant L1 categories. For example, successful discrimination of [p] and [p^h] may leverage existing English phonemes/allophones, as long-lag [p^h] occurs word-initially in English (Lisker & Abramson, 1964), and [b], occurring in multiple contexts in English (including word-initial position), is short-lag or at times has a negative VOT value (Scobbie, 2006). Although voiced, /b/ is considered a stop, with the only distinguishing characteristic being voicing for /p/ and /b/, as both are bilabial stops (Hammond, 2001). The ability to rely on both their /p/ and /b/ English phonological categories, according to PAM, may explain why they were able to perceive the short-lag [p] variant. However, had these allophones been a set of sounds that exist in English and another set of sounds that do not exist in English and do so in Spanish, it is possible that category assimilation could occur, thus being explained through the SLM theory. More research is be conducted to investigate this notion.

As discussed in the previous section, participants in all groups did not experience difficulty in perceiving allophonic differences for /p,t,k/, which is possibly due to L1 English speakers' ability to have already developed phonological categories separating the allophones of /p,t,k/ since short-lag variants exist in some word-medial contexts in English (Antoniou et al., 2011; Cox & Palethorpe, 2007; Lisker, 1957). Therefore, a better predictor of what occurred with L2 learners for the discrimination task could be explained by the PAM-L2 theory, which is a speech-driven theory (Best & Tyler, 2007). The PAM-L2 predicts that while L2 learners may have a common phonological category and be able to perceive realizations of /p/, for example, in the L1 and L2, these learners may still have different phonetic realizations of this shared phoneme because of

gestural distinctions (Best & Tyler, 2007). In looking at the performance of participants in the discrimination task and performance before intervention for the production portion in Chapter 4, the PAM-L2 best explains why L2 learners were able to perceive differences well at the onset of the study while still producing more non-native-like productions prior to intervention.

Contrary to the discrimination task findings, initially it was seen in the nativeness judgment task that L2 learners are not able to accurately assign short-lag and long-lag variants as native-like (Spanish) or non-native-like (English-accented) before treatment, and that they only slightly improve, with the CI group being the only group to improve significantly. Although there is a slight improvement by the EI and VF groups, these experimental groups were not able to surpass the threshold of an overall rating of 5 on the Likert-scale for either the pretest or the posttest (1 = inaccurate, 4 = neither accurate nor inaccurate, 7 = accurate; higher rating = more improvement). These mean values, only slightly above chance, suggest that they were unable to assign native-like judgments to the short and long-lag variants, and showed little overall improvement. Conversely, the CI group showed some improvement following training, indicating an improved ability to categorize sounds as native or non-native.

However, it should be noted that a task such as this and how it applies to perceptual skills of L2 learners is something that is lacking in the SLM (Flege, 1987), PAM (Best, 1994), and PAM-L2 (Best & Tyler, 2007) theories. In other words, theory should not only hypothesize whether or not L2 learners can discriminate and develop new phonemic categories, but it should also seek to inform research in terms of how L2 learners associate sounds (e.g., native-like or non-native-like) with the L1 and L2 after being exposed to the L2. For the SLM, PAM, and PAM-L2 theories, these theories primarily hypothesize about how sound categories are developed for L2 learners, along with how and to what extent the L1 affects categorical development; yet there is no mention in any theory about what L2 learners are able to do with these new categories that have been developed. Even for the more perceptually driven model of the PAM, there exists no discussion of how L2 learners categorize allophonic sounds in the L1 and the L2. This is crucial in that if learners are able to distinguish sounds in the L1 and the L2, but do not know how to assign these sounds to the appropriate language, it is predicted that they will also have difficulty producing the appropriate sounds in the L2.

Noticing Hypothesis

Mentioned earlier in Chapter 4, the notion of *noticing* (Izumi & Bigelow, 2000; Schmidt, 1990; Schmidt, 1992; Swain & Lapkin, 1995) from the Noticing Hypothesis (Schmidt, 1990) was proposed as the theoretical rationale for L2 learners' ability to improve in their L2 productions after intervention. This idea is again proposed as a plausible explanation for the CI group being the only group that improved significantly from pretest to posttest for the nativeness judgment task. Also previously discussed, Derwing and Munro (2005) suggest that various explicit pronunciation instruction methods are possibly able to direct learners' attention to pronunciation features in the L2. Tyler (2019), in discussing how to implement the principles of the PAM (Best, 1994) explicitly suggests the use of IPA symbols in place of graphemes to teach L2 learners how to perceive differences in the L1 and the L2, starting as early as possible. Tyler (2019) suggests that these types of intervention promote "awareness of phonological differences that are difficult to perceive, and it may provide a point of focus to help students to learn the phonetic differences between L2 phonemes" (p.625).

Further, Olson (2014b) posits that visual feedback can assist learners in visually noticing differences in the L1 and the L2 if learners' auditory perceptual capacity is not sufficient in aiding these learners to notice differences. In this specific case, no group is able to accurately categorize the long-lag and short-lag variants of /p,t,k/ in word-initial position as being associated with English or Spanish at the onset of the study. Because only learners in the CI group improved significantly, it is perhaps that the combination of an explicit approach coupled with a visual feedback component aids L2 learners of Spanish to better notice and perceive differences of long-lag and short-lag productions of /p,t,k/ as they relate to categorization of being native-like (Spanish) or non-native-like (English-accented). This is possibly due to the incorporation of two different types of L2 modalities, auditory (here, associated with explicit instruction) and visual (here, associated with visual feedback) (Baran-Łucarz, 2012). While this study suggests that a CI treatment is more beneficial in terms of *noticing* for these particular aspects of L2 Spanish, serving to improve both L2 production and perception, additional research should seek to confirm these findings. Specifically, a comparative analysis between EI, VF, and CI treatments for other segmental features of Spanish and English, as well as other features of multiple language pairings, would be of interest.

5.2.4 **Conclusion of Discussion**

As shown in this section, no group improved in their ability to discriminate between short-lag and long-lag productions of /p,t,k/, possibly due to the fact that both variants exist in participants' L1 (English), albeit with different distributions. With respect to these findings, it was it was thought that the PAM-L2 best explains the relationship between L2 learners' ability to discriminate between allophonic differences in L1 English and L2 Spanish, yet a lack of ability to produce these differences (prior to intervention).

Further, it has also been concluded that although two experimental groups show strides towards improvement for the nativeness judgment task, with only one group, the CI group, significantly improving from pretest to posttest. Moreover, it was proposed again that the Noticing Hypothesis is a likely explanation for why L2 learners in the CI group were able to more accurately assign sounds as being native-like (Spanish) or non-native-like (English-accented) post-intervention. This, again, suggests that the combination of explicit instruction and visual feedback causes L2 learners with different modality preferences to more successfully notice differences between the L1 and the L2 for social categorization of sounds.

Finally, it should be recognized that while all groups performed well above chance (i.e., near ceiling effect) in the discrimination task, performance for the nativeness judgment task was much less successful (i.e., near chance). These findings highlight the fact that perception relies both on discrimination and the ability to associate the different L2 variants with the appropriate language, here as native-like (Spanish) or non-native-like (English-accented). This key subcomponent of L2 perception is noticeably absent from existing models of L2 perception theories.

The following chapter, Chapter 6. Discussion of Production and Perception Results, discusses more in-depth the relationship between L2 production and L2 perception.

CHAPTER 6. DISCUSSION OF PRODUCTION & PERCEPTION STUDIES

This chapter reviews the research question regarding the relationship between L2 production and L2 perception, along with the accompanying hypothesis. To follow, a discussion of the theoretical implications of both studies regarding the relationship between L2 production and L2 perception is expanded on.

6.1 Review of Research Questions for Perception & Production Relationship

(RQ4) Does there exist a relationship between perception improvement and production improvement for all experimental groups? That is, does L2 perception improve after production training, indicating that perception does not always precede production?

> Hypothesis 4: Drawing on research stating that it is not clear what the relationship is between L2 production and perception (Flege et al., 1997), it is hypothesized that all groups will improve in producing more target-like productions of /p,t,k/ in Spanish post-treatment, as well as improve in their discrimination between and assignation of short-lag and long-lag variants. Again, the Noticing Hypothesis (Schmidt, 1990) is proposed and an explanation for improvement, as explicit instruction (Tyler, 2019) and visual feedback (Olson, 2014b) are both proposed as helping learners notice differences between the L1 and the L2. This would provide evidence that it is possible that L2 production and L2 perception may develop simultaneously for adult L2 learners after a form of production treatment.

6.2 Relationship between L2 Production & L2 Perception

Returning to the literature examined in Chapter 2, previous research has claimed that L2 perception always precedes L2 production for adult learners (Levy & Law, 2010); conversely, several studies made claims that the relationship between L2 production and L2 perception is unclear for adult learners, arguing that learners sometimes improved in production, but still revealed a deficit in perceptual skills (Flege et al., 1997; Kartushina et al., 2015; Zampini 1998). The following subsections examine the relationship between L2 production and L2 perception,

separated by L2 perception in the form discrimination capabilities and how it relates to L2 production (6.2.1), followed by social assignation capabilities and how this relates to L2 production (6.2.2).

6.2.1 L2 Discrimination and L2 Production

Detailed in Chapter 5, it appears that L1 English-speaking learners of L2 Spanish are able to successfully discriminate between long-lag and short-lag /p,t,k/[19]. Again, it was posited that this may result from the existing phonological inventory of English, which includes both variants of /p,t,k/, with the only difference being word context (initial vs. post-vocalic medial) (Antoniou et al., 2011; Cox & Palethorpe, 2007; Lisker, 1957). Thus, it is not unexpected that none of the groups improved in discriminating between allophonic productions of long-lag (English) and short-lag (Spanish) /p,t,k/ in the discrimination task, as it appeared they were already capable of discriminating between the majority of these productions at the pretest.

With respect to the relationship between L2 production and L2 perception, the results from the discrimination task provide evidence that, in this case, perception precedes production for this particular group of sounds, the voiceless stops. Specifically, at the pretest, while participants are highly successful at discrimination between the short-lag and long-lag stops, they fail to produce the native-like, short-lag variant in word-initial position. The 'perception-first' notion has either been explicitly stated in previous research (Levy & Law, 2010) or implied by other theoretical frameworks that describe the development of L2 perceptual skills and how these skills then affect L2 production, such as speech-driven models like the SLM (Flege, 1987) and the PAM-L2 (Best and Tyler, 2007). Both theories imply promotion of the perception-first notion by initially focusing on how the L1 will affect category formation of the L2 or assimilation to the L1, and that these aspects influence how L2 speech is produced.

Further, and as mentioned earlier, with L2 learners in this study already being able to distinguish VOT differences for /p,t,k/ because of their existence in different word-contexts in English (Antoniou et al., 2011), this explains why L2 learners were able to first perceive the differences in word-initial position, but were not successful in producing Spanish-like realizations

[19] It could be that L2 learners possibly responded to differences in sound due to phonemic category formations of /p,t,k/ vs. /b,d,g/ in English; however, further testing would need to be conducted to examine if this was the case as well as investigating how the strength of real-word bias plays a role in L2 discrimination.

prior to intervention (as long-lag variants in word-initial position exist in English, not in Spanish) (Hammond, 2001). Not only were L2 learners already capable of discriminating between the variants at the pretest, but it was only after a form of intervention that they were able to produce more native-like (Spanish) realizations of the stops. Additionally, these results suggest that the perception-first notion may apply in this particular case since the experimental groups' discrimination abilities did not improve after an intervention, nor do these learners display need for improvement at the pretest. This implies that production treatment did not affect this perceptual skill, and learners were capable of discrimination prior to enhancing their production skills. Further research may serve to confirm this interpretation across a range of L2 sounds.

However, it should be noted that more studies should be run to further test the notion that perception comes before production in adult L2 learning, as this study has only tested one group of phonemes and their allophonic realizations occurring in word-initial position for English and Spanish.

6.2.2 Nativeness Judgment of Sounds and L2 Production

Contrary to the discrimination task findings, initially it is seen in the nativeness judgment task that L2 learners are not able to accurately assign short-lag and long-lag variants as Spanish- or English-like before treatment. After treatment was carried out for the experimental groups, the CI group was the only group to improve significantly, and the effect remained small. In addition, while EI and VF groups showed non-significant improvement, all groups produced Likert-scale judgements in the middle range of the seven-point scale at the pretest (range = 4.1-4.7). As opposed to the findings for the discrimination task, it does not appear that participants initially possess the ability to accurately assign short-lag productions as native-like and long-lag productions as non-native-like, indicating that there appears to be a deficit in this perceptual skill for L2 learners.

With the CI group having significantly improved on this perceptual task (i.e., nativeness judgment task) following production training, there is evidence that production training does influence perception skills. This improvement in perceptual performance provides some preliminary evidence for L2 production to precede L2 perception, or possibly occur simultaneously, for this aspect of social categorization of sounds for the L1 and the L2. For the EI an VF groups, although they did improve in production, here they did not significantly improve for the nativeness judgment task; this could possibly serve as evidence that L2 production does

precede L2 perception for socially assigning sounds to the L1 and the L2. Additionally, perhaps with more intervention or a more longitudinal study, it could be shown that both the EI and VF groups would improve significantly for the nativeness judgment task.

However, as discussed in Chapter 5, L2 perception theories such as the SLM (Flege, 1987), the PAM (Best, 1994), and the PAM-L2 (Best & Tyler, 2007) do not discuss how L2 learners socially categorize sounds in the L2 after acquisition of new sounds or assimilation of similar sounds in the L1 and L2. Therefore, it is difficult to utilize a perception-specific theory to expand on whether the perception-first notion is supported. For this reason, the Noticing Hypothesis (Long, 1990) has been proposed in Chapter 4 and Chapter 5 for explaining learner improvement of both L2 production and L2 perception skills. Although L2 learners in this study are able to discriminate between long-lag and short-lag variants in word-initial position, it is not readily apparent that these learners are able to perceptually categorize these sounds as either native-like (Spanish) or non-native-like (English-accented). Moreover, they either do not improve in this perceptual skill (EI and VF groups), or those in the CI group do not improve until after taking part in the CI treatment, which is considered a dual modality (auditory and visual) treatment. For the nativeness judgment task, the evidence is two-fold: (1) results from this study imply that L2 perception as it relates to social categorization may not precede L2 production, as L2 production and this facet of L2 perception possibly occurred simultaneously post-intervention, although a more in-depth study and analysis would need to be conducted to closely examine which, if either, precedes the other; (2) results indicate that native English-speaking L2 learners of Spanish are able to improve this perceptual skill after a combination instruction approach, involving both explicit instructional methods and a visual feedback paradigm.

6.3 Conclusion

Taking both perception tasks and results into consideration, there is not sufficient evidence to support the notion that L1 English-speaking L2 learners of Spanish improve in their discrimination of /p,t,k/ variants after a form of production training. With largely successful discrimination between long-lag and short-lag /p,t,k/ found at the pretest, results in this study support the notion that perception, defined as discrimination, precedes production for adult, English-speaking, L2 learners of Spanish. Nevertheless, there is more research to be done in this

area for different phonemic structures that exist for different adult L2 learners with varying L1 backgrounds.

Although perception appears to come before production with respect to the discrimination task, this does not necessarily mean that adult L2 learners are able to achieve more native-like productions without some sort of pronunciation intervention. Therefore, it appears that even with an ability to discriminate between sounds, this does not correspond to nor imply an ability to produce more target-like pronunciation without intervention. Additionally, it also appears that perception does not necessarily come first in terms of a social categorization task (tested here through the nativeness judgment task), as participants were largely unable to successfully ascribe the appropriate social categorizations at the pretest. It was only after a specific type of intervention, the CI treatment, that learners were able to both improve their productions and their social categorizations of /p,t,k/. Therefore, the perception-first idea is only partially supported in this study, although more research is to be done to more thoroughly examine the relationship between L2 production and L2 perception, and more notably for the separate components of L2 perception.

CHAPTER 7. CONCLUSION

This chapter will summarize the current study and restate various conclusions from both the production and perception portions (Section 7.1). Additionally, this chapter details the pedagogical implications of the study (Section 7.2), along with discussing the limitations of the study (Section 7.3) and future directions (Section 7.4).

7.1 Summary of Study

Broadly, this study examined the effect of three different types of pronunciation instruction on L2 production and L2 perception of word-initial voiceless stops. Four groups of participants, all adult, English-speaking L2 learners of Spanish, received either explicit instruction (EI), visual feedback (VF), a combination of these two instructional methods (CI), or a supplementary task (not involving pronunciation instruction) in the form of a cultural activity of similar duration (control group [CO]). L2 production, namely voice onset time (VOT), was evaluated both prior to and following the three interventions in four tasks, ranging from most controlled (carrier phrase task) to most spontaneous (spontaneous speech task). L2 perception was evaluated through both a discrimination task (AXB task) and a nativeness judgment task (Likert-scale rating task).

Production results demonstrated that participants in all three experimental groups (EI, VF, CI) significantly improved their productions of /p,t,k/ (i.e. reduction of VOT values for the voiceless stops) for all four tasks in the production portion of this study relative to the matched CO group. Conversely, none of the experimental groups nor the CO group improved in their ability to discriminate between long-lag (English) and short-lag (Spanish) productions of /p,t,k/, and it was posited that these learners may already be skilled at discriminating between long-lag and short-lag variants, as both exist in English in differing word positions (i.e. word-initial and word-medial). It was also found that the only group that improved for the nativeness judgment task after production training was the CI group.

In a comparative analysis of the effectiveness of the different types of instruction, the CI group outperformed the CO group in all production tasks and outperformed the EI and VF groups for the majority of production tasks. Similarly, the CI group was the only group to improve significantly in their assignation of allophonic productions of /p,t,k/ for the nativeness judgment

task. The CI group's performance was explained within the theoretical framework of *noticing*. It is possible that this type of instruction appeals to different learner modalities, such as auditory and visual learners, further enhancing their noticing of differences between English and Spanish with respect to the segments /p,t,k/ in word-initial position.

7.2 Pedagogical Implications

In drawing on the previous theoretical implications of this study, several pedagogical implications are discussed for both L2 production and L2 perception. Both are discussed in the following subsections, along with the preferred method based on the results presented in Chapter 4 and Chapter 5. Finally, this section will conclude with a discussion regarding lower-level learners' ability to improve their L2 pronunciation following a type of production training.

7.2.1 Pedagogical Implications for L2 Production

For the production portion of this study, again, for all four production tasks, all three experimental groups improved significantly in productions of word-initial /p,t,k/ by reducing VOT values at the posttest (as reduction in VOT values is associated with improvement). Additionally, all three experimental groups not only improved from pretest to posttest, but each group also outperformed the CO group for all four tasks. This reinforces previous research that has found that intervention aids L2 learners in improving their L2 pronunciation (Derwing & Munro, 2005), with the specific types of instructional approaches including explicit instruction (e.g., Lord, 2010; Miller, 2012), visual feedback (e.g., deBot, 1983; Offerman & Olson, 2016; Olson, 2014b), and combination instruction (e.g., Kartushina et al., 2015). Specifically for /p,t,k/ in Spanish and English, all three treatments were shown to benefit learners in their pronunciation learning. Although the results from this study provide support for the effectiveness of all three methods, consideration of these three approaches in relation to other features and languages is crucial. As this study only provides evidence for a specific group of segments (voiceless stops) in Spanish vs. English for a particular context (word-initial), it would be premature to claim that all methods are beneficial for the learning of all phonetic features in varying contexts for multiple languages.

With all instructional approaches in this study regarded as being beneficial, it was of particular interest to investigate which, if any, instructional type outperformed other instructional

types. With the rise in L2 pronunciation research, there was a specific call in the literature for a comparative analysis of instructional-types to be conducted to ascertain which method(s) should be considered for best practices for L2 learners (Derwing & Munro, 2015; Lee et al., 2015). This study begins to answer this call, detailing a comparative analysis between the EI, VF, and CI treatments. As mentioned in Chapter 4, the CI treatment outperformed the other two types of interventions, resulting in the most successful treatment in the production portion of this study.

As the CI method combined two previously tested and successful methods, namely explicit instruction (e.g., Elliott, 1997; Kissling, 2013; Lord, 2010) and visual feedback (e.g., deBot, 1983; Offerman & Olson, 2016; Olson, 2014b), it was hypothesized that this combination approach would allow learners to improve more significantly than learners in the EI and VF groups. In seeing that this hypothesis is supported in this study, it can be suggested that a combination approach is a more beneficial approach to teaching pronunciation to L2 learners of Spanish when evaluating performance for segments such as /p,t,k/. Additionally, as mentioned in Chapter 4 and Chapter 5, this type of training appeals to an array of learner modalities, such as auditory and visual learner modalities (Baran-Łucarz, 2012; Murphy, 1997), providing a multi-faceted type of training in which L2 learners have access to a variety of materials and examples to learn and notice (Schmidt, 1990; Schmidt, 1992; Swain & Lapkin, 1995) pronunciation features in their L2.

Although this approach is supported as the most advantageous treatment-type in this study, it is important, again, to consider that one treatment-type is not always the most beneficial for every phonetic feature of a specific L2 nor for all L2 learners with a wide variety of L1 backgrounds. For example, while visual feedback has been shown to be beneficial to L2 learners in a number of suprasegmental studies (e.g., Chun, 1989; deBot, 1983; Hardison, 2004), as well as segmental studies for (e.g., Offerman & Olson, 2016; Olson, 2014b), there exist mixed results for improvement of vowels (e.g., Motohashi-Saigo & Hardison, 2009; Ruellot, 2011). Olson and Offerman (under review) suggest that visual feedback works well for durational contrasts, as these are more visually intuitive. For vowels, it could be that focusing on formants is not as visually intuitive (Ruellot, 2011). It is possible for this to be the same case for a combination approach; while this method outperformed other methods for helping L2 learners reduce VOT values, this does not necessarily imply that L1 English-speaking L2 learners of Spanish will improve productions of another type of segment in Spanish (e.g., the Spanish vowels /ɑ,e,i,o,u/) nor segmental features in another language. For example, research has shown that explicit instruction

may serve to improve vowels for L2 Spanish (e.g., Elliott, 1997), although a comparative analysis similar to the current study would be a welcome addition to the field. While the combination approach is the recommended approach with respect to the results from this study, there is a need for more comparative analyses to be conducted for a variety of Spanish phonetic features, as well as features in other languages.

7.2.2 Pedagogical Implications for L2 Perception

Beginning with L2 perception, it appears that in this study, as elaborated on in Chapter 5 and Chapter 6, that L2 learners do not demonstrate difficulty in discriminating between the allophonic differences for /p,t,k/ in word-initial position in Spanish vs. English; however, this cannot be necessarily be applied to all differences in Spanish vs. English. For example, L1 English-speaking learners of Spanish often produce [b,d,ɡ], characteristic of L1 English, in intervocalic positions in their L2 Spanish productions instead of the fricative allophones (approximants) [β,ð,ɣ], characteristic of NSs of Spanish (Hammond, 2001; Lord, 2005). Although [ð] does exist in different contexts in English, [β] and [ɣ] do "not exist in the English sound inventory" (p.204) (Hammond, 2001). Therefore, it would be beneficial to further investigate allophonic sounds in Spanish to examine whether L1 English speakers can discriminate between this set of allophones before and after different types of intervention (EI, VF, CI). Studies of this type will better inform L2 theory, specifically for Spanish, as well as how L2 perception interacts with L2 production. Moreover, further studies may reveal that more production training helps improve or enhance L2 learner discrimination skills. As Tyler (2019) suggested, the use of the IPA to teach phonemic differences could be beneficial to L2 learners in their phonological development of the L2. However, again, more research is needed to be done in this area to provide concrete evidence regarding to what extent (if any) pronunciation instruction in the L2 classroom impacts L2 perceptual development.

With respect to the nativeness judgment task, it appears that L2 learners lack the ability to associate long-lag productions of /p,t,k/ with English and short-lag with Spanish in word-initial position, even after training. Although the CI group did improve from pretest to posttest for judging tokens as being either native (Spanish) or non-native (English) productions, this does not necessarily indicate that this type of training will aid L2 learners with other features of Spanish in categorizing them as either native-like (Spanish) or non-native-like (English). Again using this

example of the allophones [β,ð,ɣ] that occur intervocalically in Spanish vs. their respective allophones [b,d,g] in English (Hammond, 2001), it would be of interest to perform a similar judgment task utilizing all three experimental methods as separate treatments for different groups to see if (1) learners are able to assign each set of allophones to their respective language at the pretest and if (2) learners in the CI group benefit most from this type of treatment. While the current study's results provide evidence that the CI treatment is the most beneficial in helping L2 learners notice (Schmidt, 1990; Schmidt, 1992; Swain & Lapkin, 1995) differing sounds in this type of social categorization, more testing needs to be done to support the CI treatment as most beneficial.

In sum, the three production-oriented instructional methods utilized in this study do not seem to improve the perceptual capacity of discrimination for the voiceless stops. Again, perception here is taken as a two-part concept; while they are able to discriminate, participants are not able to accurately assign allophonic productions as native-like or non-native-like. At the posttest, however, L2 learners are able to improve in their ability to accurately assign allophonic productions with a combination instructional approach, indicating that improvement of this perceptual skill is attainable. With that said, further research should be conducted to discover which production training-types, if any, aid in the development or enhancement of L2 perception skills, as this study focused on only one subset of phonemes for two languages.

7.2.3 Pedagogical Implications for Lower-Level L2 Learners

A final pedagogical implication considers the impact of pronunciation instruction on L2 learners of a lower level. In much previous research, the recipients of pronunciation instruction are often L2 learners that are involved in a higher level L2 course (e.g., Aliaga-García & Mora, 2009; Lord, 2010) or in a L2 course that specifically concentrates on L2 phonetics (e.g., Kissling, 2013; Lord, 2005). Although these appear to be appropriate courses in which L2 pronunciation should be taught and tested, other researchers more recently have looked at incorporating L2 pronunciation instruction into lower-level courses (e.g., Miller, 2012; Offerman & Olson, 2016). The current study incorporated three different types of instruction into a lower-level course, with all treatments aiding learners in the significant improvement of their productions of /p,t,k/. Considering that all experimental groups improved, this further adds to the discussion that lower-level L2 learners are capable of learning pronunciation earlier than in a more advanced or

phonetics-specific course. Moreover, lower-level learners not only should be given the opportunity to learn pronunciation at an earlier stage in their L2 process, but it also does not appear that learning pronunciation at an earlier stage is something that is a hindrance nor detrimental to language learning as an adult. Further, pronunciation instruction should also be considered an essential component to L2 teaching and practices, especially at the lower level when a learner is first encountering their L2.

7.3 Limitations

While this study examines varying L2 pronunciation instructional types, informs pedagogical practices, and provides further support for integrating pronunciation instruction into the L2 classroom, there are several limitations that should be noted. This section focuses on two such limitations: (1) each group had a different instructor which could be a confound for performance analysis, and (2) there was no delayed posttest performed.[20]

With respect to instructor and group, all groups were assigned different instructors. This was a consequence of using intact classrooms, as there was not one instructor teaching four separate, intermediate-level Spanish courses. While use of a single instructor would have been ideal to eliminate any confound, this could not be avoided due to constraints of time and instructor availability. The broad goal was for every instructor to carry out their specific treatment-type equally with respect to presentation of the specific treatment to participants, as several trainings were conducted (see Section 3.3.7) to attempt to mitigate this potential confound. However, it is acknowledged that individual differences may exist for each instructor. These individual differences could impact results in various ways, including style of delivery, following the allotted time as directed (15 minutes for each intervention, allowing for 20 minutes for the first intervention), as well as student motivation in relation to their instructor. Although the researcher provided training to all instructors and emphasized the manner in which each treatment was presented as well as how much time should be dedicated to each training, the researcher was not present at the time of any intervention. Further, although the different instructors for each group pose a potential confound, this replicates the reality of most natural, intact, L2 classrooms.

[20] Although one instructor was not a NS of Spanish (VF group), which has been previously posited as having less of an effect on L2 learner performance (Flege &Liu, 2001), NNS instructors have been found to benefit learners equally to NS instructors in pronunciation improvement (Levis et al., 2016). Therefore, this was not considered to be a limitation of the study.

The final limitation is the lack of a delayed posttest for the production study to investigate whether participants maintained their improved ability, as suggested by Saito and Plonsky (2019). A delayed posttest could confirm whether participants maintained or improved their newly learned pronunciation skills. Incorporating a delayed posttest could also possibly result in a larger effect size for training-type (Lee et al., 2015). While a delayed posttest was initially included as part of the methodology (conducted following the semester), only four participants total submitted the delayed posttest, which consisted of the picture task only. Limited participation was not particularly surprising, as no compensation was offered. Again, since there was no official incentive given for completing all tasks to not influence participants, it is possible that many simply were not sufficiently motivated to complete the delayed posttest. In order to respect the participant and researcher boundary, the researcher only contacted the instructors of each group and requested that the instructors remind participants to submit their delayed posttests. Since there was not a sufficient sample size for any of the groups, there were no analyses run for the four participants.

Although there are some clear limitations to this study, they are not anticipated to have significantly impacted the outcomes of learner performance in this study. Additionally, it is worth noting that previous research has acknowledged that a comparative study of this magnitude is considered particularly difficult to execute (Derwing & Munro, 2015); therefore, limitations were anticipated, but again, not to the extent to which they would dilute or greatly affect the results.

7.4 Future Directions

This section suggests various future directions that incorporate the topics discussed in the previous sections. Specifically, this section discusses and proposes how the methods in this study can be applied to future studies, as well as how these future studies can further develop L2 perception theory and inform L2 pronunciation instruction. These include testing if EI, VF, and CI treatments impact L2 learner perception and production of another phonemic pairing in Spanish, as well as the Spanish vowels for L1 speakers of English.

The current study examined the production of VOT, leveraging differences in production of /p,t,k/ in Spanish and English in word-initial position (Lisker & Abramson, 1964). While it was necessary to limit the scope of this initial comparative analysis, a truly comprehensive comparative research agenda should seek to expand the comparison of treatments beyond VOT.

An examination of different segments is crucial for shaping L2 pronunciation curriculum, as different segments have shown divergent results in following different types of instruction. For example, while visual feedback has been shown to be successful for a number of consonants (Motohashi-Saigo & Hardison, 2009; Offerman & Olson, 2016; Olson, 2014b), it was not found to result in significant improvement for vowels (Ruellot, 2011). In contrast, vowels have been shown to improve following explicit instruction (Elliott, 1997). As Spanish and English vowel inventories differ (e.g., Hammond, 2001), future research may leverage these differences to determine if these same methods (EI, VF, CI) result in improvement of vowel production. In addition, a parallel comparative analysis may investigate which method, if any, outperforms other methods. Such a study would inform research for best practices in relation to pronunciation instruction, more specifically for Spanish.

Another area for future research would seek to further examine discrimination and nativeness judgment tasks in relation to all three types of pronunciation instruction (EI, VF, CI) for different segments. In considering the links between production and perception, the current study again examined the case of the voiceless stops. While stops were useful for demonstrating learner gains for the production portion of the study, they may not have been ideal for examining the development of L2 perceptual discrimination, given their relation to the L1 phonological inventory. Examining another phonemic pairing could allow for more successful and significant results. As proposed in the previous section, choosing the allophones of phonemes in both English and Spanish would be ideal, such as the allophones [β,ð,ɣ] in Spanish and [b,d,g] in English of the phonemes /b,d,g/, as both sets of allophones occur in intervocalic position for their respective language (Hammond, 2001)[21]. As Olson (2014b) conducted a study involving these phonemes with significant improvement post-intervention, these phonemes may also be a good test case for L2 perceptual development.

Moreover, these specific phonemes are thought to be ideal foci for a study involving L1 English and L2 Spanish. As stated previously, the allophones [β,ɣ] do not exist in English (Hammond, 2001), so a discrimination task involving these allophones would better inform

[21] While this may be a good test case, it should be noted that sometimes NSs of American English produce more approximant-like variations of /b,d,g/ due to free-variation and individual differences in connected speech (Bouavichith, & Davidson, 2013). However, since various research has not found this to be the case for L1 American English-speaking L2 learners of Spanish when producing intervocalic /b,d,g/ (e.g., Hammond, 2001; Olson, 2014b), this topic is still of interest for investigating.

perceptual theory, along with testing the 'perception-first' notion (Levy & Law, 2010). Additionally, there has not been a comparative analysis of pronunciation methods for /b,d,g/, as these allophonic differences have only been tested in an explicit instruction setting (Lord, 2010) or in a visual feedback paradigm (Olson, 2014b). With respect to a nativeness judgment task, it would be equally valuable to use these same allophones of /b,d,g/ ([β,ð,ɣ] and [b,d,g]) to assess whether learners at the pretest are able to assign the fricative variants (or approximants) as Spanish and the obstruents as English in an intervocalic context, and whether or not any type of intervention enhances their ability to accurately assign these variants to their respective language appearing in an intervocalic position.

Finally, another possible direction involves a closer examination of different learner modalities. As previous research has suggested, adult L2 learners may be more auditorily or visually inclined, categorizing them as either having an auditory learning modality or a visual learning modality (Baran-Łucarz, 2012; Murphy, 1997). Again, it was posited that the EI treatment is more auditory-driven and the VF treatment is more visually-driven, with the CI treatment being audio-visual in nature. In the current study, the CI group outperformed the other groups in both the production task and the nativeness judgment task. As any intact classroom is likely composed of both auditory and visual learners, the different treatments may be more relevant to a given group of learners. While the EI and VF treatments may cater principally to one type of learner modality, the CI treatment may benefit a wider range of learners. Subsequent analysis, accounting for learner modality, would reveal whether the CI treatment truly is the most beneficial treatment for all learners or if treatment success is dependent upon individual learner modality-types.

APPENDIX A. TOKENS IN THE CARRIER PHRASE

1. Di *Paco* de nuevo
2. Di *poco* de nuevo
3. Di *pesa* de nuevo
4. Di *puse* de nuevo
5. Di *pata* de nuevo
6. Di *testigo* de nuevo
7. Di *toca* de nuevo
8. Di *población* de nuevo
9. Di *pulga* de nuevo
10. Di *capa* de nuevo
11. Di *tilde* de nuevo
12. Di *casa* de nuevo
13. Di *quepo* de nuevo
14. Di *tacaño* de nuevo
15. Di *copa* de nuevo
16. Di *quita* de nuevo
17. Di *tuviera* de nuevo
18. Di *pecado* de nuevo
19. Di *culo* de nuevo
20. Di *taza* de nuevo
21. Di *quema* de nuevo
22. Di *tela* de nuevo
23. Di *pica* de nuevo
24. Di *quise* de nuevo
25. Di *cosa* de nuevo
26. Di *toser* de nuevo
27. Di *piso* de nuevo
28. Di *tiza* de nuevo
29. Di *tusar* de nuevo
30. Di *cupo* de nuevo

APPENDIX B. TOKENS IN THE NOVEL SENTENCES TASK

1. La *taza* está rota.
2. Quiero un *poco* de agua.
3. Hay un *testigo* con el juez.
4. Yo *puse* mi ropa en la lavadora.
5. Ese material me *pica* la piel.
6. No te olvides la *tilde* cuando escribes.
7. Esa *cosa* no sirve para nada.
8. Yo esperaba que él *tuviera* el anillo para la boda.
9. Siempre se *quema* con el fuego.
10. No digas *culo* frente a tu madre.
11. No sé por qué *Paco* quiere irse.
12. *Toca* la guitarra para mí.
13. Ahora ella *pesa* mucho.
14. Llévame a *casa* por favor.
15. Yo *quise* ir con ustedes, pero no pude.
16. Mira la *pata* nadando.
17. Él vive en el segundo *piso*.
18. Ella se *quita* los zapatos.
19. Matar es el *pecado* peor.
20. Chicago tiene una *población* gigante.
21. Una *pulga* mordió el perro.
22. Escribe con la *tiza* en la pizarra.
23. Él es muy *tacaño* con respeto a su ropa.
24. Quiero una *copa* de vino tinto.
25. No creo que la ropa *quepa* en la maleta.
26. Supermán tiene una *capa* roja.
27. No es educado *toser* con la boca abierta.
28. Hay que *tusar* las ovejas.
29. Esta *tela* es preciosa.
30. El aula tiene su *cupo* lleno.

APPENDIX C. TOKENS IN THE CONTROLLED, CONTINUOUS SPEECH TASK

Me llamo (1) Paco y quiero contarte sobre mi primera experiencia con mi compañero, Pedro. Había acabado de (2) cumplir 18 años, y tuve que mudarme a Indiana para mi primer año de la Universidad. Llegué a la (3) casa de Pedro (4) con mi padre el 12 de octubre. Era un día maravilloso afuera; hacía mucho sol, las hojas se cambiaban de color, y el aire olía de manzanas. (5) Todo era perfecto, y en ese momento, no (6) quise ir adentro de la casa. Pero por fin, (7) tocamos la puerta. Estaba un (8) poco nervioso, porque nunca había vivido con alguien a pesar excepto mi padre. Cuando Pedro abrió la puerta, yo lo examiné cuidadosamente; era un (9) tipo muy alto y delgado, y llevaba una (10) pulsera de oro. Pero se veía como una persona normal con sus jeans, su camiseta gris y su bigote tan (11) tupido. "Hola, ¿qué (12) tal ustedes? Bienvenidos a mi casa," dijo él. "Tu cuarto está en el (13) quinto (14) piso – les muestro el cuarto y el resto de la casa." Era una casa enorme con un (15) techo altísimo. No podía creer lo que mis ojos veían. Mi padre de repente empezó a hablarle, "Gracias a Usted por mostrarnos la casa, es hermosa." "De nada, de nada," dijo Pedro, "y ustedes me pueden (16) tutear, ya somos familia. Yo (17) puse una (18) cosa en tu cuarto para que te sientas en casa, Paco." Cuando llegamos a mi cuarto, había una (19) taza para tomar café en mi cama. "Oí que tomas (20) tinto, o café como decimos en Colombia, entonces quería regalarte esta taza de Colombia." "¡Muchísimas gracias!" le dije. "Me siento como cometí un (21) pecado porque no te traje nada, y esta taza es buenísima." "No te preocupes, como te dije, ya somos familia. Espero que me (22) pidas para (23) cada cosa que necesitas." Pedro bajó a la cocina para empezar a cocinar la cena, y mi padre fue al baño. Empecé a descargar mis cosas de las maletas. Había puesto la taza en mi escritorio, pero no me acordé de ponerla allí. Saqué mi (24) teclado de una maleta, y sin ver, intentaba a ponerlo en mi escritorio cuando me di cuenta de que pegué la taza sin querer y no tenía el (25) poder de salvarla. Mi padre oyó el ruido e inmediatamente salió del baño. "Qué (26) paso?!" me gritó. "Bueno…" le decía con mi cara hacia el piso, "acabo de…acabo de (27) quebrar la taza. Es completamente mi (28) culpa." De repente Pedro subió a mi cuarto y nos preguntó, "¿Ustedes están bien?! Oí un ruido abajo." Le explicaba lo que paso con la taza cuando olimos algo raro. Mi padre le dijo a Pedro, "Creo que algo se (29) quema abajo…" Corrimos abajo y supuestamente, la cena se quemaba en el horno. ¡Qué (30) pena! Pedro nos

aseguró que estas cosas pasan y nos dijo que no nos preocupáramos. Desde ese momento, yo sabía que seríamos buenos amigos.[22]

[22] This story was revised and edited by a native speaker of Puerto Rican Spanish.

APPENDIX D. DISCRIMINATION (AXB) ANSWER SHEET

Nombre:_____

Instrucciones: Escucha las siguientes palabras en el PowerPoint. Después, indica (circle) la palabra que es más parecida a X.[23]

1)	A	B	11)	A	B	21)	A	B
2)	A	B	12)	A	B	22)	A	B
3)	A	B	13)	A	B	23)	A	B
4)	A	B	14)	A	B	24)	A	B
5)	A	B	15)	A	B	25)	A	B
6)	A	B	16)	A	B	26)	A	B
7)	A	B	17)	A	B	27)	A	B
8)	A	B	18)	A	B	28)	A	B
9)	A	B	19)	A	B	29)	A	B
10)	A	B	20)	A	B	30)	A	B

[23] Although this appendix only shows 30 possible answers, there were 48 total.

APPENDIX E. NATIVENESS JUDGMENT (LIKERT-SCALE) SHEET

Instrucciones: Escucha las siguientes palabras en el PowerPoint. Después, indica (circle) un número para indicar cual se parece más a un hablante nativo de español o un hablante que no es nativo del español.[24]

1)

1 2 3 4 5 6 7

Native Non-native

2)

1 2 3 4 5 6 7

Native Non-native

3)

1 2 3 4 5 6 7

Native Non-native

4)

1 2 3 4 5 6 7

Native Non-native

5)

1 2 3 4 5 6 7

Native Non-native

6)

1 2 3 4 5 6 7

Native Non-native

7)

1 2 3 4 5 6 7

Native Non-native

[24] Although there are seven responses here, there were 24 total.

APPENDIX F. IN-CLASS WORKSHEET FOR EI GROUP

Práctica de pronunciación 1: Trabajo en clase

PASO 1 - Instrucciones: Contesta las siguientes preguntas con un(a) compañero(a).

1. ¿Cómo se pronuncia 'p' en inglés al principio de una oración? ¿Y en español?

   _____

   _____

   _____

2. ¿Qué tipo de producción de 'p' en inglés (o en qué posición de una palabra) es similar a la producción de 'p' al principio de una palabra en español?_____

   _____

   _____

3. ¿Piensas que un hablante nativo de español se nota de la aspiración de 'p' por un hablante nativo de inglés? ¿Por qué o por qué no?

   _____

   _____

   _____

4. ¿Piensas que aprendiste pronunciar 'p' mejor después de la lección hoy? ¿Por qué o por qué no?_____

   _____

   _____

5. Ahora, tú y tu compañero van a repetir las siguientes palabras, tomando turnos (taking turns), diciendo cada palabra 5 veces:
   1. Pego
   2. Palo
   3. Pelo
   4. Paga
   5. Pinta

Ayuda a tu compañero(a) con su pronunciación y dale “feedback”.

6. Ahora, tú y tu compañero van a repetir las siguientes frases, dando “feedback” otra vez de pronunciación. Sólo necesitan decir cada oración UNA (1) vez:
    1. Yo pego una patada a la pelota.
    2. El palo es muy largo.
    3. Mi hermana tiene pelo muy liso.
    4. Ella siempre me paga en pesos.
    5. Esa ensalada tiene muy buena pinta.

APPENDIX G. IN-CLASS WORKSHEET FOR VF GROUP

Práctica 1 con Grabaciones: Trabajo en clase

PASO 1 - Instrucciones: Abre las primeras (4) grabaciones en Praat, y contesta las siguientes preguntas con un(a) compañero(a)

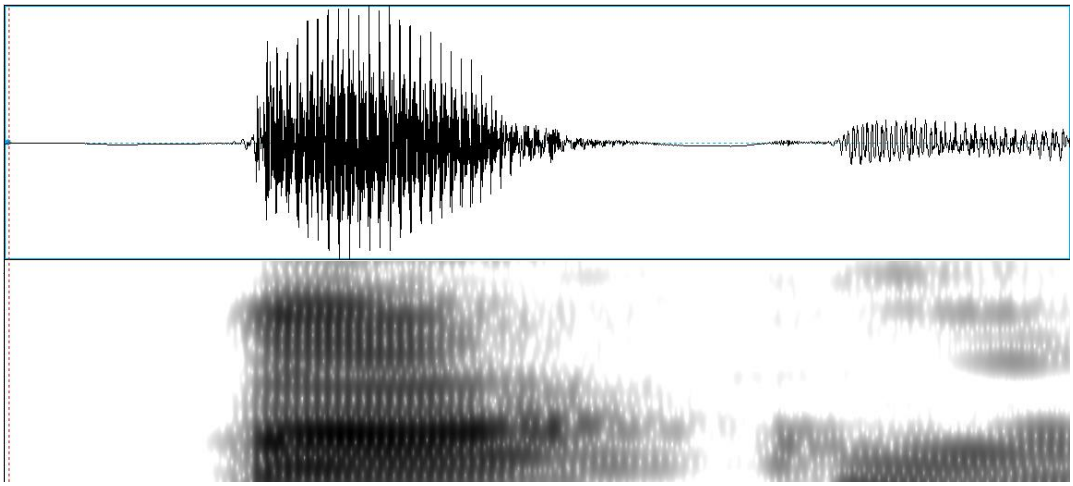
1. ¿Cómo se ve la producción de la 'p' en la foto?

_____

_____

_____

2. Marca donde piensas los límites (boundaries) deberían estar para la 'p' y la 'a' de la palabra ***Paco*** abajo:



Word 1: *Paco*

3. 'Word 1' muestra cómo un hablante nativo pronuncia la palabra '*Paco*.' ¿Has notado algo de la duración (*length*) entre la 'p' de la foto y tu 'p' en tu portátil (laptop)?

_____

_____

_____

4. ¿Son tus grabaciones similares o muy diferentes que 'Word 1'? ¿Cómo?

   _____

   _____

   _____

5. Ahora, tu profesor va poner una grabación de la palabra *'Paco'* producida por un hablante nativo de español, y también, por un hablante de inglés. ¿Cuáles diferencias notas?

   _____

   _____

   _____

6. Ahora, mira las dos grabaciones de la palabra *'Paco'* que tu profesor va a mostrar en la pantalla (*on the screen*). ¿Por qué son diferentes?

   _____

   _____

   _____

APPENDIX H. IN-CLASS WORKSHEET FOR CI GROUP

Práctica de pronunciación 1: Trabajo en clase

PASO 1 - Instrucciones: Contesta las siguientes preguntas con un(a) compañero(a).

1. ¿Cómo se pronuncia 'p' en inglés al principio de una palabra? ¿Y en español?

   _____

   _____

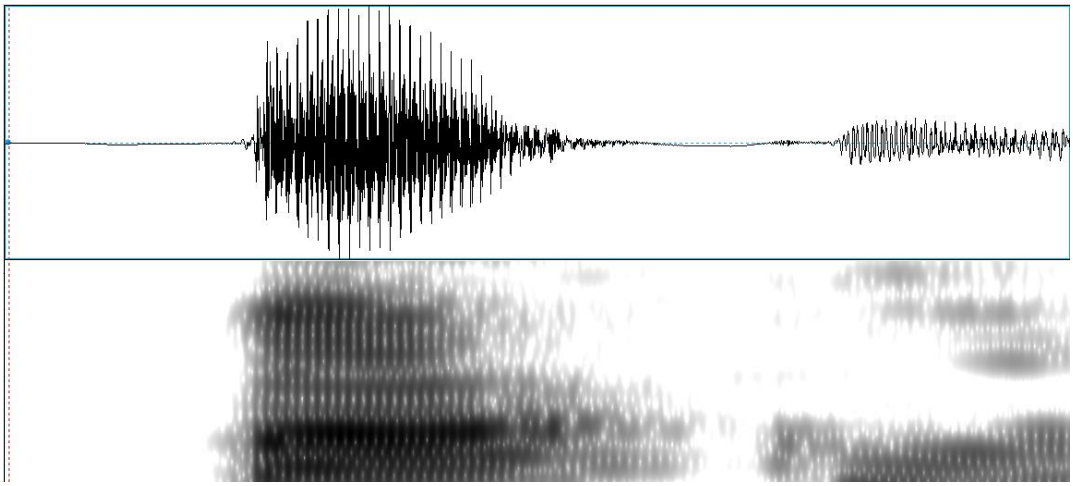
   _____

2. ¿Qué tipo de producción de 'p' en inglés (o en qué posición de una palabra) es similar a la producción de 'p' al principio de una palabra en español?_____

   _____

   _____

3. Marca donde piensas los límites (boundaries) deberían estar para la 'p' y la 'a' de la palabra ***Paco*** abajo:



   Word 1: *Paco*

4. ¿Son tus grabaciones similares o muy diferentes que 'Word 1'? ¿Cómo?

   _____

_____

_____

5. Ahora, tú y tu compañero van a repetir las siguientes palabras, tomando turnos (taking turns), diciendo cada palabra 5 veces:

    6. Pego
    7. Palo
    8. Pelo
    9. Paga
    10. Pinta

    Ayuda a tu compañero(a) con su pronunciación y dale "feedback" con su pronunciación.

6. Ahora, mira las dos grabaciones de la palabra *'Paco'* que tu profesor va a mostrar en la pantalla (*on the screen*). ¿Por qué son diferentes?

_____

_____

_____

APPENDIX I. IN-CLASS WORKSHEET FOR CO GROUP

Cultura: Trabajo en clase[25]

PASO 1 - Instrucciones: Lee el texto en voz alta en la página 14, tomando turnos con un/ una compañero(a).

PASO 2 - Instrucciones: Contesta las siguientes preguntas sobre el texto con tu compañero(a):

1. ¿Qué dice la lectura sobre los jóvenes de inmigrantes? ¿Cuál es su primera lengua?
2. En la calle 4, ¿qué pasa con los mariachis y los tacos?
3. ¿Qué piensas en general en los cambios que están pasando en Santa Ana?

[25] All exercises for the cultural activity were created based on content from the instructor's current textbook.

REFERENCES

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences, 196*(7), 569-573. Academia Prague.

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.

Allaire, J. (2012). *RStudio: Integrated Development Environment for R*. Boston, Massachusetts: RStudio Inc.

Aliaga-Garcia, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In B. Baptista, A. Rauber and M. Watkins (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2–31). Newcastle Upon Tyne: Cambridge Scholars.

Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology, 7*, 617.

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.

Arteaga, D. (2000). Articulatory phonetics in the first-year Spanish classroom. *The Modern Language Journal, 84*, 339–354.

Baran-Łucarz, M. (2012). Individual learner differences and accuracy in foreign language pronunciation in M. Pawlak (Ed.): *New Perspectives on individual differences in language learning and teaching*. (pp. 289–303). Springer.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4 [Computer software manual]. *R package version*,1-1.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289–304). Dordrecht, The Netherlands: Kluwer Academic.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The commonalities and complementarities 35 development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in Infancy Research*, *9,* 217–217.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production. A Festschrift in honour of James E. Flege* (pp.13–34). John Benjamins.

Bigelow, M., Delmas, R., Hansen, K., & Tarone, E. (2006). Literacy and the processing of oral recasts in SLA. *TESOL Quarterly*, *40*(4), 665–689.

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, *35*(4), 374–392.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249.

Boersma, P., & Weenink, D. (2019). Praat: Doing Phonetics by Computer (Computer program). Version 6.1. 02. Retrieved from http://www.praat.org/.

Bond, Z. S., & Fokes, J. (1991). Perception of English voicing by native and nonnative adults. *Studies in Second Language Acquisition*, *13*(4), 471–492.

Bouavichith, D., & Davidson, L. (2013). Segmental and prosodic effects on intervocalic voiced stop reduction in connected speech. *Phonetica*, *70*(3), 182-206.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2),136–193.

Carbó, C., Llisterri, J., Machuca, M. J., De la Mota, C. A. R. M. E., Riera, M., & Ríos, A. (2003). Estándar oral y enseñanza de la pronunciación del español como primera lengua y como lengua extranjera. *ELUA, Estudios de Lingüística de la Universidad de Alicante*, *17*, 161–180.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Chun, D. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal, 7*(1), 21–46.

Chun, D. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 61–77.

Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. John Benjamins.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489–534.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*(4), 389–405.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*(4), 331–350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, *16*, 71–77.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 20*, 1–16.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490.

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

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*, 249–273.

Díaz-Campos, M., & Lazar, N. (2003). Acoustic analysis of voiceless initial stops in the speech of study abroad and regular class students: Context of learning as a variable in Spanish Second Language Acquisition. In P. Kempchinsky & C.E. Piñeros (Eds.), *Theory, practice, and acquisition: Papers from the 6th hispanic linguistics symposium and the 5th conference on the acquisition of Spanish and Portuguese* (pp. 352–370). Cascadilla Press.

Dlaska, A., & Krekler, C. (2008). Self-assessment of pronunciation. *System, 36*, 506–516.

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*(1), 25–37.

Elliott, A. R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal, 79*(3), 356–371.

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania, 80*(1), 95–108.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227–275.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*(2), 339–368.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. [Ph.D. Dissertation, Utrecht University].

Escudero, P. (2009). Linguistic Perception of "similar" L2 sounds. In Boersma, P & Hamann, S. (Eds.), *Phonology in Perception.* (pp. 151–190). Mouton de Gruyter.

Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In *Proceedings of the 26th annual Boston University conference on language development* (pp. 208–219). Cascadilla Press.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452–465.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345–360.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics, 40*(2), 280–288.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech, 53*(3), 343–365.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly, 15*(4), 443–455.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47–65.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of Acoustical Society of America, 89*(1), 395–411.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608.

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 pro- duction: Differences and similarities* (pp. 319-355). Mouton de Gruyter.

Flege, J. E., & Bohn, O.S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11*, 35–62.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*(1), 1–34.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22.

Gilakjani, A. P. (2011). A study on the situation of pronunciation instruction in ESL/EFL classrooms. *Journal of Studies in Education*, *1*(1), 1–15.

González López, V. & Counselman, D. (2013). L2 acquisition and category formation of Spanish voiceless stops by monolingual English novice learners. Cabrelli Amaro J., Lord G., de Prada Perez A., Aaron J. E. (Eds.), *Selected proceedings of the 13th hispanic linguistics symposium*, (pp. 118–127). Cascadilla Press.

González-Bueno, M. (1997). The effects of formal instruction on the acquisition of Spanish stop consonants. *Contemporary Perspectives on the Acquisition of Spanish, 2*, 57–75.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711–2724.

Hammond, R. M. (2001). *The sounds of Spanish: Analysis and application (with special reference to American English)*. Cascadilla Press.

Hardison, D. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, *8*(1), 34–52.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge University Press.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal, 27*(1), 1–12.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America, 122*(5), 2842–2854.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition?. *TESOL Quarterly*, *34*(2), 239–278.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners?. *The Modern Language Journal, 97*(3), 720–744.

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research, 19*(3), 254–275.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366.

Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*(1), 37–63.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369–377.

Levis, J., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal, 12*(2), 13–19.

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly*, *50*(4), 894–931.

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*(3), 1290–1305.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, *33*(1), 42–49.

Lisker, L. (1972). Stop duration and voicing in English in Valdman (Ed.): *Papers in linguistics and phonetics to the memory of Pierre Delattre.* (pp. 339–343). Mouton.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical Measurements. *Word, 20*(3), 384–422.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*, 251–285.

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania, 88*(3), 557–567.

Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals, 41*(2), 364–379.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*(3), 488–503.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37–66.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1–40.

MacDonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect, 17*(3), 3–18.

MacKay, I. R., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*(1-2), 103–125.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition, 13*(3), 279–297.

McGuire, G. (2010). A brief primer on experimental designs for speech perception research. *Laboratory Report, 77*(1). 2–19.

Miller, J. S. (2012). Teaching French pronunciation with phonetics in a college-level beginner French course. *The NECTFL Review, 69*, 47–68.

Mizuno, K. (2007). *The effects of shadowing on learners' acquisition of prosody*. [Master's Thesis, Purdue University]. ProQuest Theses.

Morin, R. (2007). A neglected aspect of the standards: Preparing foreign language Spanish teachers to teach pronunciation. *Foreign Language Annals, 40*(2), 342–360.

Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285–310.

Munro, M. J., Derwing, T. M., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language Teachers. *Prospect, 21*(1), 67–79.

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, *13*(1), 34–65.

Offerman, H. M. (2014). *The effect of visual feedback on voice onset time productions by L2 learners of Spanish.* [Master's Thesis, Purdue University].

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System, 59*, 45–60.

Olson, D. J. (2014a). Phonetics and technology in the classroom: a practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*(1), 47–68.

Olson, D. J. (2014b). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning and Technology*, *18*(3), 173–192.

Olson, D. J. (2019). Feature Acquisition in Second Language Phonetic Development: Evidence from Phonetic Training. *Language Learning*, *69*(2), 366–404.

Olson, D. J., & Offerman, H. M. (Under review). Maximizing the effect of visual feedback for pronunciation instruction: A comparative analysis of three studies.

Oxford, R., Lavine, R., & Crookall, D. (1989). Language learning strategies, the Communicative Approach, and their classroom implications. *Foreign Language Annals*, *22*(1), 29–39.

Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research*, *19*(3), 209–223.

Piske, T., MacKay, I., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29*, 191–215.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

Port, R., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America, 66*(3), 654–662.

R Core Team. (2013). R: A Language and Environment for Statistical Computing (Version 3.2.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Rajab, H. (2013). Developing speaking and writing skills of L1 Arabic EFL learners through teaching of IPA phonetic codes. *Theory and Practice in Language Studies, 3*(4), 653–659.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language learning, 45*(2), 283–331.

Ruellot, V. (2011). Computer-assisted pronunciation learning of French/u/and/y/ at the intermediate level. In J. Levis and K. LeVelle (Eds.), *Proceedings of the 2nd pronunciation in second language learning and teaching conference* (pp. 199–213). Iowa State University.

Saalfeld, A. (2011). Acquisition of L2 phonology in advanced learners: Does instruction make a difference? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference,* (pp. 144–152). Iowa State University.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation in EFL settings: The case of English vowels and Japanese learners of English. *The Linguistics Journal, 3*(3), 16–40.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. *Language Awareness, 20*(1), 45–59.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842–854.

Saito, K. (2013). Reexamining effects of Form-Focused Instruction on L2 pronunciation development. *Studies in Second Language Acquisition, 35*, 1–29.

Saito, K., & Lyster, R. (2012a). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning, 62*(2), 595–633.

Saito, K., & Lyster, R. (2012b). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly*, *46*(2), 387–398.

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

Saville-Troike, M. (1973). Reading and the Audio-lingual Method. *TESOL Quarterly*, *7*(4), 395–405.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.

Schmidt, R. (1992). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226.

Scobbie, J. M. (2006). Flexibility in the face of incompatible English VOT systems. In L.M. Goldstein, C. Best and D. Whalen (Eds.), *Laboratory Phonology 8 Varieties of Phonological Competence* (pp. 367–92). Mouton de Gruyter.

Şenel, M. (2006). Suggestions for beautifying the pronunciation of EFL learners in Turkey. *Journal of Language and Linguistic Studies*, *2*(1), 11–125.

Simões, A. (1996). Phonetics in Second Language Acquisition: An Acoustic Study of Fluency in Adult Learners of Spanish. *Hispania, 79*(1), 87–95.

Sturm, J. L. (2013a). Explicit phonetics instruction in L2 French: A global analysis of improvement. *System, 41*, 654–662.

Sturm, J. L. (2013b). Liaison in L2 French: The effects of instruction. In. J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference,* (pp. 157–166). Iowa State University.

Sturm, J. L., Miyamoto, M., & Suzuki, N. (2019). Pronunciation in the L2 French classroom: Student and teacher attitudes. *French Review*, *92*(3), 60–78.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, *16*(3), 371–391.

Thomson, R. I., & Derwing, T. M. (2004). Presenting Canadian values in LINC: The roles of textbooks and teachers. *TESL Canada Journal, 21*(2), 17–33.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344.

Thorne, S., & Payne, J. S. (2005). Evolutionary trajectories, internet-mediated expression, and language education. *CALICO Journal, 22*(3), 371–397.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal, 90*, 519–535.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology, 6*, 1–12.

VanPatten, B., & Williams, J. (Eds.). (2008). *Theories in Second Language Acquisition: An Introduction*. Routledge.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539–552.

Weltens, B., & de Bot, K. (1984). The visualization of pitch contours: Some aspects of its effectiveness in teaching foreign intonation. *Speech Communication, 3*(2), 157–163.

Whitley, M. S. (2002). *Spanish/English contrasts: A course in Spanish linguistics*. Georgetown University Press.

Willis, E. W. (2006). Trill variation in Dominican Spanish: An acoustic examination and comparative analysis. In N. Sagarra & A. J. Toribio (Eds.), *Selected proceedings of the 9th hispanic linguistics symposium* (pp. 121–131). Cascadilla Press.

Yoshida, K., & Fukada, A. (2016). Effects of oral repetition on learners' Japanese word accentuation. *IALLT Journal of Language Learning Technologies, 44*(1).

Zampini, M. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania, 77*(3), 470–481.

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.

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*, (pp.18–26). Iowa State University.