THE EFFECT OF PHONETICS INSTRUCTION ON ADULT LEARNERS’ PERCEPTION AND PRODUCTION OF L2 SOUNDS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics.

By

Elizabeth Maria Kissling, M.A.

Washington, DC
March 19, 2012
THE EFFECT OF PHONETICS INSTRUCTION ON LEARNERS’ PERCEPTION AND PRODUCTION OF L2 SOUNDS

Elizabeth M. Kissling, M.A.

Thesis Advisors: Jeffrey T. Connor-Linton, Ph.D.
Cristina Sanz, Ph.D.

ABSTRACT

Traditional pronunciation instruction and instruction in L2 phonetics have been shown to improve learners’ L2 accent in some, though certainly not all, cases. Learners in intermediate and advanced Spanish FL courses have shown modest improvement in the pronunciation of some L2 phones after receiving such instruction. However, relatively less is known about how Spanish learners perceive L2 phones or how their perception changes over time and in response to instruction. Yet instruction might improve learners’ ability to perceive, not just produce, L2 phones in more target-like ways. Furthermore, target-like perception may be a necessary precursor to target-like L2 production, as is suggested by several models of L2 phonological acquisition, most notably Flege’s Speech Learning Model.

This study reports on the advantages that first-, second-, and third-year learners of Spanish as an FL (n=95) gained from explicit instruction in Spanish phonetics. Their performance was compared with a control group that received similar input, practice, and feedback but did not receive explicit instruction in Spanish phonetics. The target phones included a variety of consonants that have proven problematic for L1 English speakers: stop consonants (/p, t, k/), approximants ([β, δ, γ]), and rhotics (/ɾ, r/). Learners’ perception and production of the target phones were measured in a pretest, posttest, delayed posttest design
using multiple elicitation tasks: an AX discrimination task, an identification task, and a word list reading production task. Whereas learners in both groups improved their pronunciation equally, the explicit phonetics instruction conferred an advantage for learner’s discrimination and identification of the target phones. The results suggested that the benefit of explicit phonetics instruction was related to perceptive abilities more than productive abilities for several target phones. Instruction was equally beneficial for learners at different matriculation levels, indicating that phonetics lessons could be inserted at multiple stages along the Spanish FL curricular sequence. Finally, given that the experimental and control groups did not differ greatly for many of the tasks and phones, it was argued that a range of instructional methodologies should be explored as viable alternatives to explicit phonetics instruction for improving Spanish learners’ accent.
DEDICATION

Para Fer y Google. No sé qué haría sin vosotros.
ACKNOWLEDGMENTS

My dissertation committee has been exceptional in every way. Jeff Connor-Linton was incredibly supportive throughout my five long years in Ph.D. program. He kept an open mind and an encouraging demeanor, allowing me the freedom to bounce off him even my craziest ideas, and then talk me back down to sanity. Cristina Sanz provided many astute comments on initial drafts, along with some shrewd professional advice and many words of encouragement. Elizabeth Zsiga did her best to help bridge the great chasms in my knowledge of phonetics and phonology. I feel lucky to have worked with each of these tremendous committee members.

This dissertation would not have been possible without the help of several others, to whom I will forever be indebted. Gillian Lord provided the phonetics modules and was always available to answer my questions about her work or mine. My dear friend Germán Zárate-Sández spent many unpaid hours stuck in a recording studio and spent many more paid but equally frustrating hours rating production data, all without a single complaint. Rusan Chen graciously obliged my requests for statistics lessons and software. Beatriz Lado allowed me to poach the PSTM tasks and questionnaire language from her own dissertation. The journal Language Learning provided much-needed funding.

I want to thank my friends in the School of World Studies at Virginia Commonwealth University for allowing me to collect data there. I was encouraged by their enthusiasm about my project. I owe thanks to all the students who participated in the pilot and the main project and the Spanish instructors who encouraged them to participate: Ruth Bradner, Luis Díaz Díaz, Eugenia Muñoz, Hernán Restrepo, and Jane Wallace. I also owe a special thanks to the Media Center directors Antón Brinkwirth and Greg Hellman as well as the lab techs Maria Boykova, Eric Marquis, and Chase Trumbull for their professional assistance.

I am grateful for the many teachers and advisors that have helped me along the road towards my doctorate generally and my thesis specifically, but I would especially like to thank the following people (in alphabetical order): Glynis Cowell, Stuart Day, Robert DeKeyser, Kathryn Murphy Judy, Robert Podesva, Patrick Rebuschat, and Andrea Revesz.

The wonderful friends I have made during my years at Georgetown are just too numerous to name. Each of them has inspired me to continue working and learning in this crazy field we call linguistics. I am in awe of you (and you know who you are!).

Finally, I would like to thank my family for coming to understand why I would spend so many years in school. I am glad to know that I have made you proud (and quite confused).

Though many have helped in the creation of this final product, all mistakes are mine alone.

Many thanks,
Eli
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ................................................................. 1
1.1 Introduction ................................................................................. 1
1.2 The Speech Learning Model .......................................................... 2
1.3 Perception Training Experiments ..................................................... 10
1.4 Phonetics Instruction ................................................................... 12
1.5 Overview of the Present Study ....................................................... 17

CHAPTER 2: PRODUCTION ................................................................. 19
2.1 Introduction and Target Phones ....................................................... 19
  2.1.1 Empirical Studies of Instruction and Spanish L2 Pronunciation ........ 23
  2.1.2 The Present Study ................................................................. 31
  2.1.3 Research Questions and Hypotheses ........................................ 34
2.2 Methods ..................................................................................... 35
  2.2.1 Context ............................................................................... 35
  2.2.2 Participants ......................................................................... 36
  2.2.3 Experimental Design ............................................................ 38
  2.2.4 Phonetics Instruction (+PI) ....................................................... 39
  2.2.5 Control Instruction (-PI) ........................................................ 41
  2.2.6 Production Test ................................................................. 43
    2.2.6.1 Production Test Materials ................................................. 43
    2.2.6.2 Production Test Administration Procedures ....................... 44
    2.2.6.3 Production Test Rating and Scoring Procedures .................. 45
2.2.7 Questionnaires .................................................................54

2.3 Results .................................................................................54

2.3.1 Comparison of +PI and –PI Groups’ Demographic and Language Background ........54

2.3.2 Inter-rater Reliability .........................................................56

2.3.3 Descriptive Statistics .........................................................56

2.3.4 Distribution of Scores ........................................................59

2.3.5 RM ANOVA and Subsequent Contrasts ................................59

2.4 Discussion.............................................................................67

2.4.1 Summary of Results .........................................................67

2.4.2 Implications ......................................................................70

2.4.3 Limitations of the Production Test ....................................73

2.4.4 Conclusion .................................................................74

CHAPTER 3: PERCEPTION .................................................................77

3.1 Introduction ...........................................................................77

3.1.1 Levels of Speech Perception ...........................................77

3.1.2 Methodological Issues in L2 Speech Perception Research ..................................79

3.1.3 Instruction and Spanish L2 Perception ..............................83

3.1.4 Research Questions and Hypotheses ................................87

3.2 Methods ..............................................................................87

3.2.1 Discrimination Test .........................................................88

3.2.1.1 Discrimination Test Materials ...................................88

3.2.1.2 Discrimination Test Administration Procedures ..................90
4.1.1 The Perception-Production Link ................................................................. 125
4.1.2 Research Questions and Hypotheses ...................................................... 129
4.2 Methods ....................................................................................................... 129
4.3 Results ........................................................................................................ 130
  4.3.1 Correlation Analysis .......................................................... .......................... 130
  4.3.2 Descriptive Statistics .............................................................................. 133
  4.3.3 Distribution of Gain Scores ................................................................. 135
  4.3.4 RMANOVA and Subsequent Contrasts .................................................. 136
  4.3.5 RM ANCOVA Analysis of Production Scores ....................................... 150
4.4 Discussion ................................................................................................... 152
  4.4.1 Summary of Results ............................................................................. 152
  4.4.2 Implications ........................................................................................ 153
  4.4.3 Conclusion: A Role for Individual Differences? ................................. 161

CHAPTER 5: CONCLUSION .............................................................................. 164
5.1 Summary of Results ................................................................................ 164
5.2 Pedagogical Implications ........................................................................ 165
5.3 Limitations ................................................................................................ 169
5.4 Contribution and Future Studies .............................................................. 171

Appendix 2.1 +PI Module for \([b, d, g, \beta, \delta, \gamma]\), Introduction .............. 173
Appendix 2.2 +PI Module for \([b, d, g, \beta, \delta, \gamma]\), Explanation ....................... 174
Appendix 2.3 +PI Module for \([b, d, g, \beta, \delta, \gamma]\), Assessment of Explanation with Feedback .... 175
Appendix 2.4 +PI Module on [b, d, g, β, δ, γ], Articulation ………………………………………..176
Appendix 2.5 +PI Module on [b, d, g, β, δ, γ], Assessment of Articulation …………………177
Appendix 2.6 +PI Module on [b, d, g, β, δ, γ], Comparison with English …………………178
Appendix 2.7 +PI Module on [b, d, g, β, δ, γ], Assessment of Comparison with English …179
Appendix 2.8 +PI Module on [b, d, g, β, δ, γ], Discrimination Exercise with Feedback ……180
Appendix 2.9 +PI Module on [b, d, g, β, δ, γ], Pronunciation Practice …………………………181
Appendix 2.10 –PI Instructions Example, for Second Year Learners …………………….182
Appendix 2.11 Transcripts of Video Vignettes Used for First Year Learners in –PI ……….184
Appendix 2.12 Transcripts of Video Vignettes Used for Second Year Learners in –PI …….185
Appendix 2.13 Transcripts of Video Vignettes Used for Third Year Learners in –PI ………186
Appendix 2.14 Production Test Items ……………………………………………………………………….187
Appendix 2.15 Production Test Instructions …………………………………………………………..188
Appendix 2.16 Background Questionnaire for Spanish Learners ………………………………189
Appendix 2.17 Background Questionnaire for Native Spanish Speakers ……………………192
Appendix 2.18 Post-instructional Module Questionnaire …………………………………………..194
Appendix 2.19 Post-study Questionnaire ……………………………………………………………….195
Appendix 3.1 Discrimination Test Pairings ……………………………………………………………….198
Appendix 3.2 Acoustic Analysis of Discrimination Test Stimuli /p, t, k/ ……………………..199
Appendix 3.3 Acoustic Analysis of Vowels in Discrimination Test Stimuli …………………200
Appendix 3.4 Identification Test Items …………………………………………………………………203

References ……………………………………………………………………………………………..204
LIST OF FIGURES

Figure 2.1 Experimental Design ................................................................. 39
Figure 2.2 Spectrogram of Sample [β] Token Assigned 3 Points .................... 48
Figure 2.3 Spectrogram of Sample [β] Token Assigned 2 Points .................... 48
Figure 2.4 Spectrogram of Sample [β] Token Assigned 1 Point .................... 48
Figure 2.5 Spectrogram of Sample [ð] Token Assigned 3 Points .................... 49
Figure 2.6 Spectrogram of Sample [ð] Token Assigned 2 Points .................... 49
Figure 2.7 Spectrogram of Sample [ð] Token Assigned 1 Point .................... 49
Figure 2.8 Spectrogram of Sample [ɣ] Token Assigned 3 Points .................... 50
Figure 2.9 Spectrogram of Sample [ɣ] Token Assigned 2 Points .................... 50
Figure 2.10 Spectrogram of Sample [ɣ] Token Assigned 1 Point .................... 50
Figure 2.11 Spectrogram of Sample /ɾ/ Token Assigned 3 Points .................... 51
Figure 2.12 Spectrogram of Sample /ɾ/ Token Assigned 2 Points .................... 51
Figure 2.13 Spectrogram of Sample /ɾ/ Token Assigned 1 Point .................... 51
Figure 2.14 Spectrogram of Sample /ɾ/ Token Assigned 3 Points .................... 52
Figure 2.15 Spectrogram of Sample /ɾ/ Token Assigned 2 Points .................... 52
Figure 2.16 Spectrogram of Sample /ɾ/ Token Assigned 1 Point .................... 52
Figure 2.17 Production Test (all target phones) ............................................ 61
Figure 2.18 Production Test (8 individual phones) ......................................... 63
Figure 3.1 Discrimination Test ([t,k, β, ð, ɣ]) ................................................ 100
Figure 3.2 Discrimination Test (5 individual phones) ..................................... 102
Figure 3.3 Identification Test (all target phones) .......................................... 109
Figure 3.4  Identification Test (8 individual phones) .......................................................... 111
Figure 4.1  Gain Scores from Pretest .................................................................................. 135
Figure 4.2  Gain Scores (all target phones) ........................................................................ 137
Figure 4.3  Gain Scores for /p/ .......................................................................................... 139
Figure 4.4  Gain Scores for /t/ ........................................................................................... 140
Figure 4.5  Gain Scores for /k/ .......................................................................................... 142
Figure 4.6  Gain Scores for [β] ........................................................................................... 143
Figure 4.7  Gain Scores for [ð] ........................................................................................... 144
Figure 4.8  Gain Scores for [ɣ] ........................................................................................... 145
Figure 4.9  Gain Scores for /ɾ/ ........................................................................................... 147
Figure 4.10 Gain Scores for /ɾ/ .......................................................................................... 148
LIST OF TABLES

Table 2.1 Empirical Studies Describing Learners’ L2 Spanish Development .........................21
Table 2.2 Empirical Studies Assessing Learners’ Pronunciation after Phonetics Instruction ......25
Table 2.3 Unique Tokens of Target Phones in Instructional Modules .................................42
Table 2.4 Auditory and Acoustic Properties Used for Scoring of Production Test .....................47
Table 2.5 Average Voice Onset Time in Production of Stop Consonants (milliseconds) ..........58
Table 2.6 Average Rating (1-3 points) of Approximant and Rhotic Phones .........................58
Table 2.7 RMANOVA of Production Test: Aggregate of 8 Phones .................................60
Table 2.8 Pairwise Comparisons of Production Test: Aggregate of 8 Phones .........................61
Table 2.9 RMANOVA of Production Test: Comparison of 8 Phones ..................................62
Table 3.1 Pairings of Target Phones in Discrimination Test ..............................................89
Table 3.2 Scoring of Identification Test Items ....................................................................94
Table 3.3 Average Score Per Phone on Discrimination Test ..............................................96
Table 3.4 RMANOVA of Discrimination Test: Aggregate of [t,k, β, δ, γ] ..............................100
Table 3.5 Pairwise Comparisons of Discrimination Test: [t,k, β, δ, γ] .................................100
Table 3.6 RMANOVA of Discrimination Test: Comparison of 5 Phones .........................102
Table 3.7 Average Score Per Phone on Identification Test ..............................................106
Table 3.8 RMANOVA of Identification Test: Aggregate of 8 Phones .................................108
Table 3.9 Pairwise Comparison of Identification Test: Aggregate of 8 Phones ......................109
Table 3.10 RMANOVA of Identification Test: Comparison of 8 Phones ......................111
Table 3.11 Most Frequent Error Types on Identification Test ............................................114
Table 4.1 Kendall’s Tau Correlations Between Pretest Scores ...........................................131
Table 4.2 Gains Pretest to Posttest ............................................................134
Table 4.3 Gains Pretest to Delayed Posttest ..............................................134
Table 4.4 RMANOVA of Gain Scores on Three Test Types: Aggregate of 8 Phones ..........137
Table 4.5 Pairwise Comparisons of Three Test Types: Aggregate of 8 Phones ............137
Table 4.6 RMANOVA of Gain Scores on Three Test Types: /p/ ..........................138
Table 4.7 Pairwise Comparisons of Three Test Types: /p/ .................................139
Table 4.8 RMANOVA of Gain Scores on Three Test Types: /t/ ............................140
Table 4.9 Pairwise Comparisons of Three Test Types: /t/ ..................................140
Table 4.10 RMANOVA of Gain Scores on Three Test Types: /k/ .........................141
Table 4.11 Pairwise Comparisons of Three Test Types: /k/ ...............................142
Table 4.12 RMANOVA of Gain Scores on Three Test Types: [β] ...........................142
Table 4.13 Pairwise Comparisons of Three Test Types: [β] ...............................143
Table 4.14 RMANOVA of Gain Scores on Three Test Types: [ð] ...........................144
Table 4.15 Pairwise Comparisons of Three Test Types: [ð] ...............................144
Table 4.16 RMANOVA of Gain Scores on Three Test Types: [ɣ] ...........................145
Table 4.17 Pairwise Comparisons of Three Test Types: [ɣ] ...............................145
Table 4.18 RMANOVA of Gain Scores on Three Test Types: /ɾ/ ..........................146
Table 4.19 Pairwise Comparisons of Three Test Types: /ɾ/ ...............................146
Table 4.20 RMANOVA of Gain Scores on Three Test Types: /ɹ/ .........................148
Table 4.21 Pairwise Comparisons of Three Test Types: /ɹ/ ...............................148
Table 5.1 Summary of Findings .....................................................................167
CHAPTER 1: INTRODUCTION

1.1 Introduction

Foreign accent is very typical of second language (L2) and foreign language (FL) learners.\(^1\) There is little doubt that learners are often concerned with reducing the “foreign” quality of their accent, even though it has been argued that comprehensibility and intelligibility are more important to communicative competence than accentedness (Derwing & Munro, 2009). Comprehensibility refers to listeners’ perceptions about how difficult it is to understand speech, intelligibility is listeners’ actual (objectively measured) comprehension of speech, and accentedness is how different a speaker sounds from a particular local variety (Derwing & Munro, 2009). In some cases, a strong accent may result in reduced comprehensibility and/or intelligibility (Derwing & Munro, 2009), but accent itself is also worthy of study because accent can arouse prejudice on the part of listeners (Eisenstein, 1983; Derwing & Munro, 2009) and anxiety on the part of speakers, who often “worry about pronunciation a great deal because they feel insecure about how they sound to other people” (Harlow & Muyskens, 1994, p. 146). In fact, studies that investigated EFL and ESL learners’ attitudes about pronunciation found that the vast majority of learners wanted to sound like a native speaker (e.g., Derwing, 2003; Timmis, 2002). There is evidence that while native speakers are relatively accepting of low proficiency learners’ “foreign” accents, the possibility that learners will suffer negative evaluations because of their accent increases with increasing L2 proficiency (Galloway, 1980). A wealth of second language acquisition (SLA) research has investigated the phenomenon of foreign accent, with much research falling into one of three broad areas: critical or sensitive period research exploring

---

\(^1\) Except in instances when making a distinction between the two is relevant, henceforth both second language (L2) and foreign language (FL) will be generally referred to as L2, following current norms of nomenclature in the field of second language acquisition (Ortega, 2009, p. 6.).
the relationship between age of acquisition of the L2 and foreign accent (see Flege, 2003a, for a review), research testing the potential of instruction to reduce foreign accent (see Pardo, 2004, for a review), and development of various theoretical models that account for empirical data on incomplete acquisition of the L2 (see Colantoni & Steele, 2008).

1.2 The Speech Learning Model

Researchers have developed a variety of theoretical frameworks to account for the complex process of how second language sound systems are acquired. Their frameworks tend to model either perception or production of the second language (see Eckman, 2004; and Leather, 1999, for reviews). Many prominent perception-based models, such as Flege's Speech Learning Model (Flege, 1988, 1992, 1995, 1999, 2002), Best’s Perceptual Assimilation Model (Best, 1994; 1995; Best & Tyler, 2007), and Kuhl’s Native Language Magnet Model (Kuhl, 1993; Kuhl et al., 2008), assign a critical role to the first language (L1) of the learner and make (sometimes contradictory) predictions about acquisitional stages based on the similarities and differences between the L1 and L2 phonological systems. The Perceptual Assimilation Model claims that the perceptual primitive in speech perception is distal articulatory events, i.e. gestural constellations, as opposed to proximal acoustic cues, i.e. spectral patterns (Best, 1995). According to the Perceptual Assimilation Model, throughout her lifespan a listener’s perceptual system becomes attuned to detect the higher-order invariants in her L1-specific gestural constellations, thereby becoming increasingly worse at detecting what is not contrastive in the L1. The Native Language Magnet Model (Kuhl, 1993; Kuhl et al., 2008) focuses on L1 acquisition during the first year of life when phonetic prototypes are created and the impact that

---

2 “Most dominant language(s)” may be a better way to encapsulate previous language experience than “language first learned” (L1). However, in order to faithfully represent the literature cited, the term “L1” will be used here.
this development has on L2 perception later in life, which is to effectively shrink the acoustic space in towards the L1 phonetic prototypes such that L2 sounds are perceived as more or less acceptable exemplars of the prototypes.

Other models, such as Eckman’s Markedness Differential Hypothesis (Eckman, 1977, 1991), Major’s Ontogeny Phylogeny Model (Major, 1986, 1987, 2001), and various Optimality Theoretic models (e.g., Escudero & Boersma, 2004)\(^3\), while not denying the critical role that L1 plays, particularly in the early stages of L2 acquisition, also posit a relatively greater role for universal tendencies and factors of markedness. The Markedness Differential Hypothesis, a production-based model, posits that structures that are different in the L2 and the L1 will be difficult to acquire only if they are also more marked in the L2, with markedness being explained in terms of universal implicational hierarchies. The Ontogeny Phylogeny Model posits that production errors in L2 phonology are of two types: transfer from the L1 or developmental errors, i.e., errors related to markedness (Major, 1986, 1987, 2001). According to this model, in the initial stages of learning L2 learners tend to display high rates of transfer errors. Over time the transfer errors decrease while developmental errors increase, and finally developmental errors decrease as well, such that learners’ phonology becomes more target-like.

While all of the models discussed thus far describe particular aspects of L2 phonological acquisition quite well, their applicability to the present study is limited. The models of L2 perception have been developed to explain empirical evidence drawn from populations of naïve language learners – infants in the case of the Native Language Magnet Model and L2 learners  

\(^3\) Though optimality theory is promising, “it has, to date, had little impact on the field of L2 acquisition. Relatively few studies in L2 phonology have adopted an OT approach to understand observed phenomena in L2 learning, and the majority of these studies focus on L2 syllables” (Hancin-Bhatt, 2008, p. 164).
with brief (i.e., generally 6 months or less) naturalistic exposure to the L2 in the case of the Perceptual Assimilation Model. In contrast, the present study recruited adult, relatively experienced foreign language learners in an instructed learning context. Thus the Perceptual Assimilation Model and the Native Language Magnet Model are unsuitable models for the population under investigation here. The production models (the Markedness Differential Hypothesis and the Ontogeny Phylogeny Model) make no predictions about learners’ perception of L2 sounds and therefore are unsuitable for the present study, which undertook an exploration of learners’ perceptions. On the other hand, the Speech Learning Model (SLM) is a model that accounts for both production and perception data and even explicitly draws a link between the two. The SLM has focused mainly on learners who acquired the L2 in a naturalistic setting and have reached their ultimate attainment, but it is still the model most commonly used in research on L2 phonological acquisition within formal instructional settings (Fullana, 2006).

The SLM was developed by James Flege (e.g., Flege, 1988, 1992, 1995, 1999, 2002) to explain the acquisition of L2 speech and age-related effects seen across a variety of L2 learner populations. The postulates of the SLM claim that “language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories,” that the phonetic categories of the L1 and the L2 exist in one common phonological space, such that the categories will “evolve over the life span to reflect the properties of all L1 or L2 phones identified as realizations of each category,” and that “the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning” (Flege, 1995, p. 239).
The focus of the SLM has been speech production, which it hypothesizes to be constrained by perception. That is, a learner must attune her perception of L2 speech sounds to become more native-like before her production of L2 speech sounds can become more native-like. Such attunement often requires the creation of new categories for L2 sounds. The SLM is not the only perception-based model (see Munro, 2008 for review) that posits a relationship between perception and production and suggests that perception somehow leads production (e.g., Strange & Shafer, 2008). The nature of the link between perception and production has been characterized in rather different ways, but stated in perhaps the most simple and general terms, some attentional detection (perception) is thought to be necessary for further processing and storage in long term memory which, in turn, contains the phonological knowledge used for production. The SLM makes the strongest of the theoretical claims, arguing that target-like perception of the L2 is a necessary precursor to target-like production.

Perception is not a monolithic construct. Rather, there are multiple levels of processing involved in the act known as perception. Wode (1981; 1990) described two modes of processing, one “continuous” and one “categorical.” The “categorical” mode is dependent on the phonemic categories built up over time with exposure to one’s first language(s), whereas the “continuous” mode is relatively independent and is used for non-speech sounds, among other things. Strange (2002) referred to the same distinction as “categorial” versus “categorical.” Werker and colleagues studied the development of infants’ perception of speech and concluded that infants begin life processing phonetically and acoustically (akin to processing in Wode’s “continuous” mode) and over time transition to more phonemic (“categorical”) processing (e.g., Best, McRoberts & Sithole, 1988; Werker, Gilbert, Humphrey & Tees, 1981; Werker & Tees, 1983), with the ambient/native language(s) serving as a magnet that warps perception around L1-
informed phonemic categories. Dehaene-Lambertz and colleagues (Dehaene-Lambertz et al., 2005; Dehaene-Lambertz & Gliga, 2004) have presented behavioral and neurological evidence indicating that adults maintain the ability to process speech using either perceptual mode, but under normal conditions the phonemic (“categorical”) network tends to inhibit the acoustical (“continuous”) network.

The problem in L2 speech perception, then, is that adults tend to use categorical/phonemic processing in the L2 even though their phonological knowledge of the L2 is imperfect. Adults tend to use the same categories they have developed over years of exposure to their L1(s). Thus models of L2 phonological acquisition such as the SLM imply that learners must switch to continuous or phonetic processing at some point in order to be able to discern subtle phonetic differences between L2 sounds and analogous sounds in the L1. Said another way, the process of acquiring an L2 sound system begins first by detecting differences between native and non-native sounds and then developing the appropriate selective perceptual routines (SPRs) to “hear” them reliably (Strange & Shafer, 2008).

Unfortunately, not all adults are equally well equipped for the task. Munro (2008) found that cognitive factors, specifically phonological short term memory, underpins continuous processing abilities. Munro argued that the perception-production link is never broken, but “that individuals likely differ widely in numerous aspects including their perceptual abilities, their spoken accent, and the degree to which these interdepend. In sum, the ability to pronounce novel phones correctly may not only be limited by capacity for motoric rehabituation, but also antecedently on perceptual propensity for capturing a detailed likeness of the incoming phonetic signal which is faithful to the native model” (Munro, 2008, p. 29), that is, their ability to switch from a categorical to a continuous mode of processing in order to learn new L2 categories.
Of course, though target-like perception may be a prerequisite for target-like production, perception alone is not sufficient. The “motoric rehabilitation” to which Munro (2008) referred was the learning of new articulatory patterns for L2 phones. Chastain (1976) observed long ago that L2 learners commit at least two types of speech errors. They mispronounce sounds because they cannot hear them, and they mispronounce sounds that they can hear because they cannot articulate them. Recently Colantoni and Steele (2008) argued for developing a model of second language phonological acquisition that integrates articulatory constraints. Even the more comprehensive model they advocate continues to contribute great importance to perception, but it also accounts for other factors that impinge upon production in addition to perception.

Though the SLM does not characterize the acquisition process in terms of continuous/phonetic versus categorical/phonemic processing, the SLM delineates two perceptual stages in the acquisition process: first, a learner must detect at least some of the phonetic differences between L2 sounds and their analogous L1 counterparts, and then she can create new phonetic categories for L2 sounds. Depending on the L1/L2 phones in question, the phonetic differences that distinguish them may consist of voice onset timing, aspiration, duration, or any number of other phonetic characteristics in isolation or in combination. While the SLM does not stipulate a threshold of phonetic difference beyond which the L1/L2 contrast will necessarily be discerned by L2 learners (Flege, 2003a, p. 329), it does posit that the greater the phonetic difference between the L2 and analogous L1 sounds, also called “diaphones,” the more likely a learner is to detect the difference. For instance, Japanese speakers are better able to discern a difference between Japanese /ɾ/ (an apical post alveolar flap undefined for laterality) and American English /ɹ/, which are phonetically more dissimilar, than they are able to discern Japanese /ɾ/ from American English /l/, which are phonetically more similar (Sekiyama & Tohkura, 1993; Takagi,
The SLM further hypothesizes that the perceptual relationship between L1 and L2 sounds lies at a position-sensitive, allophonic level, not the more abstract phonemic level. Some evidence to that effect is that the target-likeness with which Japanese speakers produce and perceive English /l/ and /ɹ/ varies across phonological environments (Strange, 1992; Sheldon & Strange, 1982).

Flege and colleagues as well as several other researchers have cited empirical evidence that perceptive abilities lead productive abilities (e.g., Flege, 1988; Flege et al., 1997; Newman, 1996; Rochet, 1996). For instance, Munro (2008) conducted a series of highly controlled experiments with Mandarin speaking (n=37) learners of English and concluded that “spoken accent is dependent upon the ability to perceive subtle phonetic differences,” for a wide variety of problematic L2 English phones (p. 78). Flege and colleagues (Flege, MacKay & Meador, 1999; Flege & MacKay, 2004) tested Italian participants’ (n=72) ability to discriminate English and Italian vowel pairs and produce English vowels (Flege, MacKay & Meador, 1999). They found a much stronger correlation between participants’ ability to produce and perceive (i.e., discriminate) English vowels than between their ability to perceive Italian vowels and produce English vowels. While Flege and colleagues admitted that some learners were simply better at perceiving and producing sounds overall, they interpreted their findings as evidence that a link exists between the perception and production of language-specific phonetic segments, a relationship that exists regardless of learners’ individual aptitudes. Furthermore, the subgroup (n=41) of participants that had relatively target-like production of particularly difficult vowels such as /ʌ/ also discriminated those same vowels better, though their overall vowel discrimination scores did not differ significantly from the rest of the participants. Empirical
evidence from those studies as well as others prompted Flege to conclude that “L2 segmental perception may 'lead' (i.e., be more advanced than) production” (Flege, 2003a, p. 345).

The SLM postulates that the mechanisms and processes underlying phonetic category creation remain intact over the life span and are the same for L1 and L2 learning, thus eschewing the notion of a critical period for L2 speech learning. However, the SLM does not ignore age effects altogether. According to the SLM, age effects obtain because the likelihood that learners will detect phonetic differences between analogous L1 and L2 sounds, as well as between L2 sounds that are non-contrastive in the L1, diminishes as age of learning increases (Flege, 1995, p. 239). For example, studies (e.g., Butcher, 1976) suggest that younger German learners of English are able to discern a difference between English /æ/ and the closest German vowel (/ɛ/) relatively well, but older learners are not. When learners are unable to detect phonetic differences between analogous L1/L2 sounds, formation of a new L2 phonetic category is blocked by a process the SLM calls category assimilation. Learners assimilate instances of the L2 sound into an existing L1 phonetic category, which Flege describes alternatively as a “filtering” into the L1 or a “warping” by the L1 (e.g., Flege, 2003a). Category assimilation may persist indefinitely for some, resulting in a persistent “foreign accent” in both perception and production, but learners can overcome category assimilation if they are given enough input (Flege, 2003a, p. 327).

In sum, the SLM delineates what the process of acquiring the L2 sounds entails: learners must first be able to discern subtle phonetic differences between L2 sounds and analogous sounds in the L1 as they occur in a variety of phonological environments, at which point they are able create new categories for L2 sounds, and finally they may become able to produce the L2 sounds in target-like ways. Though the SLM does not directly address the role that instruction
could play in accelerating the acquisition process, its postulates have been interpreted as advocating for formal instruction (e.g., Fullana, 2006). The current study aims to elucidate the role of instruction in learners’ L2 production and perception.

1.3 Perception Training Experiments

Having argued that perception and production of an L2 are closely linked abilities and that both contribute to the phenomenon of “foreign” accent, the question now at hand is whether instruction is equally necessary or even helpful for improving both abilities. Many empirical studies on L2 speech perception have been conducted in laboratories using experimental paradigms such as the High Variability Phonetic Training Paradigm or the Perceptual Fading Technique (see Logan & Pruit, 1995, for a discussion of procedures). Researchers found that, following such intensive exposure, participants generally improved their ability to perceive L2 contrasts (e.g., Golestani & Zatorre, 2009; Jamieson & Morosan, 1986, 1989) and they could retain this ability for several years (Cenoz & García Lecumberri, 1997), although there was tremendous variation across individuals in initial abilities and improvement post-training (Takagi, 2002). Perhaps most well studied is the English /ɹ/-/l/ contrast among Japanese speakers (e.g., Bradlow, Pisoni, Yamada & Tohkura, 1997; Bradlow, Yamada, Pisoni & Tohkura, 1999; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994; Logan, Lively & Pisoni, 1991; McCandliss et al., 2002). This research tended to focus on perception training, but several studies showed that training in either modality –production or perception– could result in gains in the other, at least with some linguistic targets, including Chinese lexical tones (Leather, 1990), reduced vowels in English (Lacabex, García Lecumberri & Cooke, 2009), and the English /ɹ/-/l/ contrast
(Bradlow et al., 1997). These studies suggested that perception and production instruction might be mutually facilitative.

The evidence was particularly strong that perception training improved production (e.g., Aliaga-García & Mora, 2009; Bradlow, Pisoni, Yamada & Tohkura, 1997; Callan et al., 2003; Jamieson, 1987; Rochet, 1996; Wang, Jongman & Sereno, 2003; Yamada, Tohkura, Bradlow & Pisoni, 1996). This effect was found for different languages and linguistic targets. For instance, Japanese listeners trained to identify English words with /ɹ/ and /l/ not only improved their ability to perceive the contrast but also became more accurate in their production of the contrast after training (Bradlow et al., 1997; Yamada et al., 1996). Rochet trained Mandarin speaking learners of French (n=12) with the Perceptual Fading Technique to better perceive and pronounce French /y/ (1996). EFL learners trained to discriminate /p/-/b/ and /t/-/d/ based on voice onset timing (VOT) and to discriminate /i:/-/I/ and /æ/-/ʌ/ based on spectral cues improved their perception and production of the stop consonant contrasts (but not vowels) after training (Aliaga-García & Mora, 2009). Mandarin leaners produced tonal contrasts better after perceptual training (Wang, Jongman & Sereno, 2003). Even though not all learners benefited equally from perceptual training (Munro, 2008, p. 29), it has been suggested than an extended period of perception training is optimal in the early stages of L2 learning, before learners produce (Bongaerts, van Summeren, Planken & Schils, 1997; Neufeld, 1997).

With few exceptions, the abovementioned studies were carried out in a laboratory context, providing extensive and intense training with highly controlled stimuli and one of several established perception training techniques. For example, Aliaga-García and Mora (2009) recruited 36 advanced Spanish/Catalan EFL learners. They used the High Variability Phonetic Training Paradigm and exposed learners to stimuli with four pairs of contrasting phonemes: /p/-
/b/, /v/-/d/, /iː/-/I/ and /æ/-/ʌ/. This training was continued for 12 hours over the course of several weeks and was found to significantly improve learners’ perception and production of the target phones. Yet replicating such conditions, with 12 or more hours of highly controlled auditory exposure to just 4 sets of contrastive phonemes, is unfeasible for most L2 classes. The same techniques might be modifiable and incorporable in some classes (Sawallis & Townley, 2007), but it is important to caution that the results of most laboratory training studies not be viewed as indicative of the effects of formal instruction. Formal phonetics instruction is prototypically less time intensive and more explicit in nature. Formal phonetics instruction also typically targets a wide range of L2 features, including suprasegmentals and phones that may not contrast in the L2. All of the laboratory studies reviewed thus far have tested contrastive phonemes only, but L2 learners are also concerned with many aspects of their accented production that go beyond acquiring L2 phonemic contrasts (Munro & Derwing, 2011). Non-contrastive phones are often emphasized in traditional phonetics instruction and are the linguistic targets of the present study as well. Typical classroom instructional conditions are still relatively under studied (Derwing & Munro, 2009; Lord, 2005), and there is need for more research investigating the effects of the less intensive, classroom-based phonetics instruction that could be incorporated into foreign language curricula. Classroom-based instruction has been shown to improve learners’ pronunciation moderately in some cases (Pardo, 2004; Piske, MacKay & Flege, 2001), but perception is rarely addressed in this context. The present study aimed to fill these gaps.

1.4 Phonetics Instruction

After reviewing decades of empirical studies on pronunciation instruction, Pardo made several conclusions: 1) Instruction changes learners’ pronunciation; 2) type of instruction is a
determining factor; 3) learners’ needs play a strong role, 4) there are specific teaching techniques, pronunciation is not simply “picked up;” and 5) input and access to input may play a more significant role than previously assumed (2004). The last two conclusions are not, as they might seem, at odds. Both rich input, i.e., exposure to target-like speech, and formal instruction are beneficial. Investigations of the effect of formal instruction on accent found that while general language instruction was not related to global foreign accent (Piske, MacKay & Flege, 2001), pronunciation instruction had a significant effect on L2 production accuracy (Neufeld, 1977; Piske, MacKay & Flege, 2001) in several FL contexts, including English (Pennington, 1992; Pennington & Richards, 1986), French (Clark, 1967; Walz, 1980), German (McCandless & Winitz, 1986; Moyer, 1999), and Spanish (Elliott, 1995; Elliott, 2003; Lord, 2005). In fact, it was suggested that adult L2 learners could not achieve native-like pronunciation without the help of instruction (Bongaerts, van Summerin, Planken & Schils, 1997; Fullana, 2006) and that instruction was required to improve the pronunciation of “fossilized” learners (Derwing, Munro & Wiebe, 1997). Even very brief pronunciation instruction was touted as beneficial (Wipf, 1985). Drawing learners’ attention to particular acoustic features of the L2 system may be more expedient than merely exposing them to L2 sounds in the hopes that they will discover those relevant acoustic features. Guion-Anderson described the learning process in this way: “classroom teachers provide explicit instruction about what to ‘listen for’ in order to allow students to orient their attention toward relevant aspects of the L2 speech and develop their ‘hooks’ into the L2. Once this is accomplished, learners can break out of their automatic processing routines and take in the acoustic information relevant to the L2 for further processing” (in press, p. 6).
The instruction employed in the abovementioned studies included several components considered important for facilitating learners’ acquisition of the L2 phonological system and thereby, it was hoped, improving their pronunciation. The instruction was usually characterized as “pronunciation instruction,” though it will be referred to in the present study as phonetics instruction. The central component was explicit teaching of L2 phonetics, typically at the segmental level and emphasizing phonetic parameters relevant to segments, e.g., airstream mechanisms, places and manners of articulation, voicing, etc. Phonetics instruction in the FL classroom emphasized the differences between learners’ L1 and the L2 phonological systems with regards to phonemic inventories, articulation of analogous phones, grapheme-phoneme correspondences, and phonological processes. Older instructional techniques (e.g., Clark, 1967) used static drawings of the vocal tract to illustrate articulatory targets. More contemporary phonetics instruction (e.g., Lord, 2005) added animated diagrams of the vocal tract and employed acoustic analysis software to provide visual representations of spectral features, contrasts of duration, intonation contours, and other relevant phonetic features. In addition to providing explicit information aimed at building learners’ explicit knowledge of L2 phonology, traditional instruction also included activities that allowed learners to practice their perceptive and productive skills. Phoneme discrimination and identification exercises were sometimes employed, usually with feedback. Pronunciation practice was always employed, though the format varied considerably from word list reading to tongue twisters and jazz chants. Pronunciation feedback varied as well, from simple teacher-fronted modeling to the highly individualized, immediate visual and auditory feedback provided by acoustical analysis software packages. All these exercises, in addition to the core component of explicit teaching of the L2
phonology, were thought to facilitate acquisition of target-like L2 pronunciation (Arteaga, 2000; Elliott, 2003, p. 38).

Though phonetics instruction traditionally emphasized segmentals, several researchers argued that suprasegmentals and more global aspects of speech were more important. Anderson-Hsieh found suprasegmental instruction beneficial for reducing the perceived accent of international teaching assistants (1990). Derwing, Munro, and Wiebe (1997) compared instruction with a segmental focus to more global instructional techniques, such as tapping out beats and finding rhythm in nonsense syllables. The instruction was provided for 20 minutes each day over a 12-week speech course for adult ESL learners (n=48). Both types of instruction resulted in improved comprehensibility, intelligibility, and accent compared to a control group, but the global orientation was associated with greater gains. Derwing and Rossiter (2003) reanalyzed the narrative task data of the previous study and concluded that an emphasis on prosody was most favorable for improving the comprehensibility and fluency of learners’ speech. Missaglia (1999) likewise compared “prosody-centered” instruction to “segment-centered” instruction for inexperienced Italian speaking learners of German. The prosody-centered instruction was associated with more improvement in learners’ suprasegmental and segmental production. It has been suggested that the tendency of teachers and to focus on segmentals has more to do with their teachability than importance for communication (Dalton & Seidlehofer, 1994). That is, teachers find segmentals easier to teach because segments are generally easier to isolate, define, and generalize than are suprasegmental features such as intonation (Dalton & Seidlehofer, 1994).

For all the apparent benefit of instruction, both segmental and suprasegmental, a thorough review of the empirical studies indicates that their results were in fact complex and sometimes
even contradictory (Elliott, 2003; Piske, MacKay & Flege, 2001). Some studies reported that pronunciation instruction had little to no effect on learners’ pronunciation accuracy (Purcell & Suter, 1980; Suter, 1976; Tominaga, 2009). Others concluded that instruction improved segmental production but not comprehensibility (Derwing, Munro & Wiebe, 1997) or vice versa (Saito, 2011). Still others indicated that instruction was not equally effective for all the phones targeted by the instruction (e.g., González-Bueno, 1997; Lord, 2005). Instruction may be best at building up explicit knowledge about the L2 phonological system that supports execution of only certain types of tasks, such as reading words from a list (Venkatagiri & Levis, 2007), which are not useful beyond the classroom context.

Finally, there is simply too little empirical evidence to justify any grand claims about the effectiveness of phonetics instruction (Munro & Derwing, 2011). In fact, “a large body of speculative and opinionated commentary on pronunciation has been published, much of which has never been submitted to empirical test” (Munro & Derwing, 2011, p. 316). It seems that second language acquisition researchers still disagree as to whether learners can actually employ the kind of explicit knowledge about L2 phonology they learn from formal instruction to effectively change the patterns of their speech. A range of factors likely effect the results of instruction, including but not limited to the developmental readiness of learners, the languages pairs and phones targeted, differences in instructional techniques, and time on task.

The evidence is a particularly sparse with regards to how phonetics instruction might impact learners’ perception of the L2. Nearly all the published research on classroom-based instruction reported changes in learners’ production rather than perception, with few exceptions (e.g., Champagne-Muzar, Scheneideran & Bourdages, 1993; Matthews, 1997; Yule, Hoffman & Damico, 1987). Champagne-Muzar, Scheneideran and Bourdages (1993) provided French
learners with 12 hours of phonetics instruction focused on both segmentals and suprasegmentals, the first half of which was devoted to an “auditory sensitization” period in which learners remained silent. Compared to a control group, the instructed learners made significant improvement on both phoneme discrimination tasks and production tasks. Matthews (1997) provided instruction on 6 English segmental contrasts (/ŋ/-/l/, /s/-/ʃ/, /b/-/v/, /θ/-/f/ and /s/-/θ/) to Japanese speaking learners of English once each week for 5 weeks. The instruction emphasized articulatory explanations, visual demonstrations, and mimicking. Instructed learners made significant improvement in their ability to discriminate segments, as compared with a control group, though they only improved some contrasts (/b/-/v/, /θ/-/f/ and /s/-/θ/). Thus Matthews (1997) provided evidence that phonetics instruction could improve learners’ perception of L2 sounds. However, Yule, Hoffman and Damico (1987) found just the opposite, that the immediate impact of instruction was a decrease in learners’ accuracy on a phoneme identification task, along with an improvement in learners’ ability to monitor their own accuracy. The authors concluded that their learners had entered a transitional stage in phonological development, so the effect of instruction was positive in terms of L2 development.

1.5 Overview of the Present Study

The present study attempted to bridge several gaps in the literature by empirically testing the effectiveness of traditional, classroom-based phonetics instruction for improving both the productive and perceptive skills of instructed learners of Spanish. Perception was assessed with discrimination and identification tasks. The effect of explicit phonetics instruction was measured against a control condition that provided the same putatively beneficial activities and practice as the experimental condition but lacked explicit instruction in L2 phonetics. Learners were
recruited from three different matriculation levels in order to address issues of developmental readiness and curricular sequencing.

The rest of the thesis is organized into four chapters. As pronunciation has figured more prominently in the literature than perception, the production data are presented first. Chapter 2 explains the methods and experimental design of the entire study and then presents analyses of the production data. Chapter 3 presents analyses of the perception data. Chapter 4 discusses the relationship between perception and production. The introduction of each of these three analytic chapters reviews the literature that is most pertinent to the chapter’s analysis, with an emphasis on Spanish L2 acquisition. Chapter 5 summarizes the results, discusses the pedagogical implications of the study, lists limitations, and suggests ideas for future research.
CHAPTER 2: PRODUCTION

2.1 Introduction and Target Phones

Chapter 1 provided a review of studies that measured the effect of general language instruction and phonetics (or “pronunciation”) instruction for a variety of second languages. The current chapter focuses specifically on Spanish, the target language of the present study.

The targets phones addressed in the current study were 8 Spanish consonantal phones, [p, t, k, β, ð, γ, r, r]. These phones were chosen because they are widely recognized as late acquired by native speakers of English (e.g., Castino, 1992; Dalbor, 1997; Díaz-Campos, 2004; Face & Menke, 2010; Reeder, 1998; Rose, 2010; Simões, 1996; Waltmunson, 2006; Zampini, 1993) and because they have been examined repeatedly in investigations of pronunciation instruction (Castino, 1996; Elliott, 1995, 1997; González-Bueno, 1997; Lord, 2005)\(^4\). What follows is a description of the phonetic properties of the target phones, a comparison of the phones with analogous English phones, and a summary of the empirical evidence regarding English speakers’ acquisition of the phones. This should not be taken as a suggestion that L1 is the only factor that impacts production or perception of the L2. Nevertheless, L1 figures as enormously important in the L2 speech and perception literature, especially for L2 learners who are immersed in a heavily L1-dominant environment, as were the learners in the present study. Therefore, the ensuing depiction of the target phones will focus on differences between English and Spanish and will review only studies that recruited English speakers learning Spanish in instructed settings. The discussion of the target phones will be brief, but the reader may consult Waltmunson (2006) for more detailed descriptions of the articulatory and acoustic properties of the target phones as well

---

\(^4\) Much less research has investigated Spanish learners’ acquisition of suprasegmental features Nibert (2005, 2006).
as a more detailed discussion of English speakers’ difficulties pronouncing the target phones. Table 2.1 summarizes the results of studies that measured Spanish learners’ phonological development with regards to the target phones.

The phonemes /p, t, k/ are aspirated in English when they occur in syllable-initial position and in stressed syllables, creating the allophones [pʰ, tʰ, kʰ]. In Spanish, however, /p, t, k/ are not aspirated. The underlying difference is in voice onset timing (VOT), which is the time between the release of the stop closure and the start of vocal fold vibration. Long VOT values are associated with aspiration (Hualde, 2005). In Spanish, /p, t, k/ are realized with short VOT, while in English /p, t, k/ in stressed syllables are realized with long VOT. It has been reported that the average VOT, in milliseconds, is 4 ms for Spanish [p] as compared to 58 ms for English [pʰ], 9 ms for Spanish [t] as compared to 70 ms for English [tʰ], and 29 ms for Spanish [k] as compared to 80 ms for English [kʰ] (Lisker & Abramsom, 1964). Native speakers of English tend to aspirate /p, t, k/ in stressed syllables when speaking Spanish, leading to a noticeable foreign accent (Hualde, 2005; Lord, 2005). There is evidence that learners’ VOT diminishes over time as they gain more experience with Spanish but does not fall within native speaker range even at very advanced levels of L2 Spanish proficiency (Reeder, 1998). For instance, the advanced learners in Lord’s (2005) study produced an average VOT of 29 ms for /p/, 30 ms for /t/, and 43 ms for /k/. Though the discrepancy between learners’ VOTs and target-like VOTs may seem small, listeners are in fact quite sensitive to variations in VOT and have been known to detect “accented” speech with only 30 ms of speech on which to base their judgments (Flege, 1984).
<table>
<thead>
<tr>
<th>Author</th>
<th>Participants' Level(s)</th>
<th>N</th>
<th>Targets</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castino (1992)</td>
<td>3rd semester</td>
<td>20</td>
<td>/ptkβγηγγx/</td>
<td>Markedness predicted difficulty.</td>
</tr>
<tr>
<td>Díaz-Campos (2004)</td>
<td>study abroad program or 'advanced' at home language course</td>
<td>46</td>
<td>/ptkβδγγγγ/</td>
<td>No advantage for study abroad. [β δ γ] were highly resistant to improvement.</td>
</tr>
<tr>
<td>Face &amp; Menke (2010)</td>
<td>4th semester to professional level</td>
<td></td>
<td>/ηη/</td>
<td>Experience correlated with target-like production, but even the most advanced were not always target like.</td>
</tr>
<tr>
<td>Reeder (1998)</td>
<td>novice to advanced levels</td>
<td>70</td>
<td>/aeiouptksr/, orthographic /h/</td>
<td>All phones improved with increasing experience, but even very advanced learners were not within NS range.</td>
</tr>
<tr>
<td>Rose (2010)</td>
<td>3rd semester to graduate level</td>
<td>24</td>
<td>/ηη/</td>
<td>Delineated several developmental stages in acquisition of /ηη/.</td>
</tr>
<tr>
<td>Simões (1996)</td>
<td>5 weeks study abroad program</td>
<td>5</td>
<td>/aeiou/, overall fluency</td>
<td>Vowels were inconsistent. Fluency generally improved, but not significantly for all learners.</td>
</tr>
<tr>
<td>Watsonmunson (2006)</td>
<td>1st-6th semesters</td>
<td>22</td>
<td>/tt γγ/</td>
<td>Assessed relative degree of difficulty.</td>
</tr>
<tr>
<td>Zampini (1998)</td>
<td>advanced phonetics course</td>
<td>13</td>
<td>/p b/</td>
<td>No correlation found between learners' perceptual abilities and production of /p b/.</td>
</tr>
<tr>
<td>Zampini (1993)</td>
<td>2nd &amp; 4th semesters</td>
<td>32</td>
<td>/bdgβδγ/</td>
<td>Learners were highly accurate in producing /bd g/ but produced [β δ γ] in fewer than 25% of required contexts.</td>
</tr>
</tbody>
</table>

*Note that results refer only to production, not perception, with the exception of Zampini (1998).*
The English and Spanish phonemes /b, d, g/ also contrast in terms of VOT. However, the studies that investigate learners’ production of the voiced Spanish stops have paid much less attention to VOT and much more attention to the allophonic variation of /b, d, g/ with \([\beta, \delta, \gamma]\). In Spanish, after a pause or a nasal, as well as after /l/ in the case of /d/, the phonemes /b, d, g/ are realized as stops, but in all other contexts they are realized as the approximants \([\beta, \delta, \gamma]\), which are sometimes less precisely described as fricatives having undergone a process of spirantization (Hualde, 2005). Native speakers of English tend to produce stops for Spanish /b, d, g/ in all phonological environments and avoid pronunciation of \([\beta, \delta, \gamma]\), resulting in a noticeable foreign accent. Many learners fail to produce the bilabial \((\beta)\) and velar \((\gamma)\) approximants, which are not part of their English repertoire, and fail to assign \([\delta]\) allophonic status and produce it in the appropriate contexts in Spanish. In fact, second-semester college Spanish learners (n=17) have been reported to produce \([\beta, \delta, \gamma]\) in only 16%, 5%, and 12% of required contexts, respectively (Zampini, 1993). By the fourth semester of Spanish instruction, learners (n=15) made some improvement but still produced \([\beta, \delta, \gamma]\) infrequently (24%, 7%, and 18%, respectively) in the required contexts (Zampini, 1993).

The Spanish rhotics, /ɾ/ (an alveolar tap) and /r/ (an alveolar trill), are typically late acquired by English speakers. The difficulties English speakers face when acquiring the Spanish rhotics are multi-faceted. The American English rhotic is a voiced alveolar approximant and retroflex sound /ɹ/ quite unlike the Spanish /ɾ/ and /r/. Substitution of /ɹ/ for Spanish /ɾ/ or /r/ is

---

5 Hereafter referred to simply as \([\beta, \delta, \gamma]\).

6 Of notable exception to this generalization are some Central American dialects, in which occlusive sonorant forms of /b, d, g/ can be realized when in combination with /s, r, l/ or glides (Canfield, 1962; Zamora Munné & Guitart, 1982).

7 Both phones, but most commonly the trill, can be realized as a voiced apicoalveolar sibilant in many regions of Latin America (Resnick, 1975; Canfield, 1962). The assibilated variant is not considered standard and generally is not taught to L2 learners, nor is it referenced in the instructional modules that will be used here.
noticeably “foreign,” yet it is a pervasive feature of English speakers’ production of Spanish (e.g. Elliott, 1997; Face, 2006; Major, 1986). The trill does not exist in English and requires substantial articulatory force to produce, making it difficult for English speakers (Lord, 2005), which perhaps explains why learners tend to produce taps with greater accuracy than trills (e.g., Reeder, 1998; Waltmunson, 2006). Relatively novice students produce the trill accurately about 10% of the time (Reeder, 1998), advanced undergraduate students produce the trill accurately about 25% (Face, 2006) to 55% (Reeder, 1998) of the time, and even professionals with more than 25 years of Spanish-speaking experience do not always produce target-like trills (Face & Menke, 2010). Production of the Spanish tap, on the other hand, is generally less problematic for L2 learners (e.g., Reeder, 1998; Waltmunson, 2006), though differentiation between the tap and the trill is still problematic for learners at least through the eighth semester of college study (Rose, 2010). An alveolar tap does exist in English, and it is produced as an allophonic variant of /t/ and /d/ in post-tonic intervocalic positions in words such as letter and ladder. Thus English speakers’ difficulty with the tap is thought not to be articulatory but rather perceptual. Most investigations of tap and trill, however, have focused on learners’ production, leaving unanswered the question of exactly how L2 Spanish learners actually perceive these phones.

2.1.1 Empirical Studies of Instruction and Spanish L2 Pronunciation

Several researchers have set out to measure the effect that explicit instruction in Spanish phonetics and phonology can have on learners’ pronunciation of difficult-to-acquire Spanish phones. Table 2.2 summarizes those studies, and the relevant methodological characteristics and results of each study will now be presented.
Castino (1996) collected data from college students (n=40) in their third or fourth year of Spanish study and enrolled in a semester-long phonetics course. No control group was recruited. The target phones were [β, θ, ñ, r, r, x]. As part of the phonetics instruction, learners were taught the place and manner of articulation for each target phone. Learners participated in extensive pronunciation drills. They also recorded themselves, transcribed the recordings, and received feedback on the accuracy of their pronunciation. Two types of tests were used to measure the change in learners’ production before and after instruction: a dialogue to read aloud and a communicative task to elicit spontaneous speech. Tokens of the target phones were rated for accuracy on a scale of 0-2 points. The results of the analysis of variance indicated that learners were more accurate at producing each of the target phones in the posttest when compared to the pretest and that they were more accurate when reading from a script than when producing the phones spontaneously. The results must be interpreted cautiously because Castino (1996) did not use a control group, provided scant detail about the methods employed in the study, and failed altogether to address important considerations such as reliability of the rating.
<table>
<thead>
<tr>
<th>Author</th>
<th>Participants' Level</th>
<th>N Exp</th>
<th>N Control</th>
<th>Targets</th>
<th>Instructional Time</th>
<th>Post-instructional Activities and Practice (where described)</th>
<th>Control condition (where present)</th>
<th>Testing (Pre and Post)</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castino (1986)</td>
<td>6-8th semester, phonetics course</td>
<td>40</td>
<td>n/a</td>
<td>[β ɣ r x]</td>
<td>full semester course</td>
<td>&quot;Extensive pronunciation drilling of Spanish allophonic variants.&quot; Audio recordings of student productions and subsequent transcribing.</td>
<td>n/a</td>
<td>1. Dialogue reading. 2. Spontaneous speech during communicative task.</td>
<td>Significant improvement for all phones. Spontaneous speech was more accurate than dialogue reading.</td>
</tr>
<tr>
<td>Elliott (1995, 1997)</td>
<td>3rd semester</td>
<td>43</td>
<td>23</td>
<td>[s e i o u p t k b g b ɣ r w]</td>
<td>10 - 15 min. per class, for 21 class periods</td>
<td>Pronunciation practice in the form of repetition, jazz chants, rhymes, and tongue twisters. Immediate and consistent correction of pronunciation errors.</td>
<td>Regular section of same course taught by same instructor. Little to no correction of production errors.</td>
<td>1. Word repetition, 2. Sentence repetition, 3. Word reading, 4. Spontaneous speech in picture task.</td>
<td>Instruction was significant predictor of aggregate post-test scores on tasks 1-3 (not 4), and of post-test scores of liquids and stops (not fricatives, nasals, or vowels).</td>
</tr>
<tr>
<td>Gonzalez-Bueno (1997)</td>
<td>4th semester. OPI determined all were intermediate</td>
<td>30</td>
<td>30</td>
<td>/p t k b g/</td>
<td>5 - 10 min. per class, 3 times per week for semester</td>
<td>Perceptual discrimination. Transcription. Production practice in the form of sentence repetition and communicative activities such as expanded dialogues.</td>
<td>Regular section of same course taught by same instructor (the experimenter).</td>
<td>OPI</td>
<td>Significant reduction of VOT of /p/ and /g/ only.</td>
</tr>
<tr>
<td>Lord (2005)</td>
<td>phonetics course</td>
<td>17</td>
<td>n/a</td>
<td>[p t k β ɣ r], diphthongs</td>
<td>full semester course</td>
<td>Use of voice analysis software to analyze spectrograms of student productions.</td>
<td>n/a</td>
<td>Literary passage reading. Same passage was used as an in-class activity for self-analysis.</td>
<td>Significant improvement in producing [β ɣ r] and diphthongs. No significant reduction of VOT for /p t k/.</td>
</tr>
<tr>
<td>Lord (2010)</td>
<td>6-8th semester, study abroad</td>
<td>4</td>
<td>4</td>
<td>[b d g β ɣ]</td>
<td>8 weeks</td>
<td>Study abroad immersion following semester-long phonetics course.</td>
<td>In same study abroad program as experimental group, but no prior phonetics instruction.</td>
<td>Word list reading</td>
<td>Group with prior phonetics instruction improved 17% more than control (not tested for statistical significance).</td>
</tr>
</tbody>
</table>

* Instruction included explanations of Spanish/English contrasts, place and manner of articulation, and grapheme-phoneme correspondences.
Elliott (1995, 1997) provided two intact classes of students (n=43) enrolled in a third-semester college Spanish course with 10-15 minutes of instruction in phonetics and phonology during each of 21 class meetings over the course of a semester. Another intact class served as the control group (n=23). The experimental group’s instructional treatment addressed 20 Spanish allophones as well as the “silent” grapheme h. Treatment tasks included modeling the phones in words; asking students to describe the point, place, and manner of articulation of the phones, correcting them as necessary and providing diagrams of the vocal tract configuration when helpful; contrasting the phones with their English counterparts; and providing pronunciation practice such as sentence reading, jazz chants, rhymes, and tongue twisters. Students worked in small groups and provided feedback on the target-likeness of their classmates’ production. Elliott (1995, 1997) characterized the treatment as “phonological instruction.”

Learners completed a pretest during the second week of the course and an identical posttest in the final two weeks of the semester. For the tests, students read lists of Spanish words and sentences. The tests also included a picture description task to elicit spontaneous speech. Three raters judged students’ pronunciation accuracy. One rater was a native speaker of Spanish, and the other two were highly proficient non-natives, one of whom was the researcher. Each token of a target phone was rated either incorrect (1 point), an approximation of the correct phone (2 points) or correct (3 points). Additionally, the spontaneous speech samples were rated with an overall score of 1 (very strong accent) to 5 (native).

Elliott (1995) reported that the pretest scores of the experimental and control groups did not differ statistically (mean=174.07 and 172.43, respectively). The posttest
score of the experimental group (mean=186.56) was significantly higher than the control group (mean=170.09). In fact, treatment group accounted for 14% of the unique variance in posttest scores. Elliott administered a survey to measure students’ attitudes towards L2 pronunciation (The Pronunciation Attitude Inventory) and a standardized test to measure students’ cognitive style in terms of relative field dependence/independence. Neither attitude nor cognitive style was found to be a significant predictor of posttest gains, but attitude was moderately correlated with pretest and posttest pronunciation scores ($r = .31$, $p < .01$), suggesting that attitude was an individual difference that may have accounted for some variation in L2 phonological acquisition rate or ultimate attainment. Elliott concluded that the multimodal instruction provided in his study was beneficial to learners’ pronunciation and that future studies should investigate the effects of similar instructional techniques with more novice learners.

Elliott (1997) then analyzed the same data to determine the effect of instruction across the different production tasks and target phones. He concluded that instruction was most beneficial for discrete-word level tasks and beneficial for sentence repetition tasks, but not more beneficial than the control treatment for spontaneous production. As for phone classes, instruction proved beneficial for liquids ($F = 16.76, p = .0001$) and stops ($F = 7.93, p = .006$) but not for fricatives ($F = 2.09, p = .15$), nasals ($F = 3.40, p = .06$), or vowels ($F = 3.81, p = .055$). Of all the phones, the trill (/r/) experienced the most improvement.

González-Bueno (1997) investigated the effect of instruction on learners’ pronunciation of Spanish stops. The students ($n=60$) were enrolled in fourth-semester Spanish conversation classes. Their Oral Proficiency Interview (OPI) ratings ranged
from novice high to intermediate high. The experimental group (n=30) received 5-10 minutes of formal instruction on the target phones during most class meetings. The target phones were syllable-initial /p, t, k, b, d, g/. The instructional treatment included a theoretical and practical introduction to articulation; contrasts of the phones with their English counterparts; analysis of spectrograms; pronunciation practice, including word lists, sentences, and expanded dialogues; and discrimination activities, in which students were presented with minimal pairs (e.g., [paca]/[baca]) and wrote down what they perceived. Students’ pronunciation accuracy was measured in a pretest and posttest, both modified OPIs that elicited 20 minutes of spontaneous speech. The OPIs were then transcribed and analyzed by the researcher. Six tokens of each target phone were sampled randomly from the pre-treatment and post-treatment OPI of each student. The tokens were in stressed syllable-initial positions and were followed by one of the vowels /a, o, u/. The voice onset timing (VOT) of the tokens were measured and averaged for each target phone, and the data were then submitted to an analysis of variance with the pretest values entered as a covariate (ANCOVA). A reduction in VOT would constitute an improvement in pronunciation for these target phones, since the VOT of the English stops is longer than that of their Spanish counterparts.

González-Bueno found that while the control group’s production of the target phones did not vary from pretest to posttest, the experimental group’s production of two target phones, /p/ and /g/, did improve significantly (mean VOT reduction of 21.9ms and 22.17ms, respectively) from pretest to posttest. However, the experimental group did not make significant improvement on their production of the other four phones. It should be noted also that students’ production was not rated in a naturalistic sense. González-
Bueno did not report whether this reduction in VOT of /p/ and /g/ would have been detected by the human ear and thus recognized as a change toward more target-like pronunciation of Spanish stops.

Lord (2005) investigated the effects that instruction given in an advanced Spanish phonetics course had on students’ (n=17) pronunciation of a variety of segmental and suprasegmental features. Students received instructional treatments similar to those in Elliott (1995) and González -Bueno (1997). The target phones in Lord (2005) were syllable-initial /p, t, k/, the approximants [β, ð, γ], the vibrant trill /r/, and one suprasegmental feature: diphthongs within and between words. Students completed a pretest on the second day of class and an identical posttest at the end of the semester. The tests required students to read aloud a literary passage. The researcher analyzed several tokens of each target phone, using voice analysis software to measure the VOT of /p, t, k/ and coding tokens of the other phones as either “correct pronunciation” or “incorrect.”

Lord (2005) reported that posttest scores for nearly all the target phones were significantly higher than pretest scores. Accuracy of pronunciation of [β] increased from 44% to 66%, accuracy of [ð] increased from 34% to 48%, accuracy of [γ] increased from 28% to 53%, accuracy of /r/ increased from 26% to 39%, and accuracy of diphthongization across words increased from 60% to 86%. While students did not reduce VOT of /p, t, k/ significantly from pretest to posttest, their pretest VOTs were significantly longer than native speaker controls, whereas their posttest VOTs were not. Lord (2005) claimed that students therefore showed demonstrable improvement on VOT reduction as well. However, Lord (2005) did not include a control group in this study,
and so it would be premature to conclude that the improvements made in pronunciation were the direct result of receiving instruction in Spanish phonetics and phonology.

Lord (2010) investigated the combined effects of immersion and instruction on L2 pronunciation. Learners (n=8) were enrolled in a 2-month summer program in Mexico. Half of the learners had previously taken a Spanish phonetics course, and half had not. Learners were recorded reading a list of 60 nonce words that included 10 tokens of each of 6 target phones: /b, d, g/ and [β, ð, ɣ]. Signalize software was used to detect occlusions in the recorded data, i.e., whether learners produced stops in contexts that required approximants. Participants who had had prior instruction in phonetics produced the target phones with greater accuracy on the pretest. Both groups improved their pronunciation accuracy during the course of the 2-month immersion experience. The improvement was greater (20.1% gain) for the group with prior phonetics instruction compared to the group without (2.8% gain). The data suggested that the combined effect of instruction and extensive exposure to native-sounding speech was greater than the effect of either alone. However, these differences could not be tested for statistical significance due to small numbers of participants enrolled in the study.

The studies summarized in Table 2.2 and described in detail above all reported positive effects of instruction, at least for some target phones in some testing conditions. However, their experimental designs may have predisposed them to finding positive effects of instruction, or suffering from Type 1 error. Castino (1996) and Lord (2005) did not include a control group of any kind, so while learners’ pronunciation improved from the beginning to the end of their semester-long phonetics course, it cannot be concluded that the improvement was due only to the instruction they received. Elliott (1995, 1997)
and González-Bueno (1997) compared their instructed learners with an intact class of learners at the same curricular level who did not receive pronunciation instruction. The experimental and control groups, then, differed not only in terms of the presence or absence of explicit instruction about Spanish phonetics and phonology but also in terms of the types of exposure, practice, and feedback they received regarding the target phones. For example, for the experimental group in Elliott’s (1995) study, “errors in pronunciation were corrected immediately and consistently” whereas “pronunciation was corrected for subjects in the control section only when the utterance was unintelligible and impeded communication. As a result, error correction rarely occurred […] The major goal for subjects in the control section was to communicate meaning and not to be overly concerned with correct pronunciation (p. 535). Elliott (1995) viewed correction of pronunciation errors as an essential component of pronunciation instruction, which is the traditional view. Yet the experimental design of studies such as Elliott (1995, 1997) and González-Bueno (1997) does not allow the effect of the explicit phonetics instruction itself to be assessed as separate from other factors that potentially contribute equally or even more toward improvement in pronunciation, such as learners’ attention being drawn repeatedly to the target phones.

2.1.2 The Present Study

In sum, prior research suggests that all these target phones, \([p, t, k, \beta, \delta, \gamma, \acute{r}, \grave{r}]\), play a role in English speakers’ accent in Spanish and are difficult to acquire. Research also suggests that some of these phones are amenable to instruction. The focus of the present study was different from the studies in Table 2.2, which attempted to evaluate
traditional pronunciation instruction as a whole. The present study evaluated the effectiveness of one foundational component of instruction, which was the explicit teaching about Spanish phonetics and phonology. A group that received lessons on Spanish phonetics and phonology was compared to a control group that did not. Exposure and attention to the target phones, pronunciation practice, and feedback were held constant so that the main difference between groups was explicit knowledge of Spanish phonetics and phonology. In other words, the present study attempted to compare pronunciation instruction with another viable methodological choice that had the potential to be equally advantageous for improving learners’ pronunciation.

Other researchers have examined various elements and methodologies of pronunciation instruction, with mixed results. Chung (2008) compared explicit, implicit, and noticing instruction for improving Chinese EFL learners’ production of English word stress and found that all groups improved equally on the posttest, but the explicit group was significantly better in the delayed posttest. Macdonald, Yule and Powers (1994) found no significance difference in the pronunciation changes of Chinese EFL learners exposed to traditional drilling activities, self-study with tape recordings, or interactive activities, although all methods were superior to a no-intervention control condition. Moyer (1999) reported that feedback at the segmental and suprasegmental levels was a significant predictor of accent for English speaking German learners, and de Bot (1983) reported that Dutch EFL learners benefited from visual feedback (seeing pictorial representations of pitch contour) in learning English intonation whereas auditory feedback alone (hearing themselves) was detrimental, yet Ducate and Lomicka (2009) found no benefit of repetitive practice and feedback in their 16-week podcasting
experiment with German and French learners. Thus there is still much work to be done unpacking the many elements of pronunciation instruction so as to determine which are the most effective, whether in isolation or in combination.

It seems particularly relevant to question the effectiveness of explicit instruction in phonetics and phonology considering that it is precisely this explicit element of pronunciation instruction that may make teaching pronunciation unattractive to the FL community. Researchers who advocate for pronunciation instruction lament that textbook authors, instructors, and administrators are reticent to include pronunciation instruction in the FL curriculum because they view it as overly form-focused and in opposition to their communicative, meaning-focused methodology (e.g., Arteaga, 2000; Morin, 2007). It has been suggested (Isaacs, 2009) that pronunciation instruction should be better integrated within communicative activities. Alternatives for bringing learners’ attention to the L2 sound system, through targeted exposure, focused listening, dictation, transcription, or other activities need to be seriously considered and empirically tested if in fact explicit phonetics instruction is not a viable methodological choice for some learning contexts.

Another ancillary but highly relevant question is when to provide pronunciation instruction, if at all. There may be an optimal stage in the L2 acquisition process or an optimal moment in the L2/FL curriculum for such instruction. In their report of an Interagency Language Roundtable investigation, Higgs and Clifford (1982) found a U-shaped curve in terms of the relative contribution that pronunciation makes towards assessments of global language proficiency. Pronunciation is most important in the early stages of L2 acquisition and in the more advanced stages, though its importance wanes in
the intermediate stage. If instruction leads to short-term improvements in pronunciation, the Higgs and Clifford (1982) study might suggest that instruction is best provided early in the curriculum and then “recycled” in advanced courses. There are proponents of providing instruction at the earliest stages possible (e.g., Arteaga, 2000). Yet researchers investigating pronunciation instruction have more often recruited participants at the intermediate to advanced levels (e.g., refer to Table 2.2) and have encouraged others to replicate their studies with more novice learners (Elliott, 1995; Lord, 2005). These studies have recruited participants from only one matriculation or proficiency level and have therefore been unable to address the issue of optimal instructional sequencing. The FL curricula that do include phonetics instruction typically offer a phonetics course at the advanced course level, for language majors and minors, but there are no empirical data available to indicate whether or not this sequencing is optimal.

2.1.3 Research Questions and Hypotheses

The present study attempted to contribute to the existing research by evaluating the explicit teaching of L2 phonetics and phonology as separate from the other putatively beneficial aspects of pronunciation instruction and by comparing the effects of instruction across curricular levels. The research questions relating to pronunciation accuracy were: 1. Does instruction in L2 phonetics and phonology improve learners’ ability to produce L2 phones?, and 2. Does the effectiveness of instruction depend on learners’ L2 curricular level? Based on the generally positive effects found in the literature, it was hypothesized that instruction would prove beneficial for learners’ production of the target phones. Though prior studies did not directly compare learners of different curricular
levels, developmental readiness has figured prominently in theories of L2 phonological acquisition, so it was hypothesized that the effectiveness of instruction would vary by target phone and would interact with learners’ curricular level in possibly complex ways.

2.2 Methods

2.2.1 Context

Participants (n=95) recruited for this study were enrolled in Spanish as a foreign language courses at a mid-sized, public university in the southeastern United States. The Spanish FL curriculum at this university was made up of introductory language courses, conversation courses, grammar and composition courses, and various upper-division literature, film, media, civilization, translation, and interpretation courses. There was no course dedicated to linguistics in general or phonetics/phonology in particular. Spanish courses at this university lasted one semester (14 weeks) and required either 4 weekly contact hours (first year sequence) or 3 weekly contact hours (all others).

Seven intact classes participated: three introductory, two intermediate, and two advanced. Five instructors taught these classes (two instructors taught multiple sections). The instructors were born and raised in the United States, Colombia, Peru, and Puerto Rico. Most were immersed in a Spanish-dominant environment at least through the age of puberty. The one instructor born in the United States spoke English, Spanish, and French at home and in school. Three instructors were female and two were male. All
instructors were college educated and had formal education in Spanish. All had lived in the United States and taught Spanish for 8 or more years.

2.2.2 Participants

A total of 124 Spanish learners were recruited. Of those, 14 missed multiple sessions or withdrew from the class, and 15 were eliminated from the analysis for not meeting background criteria. Participants were included in the analysis if they were at least 18 years of age, had not begun learning Spanish before the age of 10, and had not received instruction in Spanish phonetics and phonology before the study. As a result, 95 total participants were included in the analysis, 58 female and 37 male. Of those, 86 completed all four sessions. Participants’ mean age was 22.06 (range 18-44, mode = 19). The mean age at which they began learning Spanish was 15.66 (range 11 – 40, mode = 13).

Participants were enrolled in one of three courses: introductory level 1, intermediate conversation, or advanced conversation. The curriculum allows students flexibility in course sequencing, but most students enrolled in these courses were in their first year, second year, or third year of college Spanish study, respectively. Therefore, to avoid the term “advanced,” which refers here to a curricular sequence and not linguistic or communicative competence, participants in this study will henceforth be referred to as first year, second year, and third year learners. According to their reports on the background questionnaire, the first year students had completed on average 2.26 (year long) high school courses and 0 (semester-long) college courses in Spanish. The second year students had completed on average 2.97 high school courses and 2.75 college
courses (range: 1-6). The third year students had completed on average 3.47 high school
courses and 3.68 college courses (range: 2-7). Participants were told that the study was
designed to test the effectiveness of instruction on their listening and speaking skills.
They completed the study during class time and were not given extra credit or offered
compensation for participating in the study.

In addition to the Spanish learners, native speakers (NSs) of Spanish (n=10) were
recruited to provide baseline data. NS participants attended one session in which they
completed a background questionnaire and one version of the experimental tests. The
NSs were natives of Chile, Colombia, Mexico, and Spain, and they were in the United
States to attend graduate school or to work as professionals. Half were males and half
were females. Their mean age was 30 (range 24 – 37). Their mean time living in the US
was 4.42 years (range 2 months – 10 years). Their mean age when first exposed to
English was 8.4 years (range 5 – 15). All the NS were first exposed to English through
foreign language classes. All judged their current English proficiency to be advanced or
very advanced. None had ever taken a class in phonetics and phonology. These NS
participants were all competent Spanish/English bilinguals, and it was expected that their
pronunciation of the target phones might differ from monolingual Spanish speakers,
particularly in terms of VOT (see Flege, 1981; Williams, 1979). However, their speech
exhibited no foreign or (compared to other speakers of their native dialect) nonstandard
accent. These NSs’ pronunciation of the target phones was thus considered an
appropriate baseline with which to compare learners of Spanish.
All participants were recruited and treated under the research protocol approved by the offices of Human Subjects Protection at Georgetown University (IRB # 2010-425) and Virginia Commonwealth University (IRB # HM13058).

2.2.3 Experimental Design

The study was a pretest, posttest, delayed posttest design. Participants were randomly assigned to one of two instructional conditions: experimental or control. Intact classes came to the language lab during their normal class time. As students entered the lab on the first day, they were directed to sit at the next available computer station. The computer stations were set up in an alternating pattern to deliver either the experimental or the control treatment. Students were told that their assigned activities might look different from their neighbor’s activities on any given day, but they were not told whether they were in the experimental or control condition. Students sat at the same computer station during all the sessions of the study. All aspects of the study were undertaken during normal class time. The class instructors were present during most of the sessions, but they did not participate in or lead the instruction in any way.

The experimental group (+PI) received phonetics instruction via four online modules that explicitly taught aspects of Spanish phonetics and phonology. The control group (-PI) watched video vignettes of Spanish native speakers and took dictation but did not receive phonetics instruction. The experimental design is illustrated in Figure 2.1. All sessions were completed within one semester of instruction. The class meeting times overlapped, and so not all classes could complete the study concurrently. The third year
learners began the study in the first week of September, the second year learners began
the study 3 weeks later, and the first year students completed the study during the last 6
weeks of the semester. In between the experimental treatment sessions, classes within
each matriculation level followed similar syllabi. The syllabi and textbooks did not
contain any lessons focused on phonetics or pronunciation. All the instructors reported
that they typically listened for learners’ pronunciation errors and corrected them on
occasion during normal class meetings, but that they emphasized communication and
fluency more than pronunciation accuracy. Instructors agreed not to provide any
additional phonetics or pronunciation instruction during the study. Instructors were not
told which Spanish phones were targeted in the study.

2.2.4 Phonetics Instruction (+PI)

The phonetics instruction group (+PI) completed 4 computer-based, interactive
modules focused on: 1) an introduction to articulatory phonetics, 2) the occlusive
consonants /p, t, k/, 3) the occlusive consonants /b, d, g/ and their approximant variations
[β, δ, γ], and 4) the liquid consonants /ɾ, r/. All learners completed the introduction to articulatory phonetics first, but the order of the other three modules was counterbalanced. The modules were created by Dr. Gillian Lord at the University of Florida and are free and open to the public (http://grove.ufl.edu/~glord/). The modules presented the following information and activities, all of which are typical of instruction in phonetics and phonology and have empirical evidence supporting their use (Arteaga, 2000; Elliott, 2003, p. 38): an explanation of grapheme-phoneme correspondences; an explanation of the point, place, and manner of articulation, with an animated diagram of the vocal tract illustrating how each sound is produced; an explanation of differences in the articulation of analogous Spanish/English sounds and the phonological environments in which the sounds are produced in each language; and identification activities which required learners to identify Spanish and English sounds or identify the manner of articulation (e.g., occlusive [edo] versus approximant [eðo]). After each of these sections of the module, there was a brief multiple-choice comprehension check. Learners received feedback about which items they answered incorrectly, and they had to re-take the assessment and answer all items correctly before they could proceed to the next section. Finally, each module contained a pronunciation practice activity that directed learners to listen to and repeat after a native speaker producing Spanish words and phrases that incorporated the target sounds until learners felt that their pronunciation approximated the native speaker. Learners received no additional feedback on their pronunciation other than their own evaluation of their production as compared with the native speaker. Appendices 2.1 – 2.9 contain screen shots displaying all the information and activities contained in one example module.
Learners spent between 15 and 40 minutes on each module. A time limit of 25 minutes per module was suggested, but the modules were designed to be self-paced, and so there was some variation among learners. The instructional time per phone was thus brief but on par with the amount of time devoted to each L2 phone in other non-specialized FL classes in similar studies (e.g., Elliott, 1995; González-Bueno, 1997). On average, the instruction exposed learners minimally to 10 unique tokens of each target phone, 3 of which were contained in the pronunciation practice section.

2.2.5 Control Instruction (-PI)

Learners in the control group (-PI) completed self-paced, computer-based, interactive online modules that exposed them to the target phones in amounts roughly equivalent to the +PI and gave them practice, but they received no explicit instruction in phonetics and phonology. The -PI modules consisted of video vignettes featuring native speakers of Spanish talking about a variety of topics. The vignettes are free and open to the public (http://lait.s.utexas.edu/spe/). Learners completed a dictation as they watched the videos. They were instructed to compare their dictation with the official transcript, read the English translation for meaning, evaluate their dictation, comment on the speaker’s accent, and also repeat a particular sentence in the video until their pronunciation was like the speaker’s. The videos chosen for the study were appropriate for each course level and were related to the current topics of study in the learners’ courses. The videos featured speakers from the capital cities of Spain, Colombia, Peru, Ecuador, Costa Rica, and Guatemala, none of whom produced the target phones in non-
standard ways. The -PI instructions are reproduced in full in Appendix 2.10. Appendices 2.11-2.13 provide full transcriptions of the video vignettes.

On average, the instruction exposed learners minimally to 10 unique tokens of each target phone, 3 of which were contained in the pronunciation practice section. On average the control group was exposed to the same amount of unique tokens of the target phones. Table 2.3 presents a side-by-side comparison of the number of tokens included in the +PI and the -PI modules⁸. However, learners in both groups could click again and listen to each token as often as they liked, and so learners may have actually heard and/or pronounced more tokens during instruction.

<table>
<thead>
<tr>
<th>Exposure (Listening)</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
<th>[β]</th>
<th>[ð]</th>
<th>[ɣ]</th>
<th>/ɾ/</th>
<th>/r/</th>
<th>Average of 8 phones</th>
</tr>
</thead>
<tbody>
<tr>
<td>+PI</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>13</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>1st Year -PI</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>24</td>
<td>11</td>
<td>31</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>2nd Year -PI</td>
<td>14</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>4</td>
<td>12</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3rd Year -PI</td>
<td>15</td>
<td>6</td>
<td>33</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>17</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Average of -PI</td>
<td>11</td>
<td>6</td>
<td>16</td>
<td>7</td>
<td>16</td>
<td>6</td>
<td>20</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pronunciation Practice</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
<th>[β]</th>
<th>[ð]</th>
<th>[ɣ]</th>
<th>/ɾ/</th>
<th>/r/</th>
<th>Average of 8 phones</th>
</tr>
</thead>
<tbody>
<tr>
<td>+PI</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>1st Year -PI</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>2nd Year -PI</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3rd Year -PI</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Average of -PI</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

⁸ The -PI featured natural, authentic language. The vignettes were varied across matriculation levels so as to be level appropriate in complexity and content. Thus the –PI modules could not be so strictly controlled as to contain the target phones in exactly the same amounts as the +PI.
It was thought that the dictation exercises could be fairly compared to the phonetics instruction in that both types of instructional modules presented the target phones in roughly equal amounts, required that learners focus their attention on the sounds (form), and provided pronunciation practice with identical feedback conditions. Time on task was also equivalent across +PI and –PI groups. The main difference was that the dictation exercises of the -PI provided no explicit instruction in phonetics and phonology. However, it must be noted that learners in the -PI were never told which phones were the linguistic targets of the study, which may have constituted an important difference between the conditions. Also, the pronunciation practice was different in that the -PI practiced with extended sentences and the +PI practiced with short phrases.

2.2.6 Production Test

The perception tests will be described in Chapter 3. The current chapter is concerned only with production.

2.2.6.1 Production Test Materials

The production test consisted of a 28-item list of words and phrases that participants read aloud. The tests items are presented in Appendix 2.14, with graphemes representing the target phones bolded. The target phones were not bolded in the list presented to students.

Researchers investigating L2 production typically use reading tasks (e.g., Lord, 2005), prompts for spontaneous speech (e.g., González-Bueno, 1997), or a combination
of both (e.g., Elliott, 1995, 1997). A word list was chosen so that beginning learners would not be cognitively overburdened by the task demands and rather would be able to focus mainly on pronunciation during the test. Of the 28 items, 20 were very frequent words selected from the active vocabulary lists in the first two chapters of the textbook used in the introductory course so that all learners would be equally familiar with them. Additionally, each target phone was included in one infrequent Spanish word (e.g., calaba [was soaking]) that paired the target phone with the vowel [a] in order to include some words with which all learners would be equally unfamiliar. In total then, the production tests included 4 tokens of each of the target phones in a variety of phonological environments. The items were presented in different orders on the pretest, immediate posttest, and delayed posttest. During the pretest and delayed posttest, participants were asked to translate the items so that their knowledge of these words could be assessed. Indeed, most participants translated the “familiar” correctly on the pretest, and no participants translated the “unfamiliar” words correctly on the posttest.\footnote{Participants’ familiarity with the production test items was assessed by calculating learners’ accuracy in a random sampling of 25% of the data. First year learners’ translations on the pretest were analyzed to indicate how many items they knew prior to beginning the study. On the pretest, first year learners translated the “familiar” items with 82% accuracy, as compared with 0% accuracy for the “unfamiliar” items. Third year learners’ translations on the delayed posttest were analyzed to indicate how many items they knew by completion of the study. On the delayed posttest, third year learners translated the “familiar” items with 95% accuracy, as compared with 0% accuracy for the “unfamiliar” items.}

2.2.6.2 Production Test Administration Procedures

Learners were seated at individual PC computer stations in the language lab as they completed all instructional modules, tests, and questionnaires. They wore noise-
canceling headphones with attached microphones. They recorded their production tests as WAV files, sampling rate of 22 KHz and sampling size of 16 bit, with the software package Sanako. NS participants completed the study components at a quiet location convenient to them (e.g., their home or a local library) and were recorded using Praat software at a sampling rate of 44 KHz and 16 bit. The instructions for the production test are given in Appendix 2.15.

2.2.6.3 Production Test Rating and Scoring Procedures

An independent rater transcribed the production data and assigned scores to the approximant and rhotic phones. The rater was a native speaker of Spanish pursuing a PhD in Spanish linguistics with an emphasis on phonetics and phonology. He was unaware of the objectives and hypotheses of the research study. The sound files were coded such that the rater was unaware of learners’ matriculation level, learners’ instructional condition, and time of test. Analyzing the approximately 400 sound files in a random order, the rater transcribed the files into IPA and rated the tokens of target phones. The rater used auditory cues as well as waveforms and wide-band spectrograms (window of .005s) in Praat. In cases where the acoustic evidence in the spectrogram and waveform was not clear, e.g., due to clipping, the rater based his decision on auditory properties alone.

The approximant and rhotic target phones were assigned 1-3 points for every token that could be heard clearly. Only a few productions were not assigned points because the participant misread the word and therefore did not include the target phone or
the sound quality was poor due to background noise, participant whispering, etc. Productions were assigned 3 points if they demonstrated all the auditory and acoustic properties that are associated with their Spanish pronunciation, 1 point if they demonstrated all the auditory and acoustic properties that are associated with an English-accented pronunciation, and 2 points if they demonstrated a combination of the auditory and/or acoustic properties of both languages. In order to construct this rating scale, the researcher and the independent rater discussed these properties at length while analyzing a representative sample of Spanish learners’ and NSs’ recordings for each test item. Table 2.4 lists the relevant properties analyzed for each phone. Figures 2.2 – 2.16 display spectrograms illustrating the acoustic evidence supporting a rating of 3 points, 2 points, and 1 point for each target approximant and rhotic phone.

Different ratings were used for the rare occasions when learners produced phones that were so unclear as to be not ratable (coded as “?”) or substituted an entirely different phone for the target phone (coded as “@”).

---

10
<table>
<thead>
<tr>
<th></th>
<th>English-like (1 point)</th>
<th>(2 points)</th>
<th>Spanish-like (3 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[β]</td>
<td>auditory</td>
<td>heard as stop</td>
<td>e.g., no formants</td>
</tr>
<tr>
<td></td>
<td>acoustic</td>
<td>full occlusion, amplitude drops precipitously, release burst possibly apparent</td>
<td>formant structures and voice bar evident; no evidence of occlusion, burst, or frication; little to no decrease in amplitude from surrounding vowels</td>
</tr>
<tr>
<td>[δ]</td>
<td>auditory</td>
<td>heard as stop (e.g., [d]) or as [ɾ]</td>
<td>e.g., no formants</td>
</tr>
<tr>
<td></td>
<td>acoustic</td>
<td>full occlusion, amplitude drops precipitously, release burst possibly apparent</td>
<td>formant structures and voice bar evident; no evidence of occlusion, burst, or frication; slight to moderate decrease in amplitude from surrounding vowels</td>
</tr>
<tr>
<td>[ɣ]</td>
<td>auditory</td>
<td>heard as stop</td>
<td>e.g., no formants</td>
</tr>
<tr>
<td></td>
<td>acoustic</td>
<td>full occlusion, amplitude drops precipitously, release burst possibly apparent</td>
<td>formant structures and voice bar evident; no evidence of occlusion, burst, or frication; moderate to significant decrease in amplitude from surrounding vowels</td>
</tr>
<tr>
<td>/s/</td>
<td>auditory</td>
<td>heard as [s]</td>
<td>e.g., closure too long (e.g., [d])</td>
</tr>
<tr>
<td></td>
<td>acoustic</td>
<td>no occlusion, formant structures evident, possibly with low F3</td>
<td>brief lightened band, no formants evident (except for para and número, in which [ɾ] may be reduced)</td>
</tr>
<tr>
<td>/r/</td>
<td>auditory</td>
<td>heard as [r]</td>
<td>e.g., heard as [ɾ] or another target-like dialectal variant of [ɾ]</td>
</tr>
<tr>
<td></td>
<td>acoustic</td>
<td>no occlusion, formant structures evident, possibly with low F3</td>
<td>e.g., low F3</td>
</tr>
</tbody>
</table>
Figure 2.2 Spectrogram of Sample [β] Token Assigned 3 Points

Figure 2.3 Spectrogram of Sample [β] Token Assigned 2 Points

Figure 2.4 Spectrogram of Sample [β] Token Assigned 1 Point
Figure 2.5 Spectrogram of Sample [ð] Token Assigned 3 Points

Figure 2.6 Spectrogram of Sample [ð] Token Assigned 2 Points

Figure 2.7 Spectrogram of Sample [ð] Token Assigned 1 Point
Figure 2.8 Spectrogram of Sample [ɣ] Token Assigned 3 Points

Figure 2.9 Spectrogram of Sample [ɣ] Token Assigned 2 Points

Figure 2.10 Spectrogram of Sample [ɣ] Token Assigned 1 Point
Figure 2.11 Spectrogram of Sample /ɾ/ Token Assigned 3 Points

Figure 2.12 Spectrogram of Sample /ɾ/ Token Assigned 2 Points

Figure 2.13 Spectrogram of Sample /ɾ/ Token Assigned 1 Point
Figure 2.14 Spectrogram of Sample /r/ Token Assigned 3 Points

Figure 2.15 Spectrogram of Sample /r/ Token Assigned 2 Points

Figure 2.16 Spectrogram of Sample /r/ Token Assigned 1 Point
The stop consonant target phones were not rated. Rather, the researcher measured the VOT of /p, t, k/ using evidence from waveforms and wide band spectrograms displayed in Praat. The VOT was measured from the release of the stop closure to the first glottal pulse, indicating the beginning of voicing. The VOT measure was considered objective and reliable, so a second rater was not used. Due to the nature of the word reading task, intra-speaker and inter-speaker differences in speech rate were considered immaterial and were not controlled for. VOT data will be presented in the results section separately from approximant and rhotic rating data.

In order to make comparisons across target phones and to calculate an 8-phone aggregate score, the VOT data were also transformed into 1-3 points rating. This transformation was performed on each test item separately, utilizing the VOTs produced by NS to determine the categorical ratings as follows: 3 points were assigned to VOT values that fell within the NS range, i.e., no longer than the longest VOT produced by any NS on that test item; 2 points were assigned to VOT values that were no longer than the NS maximum plus the value of the NS range; and 1 point was assigned to all others. For example, for the test item para, the VOT values produced by NS had a minimum of 5ms, a maximum of 21ms, and thus a range of 16ms. For para, then, learner-produced VOTs of 0 – 21ms were assigned 3 points, VOTs of 22 – 37 ms were assigned 2 points (because 21+16=37), and VOTs longer than 38 ms were assigned 1 point.

Scores on the pretest, immediate posttest, and delayed posttest were calculated by adding the points received by each learner for the 4 items relevant to each target phone on each test. Finally, overall scores were calculated for each test administration by adding the scores on all target phones. In the case of the immediate posttest, this overall score
encompassed scores from sessions 2 and 3, since learners completed half of the instructional modules, followed by production posttests, during sessions 2 and 3.

2.2.7 Questionnaires

A background questionnaire asked Spanish learners to report their language learning experiences as well as basic demographic information. The background questionnaire is presented in Appendix 2.16. The background questionnaire given to native speakers of Spanish is presented in Appendix 2.17. Post-instructional module questionnaires asked learners to evaluate the interest, difficulty, helpfulness, and usefulness of each module on Likert scales of 1 – 10 and provide additional comments. The post-instructional module questionnaire is presented in Appendix 2.18. The post-study questionnaire asked learners to evaluate the instruction they received overall, with the same categories, as well as evaluate their performance and improvement on the discrimination and identification tests, which will be discussed below. The post-study questionnaire is presented in Appendix 2.19.

2.3 Results

2.3.1 Comparison of +PI and –PI Groups’ Demographic and Language Background

The background questionnaire collected information about learners’ demographic and language backgrounds. Learners reported their sex, current age, and age at the onset of Spanish learning. They reported the number of Spanish courses taken in high school,
number of Spanish courses taken in college (organized by type: general language, conversation, composition, civilization, and literature courses), number of native speakers who taught their Spanish courses in high school and college,\textsuperscript{11} time spent outside of class using Spanish, and time spent using Spanish abroad. They also reported on languages other than Spanish: those they had studied formally, leaned to a high proficiency, and been exposed to informally. The +PI and -PI groups were compared on each of these demographic and language background variables, at each matriculation level (1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} year) using independent samples T-tests. No significant differences between the +PI and -PI groups were found (all $t < 2.27$, $p > .05$), with just one exception. The third year +PI learners spent more time abroad (mean 91.85 hours, SD 101) than the third year –PI learners (mean 12.26 hours, SD 22.04) ($t(17) = 2.30, p = .03$, CI 6.64 – 153). Time spent abroad was calculated by multiplying the number of weeks spent abroad by the approximate number of hours spent each week using Spanish while abroad\textsuperscript{12}. However, a closer look at the data showed that the group difference was related to just three individuals in the +PI who had immersion experiences of 2 weeks, 1 month, and 2 months, respectively. It was decided that these participants would be kept in the analysis.

\textsuperscript{11} This measure was collected as an approximation of non-accented exposure to Spanish. Flege’s “accented L2 input hypothesis” claims that learners will not be able to perceive and produce L2 sounds accurately if they have received accented input in the L2 (1991).

\textsuperscript{12} Learners tended to use descriptive terms to describe their use of Spanish rather than hours. When learners reported using “minimal” Spanish, “very little” Spanish, or “basic” Spanish while abroad, it was estimated that they used Spanish for 3 hours each week. Likewise, “some” or “moderate” usage was estimated as 7 hours per week; “lots”, “quite a bit” or “frequent” usage was estimated as 14 hours per week; and “constant,” “in classes,” “with host family,” or “immersion” usage was estimated as 28 hours per week.
2.3.2 Inter-rater Reliability

Only one rater measured the VOT of the stop consonants /p, t, k/ because the VOT measurement was considered relatively objective and reliable. For the approximants and rhotics, the researcher randomly selected 10% of the data (770 target phone productions) to re-rate. There was inter-rater agreement on ratings for 95% of those data (Cronbach’s alpha of .96), so the rating was deemed reliable.

2.3.3 Descriptive Statistics

The average VOTs produced for the target phones /p, t, k/ by learners and NSs are presented in Table 2.5. Note that NSs in this task produced longer VOTs than have been previously reported (e.g., Lisker & Abramson, 1964; Poch, 1984). Several factors may have coincided to produce the relatively longer VOTs. Task effects may have been at issue, as the task involved word reading rather than continuous speech and all phone items were word-initial (Torreblanca, 1988). Also, these bilingual speakers’ VOTs in Spanish may be longer due to influence from English phonology (Flege, 1981; Williams, 1980). Table 2.6 presents the average ratings assigned to learners’ and NSs’ productions of the approximant and rhotic target phones. Note that NSs’ average ratings range from 2.43 - 3.00. Recall that tokens were assigned 3 points only if they fit all the auditory and acoustic criteria of an “idealized” realization of the target phone. The NS participants were bilingual speakers with dialectal differences, and though nothing about their speech sounded “foreign,” to the experimenter’s ear, some tokens of the target phones they uttered were not acoustically “ideal.” No NS received fewer than 2 points on any token, however.
Even though the NSs had longer VOTs and lower average ratings than one might expect, learners’ pronunciation was found to be significantly different (less target-like) than the NSs using independent samples T-tests. This was true of every phone and for learners at all three matriculation levels, before and after instructional intervention. Interestingly, the T-tests suggest that the pattern of development across matriculation levels differed across phones. VOT shortened with increasing Spanish proficiency from year 1 to year 2 but leveled off at the intermediate level (no significant difference from year 2 to year 3). The same pattern was found for [ð] and [ɹ]. The phones ([β, γ, r]), in contrast, did not improve across learners’ matriculation levels at all, from year 1 to 2 or from year 2 to 3. Some approximants and rhotics may have been more slowly acquired than stop consonants, as was suggested in previous empirical studies (e.g., Diaz-Campos, 2004; Reeder, 1998; Zampini, 1993). Alternatively, the difference in measurements between the stops (continuous scale, measured in milliseconds) and other phones (rating of 1-3 points) may explain why development across learners’ matriculation levels was apparent with stops but not with other phones. A finer-grained measurement scale for the approximants and rhotics may have been necessary to capture minor changes in learners’ production of those phones.
Table 2.5 Average Voice Onset Time in Production of Stop Consonants (milliseconds)

<table>
<thead>
<tr>
<th></th>
<th>First Year</th>
<th></th>
<th>Second Year</th>
<th></th>
<th>Third Year</th>
<th></th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+PI (n=22)</td>
<td>-PI (n=23)</td>
<td>+PI (n=18)</td>
<td>-PI (n=14)</td>
<td>+PI (n=10)</td>
<td>-PI (n=9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>Pre</td>
<td>64 (25)</td>
<td>58 (21)</td>
<td>42 (21)</td>
<td>38 (29)</td>
<td>37 (17)</td>
<td>35 (12)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>54 (20)</td>
<td>52 (20)</td>
<td>36 (18)</td>
<td>38 (21)</td>
<td>37 (23)</td>
<td>27 (10)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>58 (24)</td>
<td>58 (23)</td>
<td>41 (21)</td>
<td>44 (24)</td>
<td>35 (19)</td>
<td>37 (18)</td>
</tr>
<tr>
<td>/t/</td>
<td>Pre</td>
<td>70 (24)</td>
<td>67 (23)</td>
<td>46 (20)</td>
<td>46 (26)</td>
<td>51 (33)</td>
<td>34 (17)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>65 (22)</td>
<td>63 (28)</td>
<td>46 (22)</td>
<td>46 (23)</td>
<td>41 (21)</td>
<td>28 (14)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>65 (23)</td>
<td>67 (27)</td>
<td>44 (15)</td>
<td>47 (26)</td>
<td>43 (21)</td>
<td>36 (21)</td>
</tr>
<tr>
<td>/k/</td>
<td>Pre</td>
<td>72 (22)</td>
<td>77 (18)</td>
<td>57 (16)</td>
<td>54 (17)</td>
<td>60 (30)</td>
<td>46 (14)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>66 (19)</td>
<td>68 (21)</td>
<td>51 (17)</td>
<td>54 (21)</td>
<td>58 (18)</td>
<td>44 (10)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>66 (19)</td>
<td>71 (24)</td>
<td>57 (14)</td>
<td>63 (24)</td>
<td>55 (18)</td>
<td>48 (9)</td>
</tr>
</tbody>
</table>

Table 2.6 Average Rating (1-3 points) of Approximant and Rhotic Phones

<table>
<thead>
<tr>
<th></th>
<th>First Year</th>
<th></th>
<th>Second Year</th>
<th></th>
<th>Third Year</th>
<th></th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+PI (n=19)</td>
<td>-PI (n=20)</td>
<td>+PI (n=17)</td>
<td>-PI (n=9)</td>
<td>+PI (n=10)</td>
<td>-PI (n=8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td>X (s)</td>
<td></td>
</tr>
<tr>
<td>/β/</td>
<td>Pre</td>
<td>1.33 (.34)</td>
<td>1.24 (.28)</td>
<td>1.44 (.44)</td>
<td>1.25 (.22)</td>
<td>1.35 (.43)</td>
<td>1.47 (.59)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>1.24 (.42)</td>
<td>1.13 (.36)</td>
<td>1.24 (.36)</td>
<td>1.11 (.33)</td>
<td>1.6 (.52)</td>
<td>1.38 (.35)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>1.24 (.27)</td>
<td>1.31 (.38)</td>
<td>1.32 (.26)</td>
<td>1.28 (.40)</td>
<td>1.4 (.39)</td>
<td>1.38 (.40)</td>
</tr>
<tr>
<td>/ð/</td>
<td>Pre</td>
<td>1.12 (.24)</td>
<td>1.19 (.24)</td>
<td>1.51 (.61)</td>
<td>1.27 (.35)</td>
<td>1.43 (.41)</td>
<td>1.56 (.46)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>1.11 (.25)</td>
<td>1.09 (.17)</td>
<td>1.35 (.42)</td>
<td>1.24 (.32)</td>
<td>1.17 (.25)</td>
<td>1.66 (.64)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>1.11 (.28)</td>
<td>1.23 (.32)</td>
<td>1.52 (.43)</td>
<td>1.31 (.37)</td>
<td>1.35 (.47)</td>
<td>1.83 (.61)</td>
</tr>
<tr>
<td>/γ/</td>
<td>Pre</td>
<td>1.29 (.33)</td>
<td>1.61 (.55)</td>
<td>1.54 (.53)</td>
<td>1.19 (.27)</td>
<td>1.45 (.54)</td>
<td>1.78 (.66)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>1.35 (.29)</td>
<td>1.38 (.39)</td>
<td>1.46 (.49)</td>
<td>1.31 (.27)</td>
<td>1.24 (.36)</td>
<td>1.81 (.68)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>1.41 (.45)</td>
<td>1.58 (.45)</td>
<td>1.56 (.39)</td>
<td>1.44 (.35)</td>
<td>1.58 (.55)</td>
<td>1.69 (.65)</td>
</tr>
<tr>
<td>/ɾ/</td>
<td>Pre</td>
<td>1.78 (.78)</td>
<td>2.13 (.68)</td>
<td>2.47 (.64)</td>
<td>2.67 (.47)</td>
<td>2.21 (.75)</td>
<td>2.66 (.38)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>2.06 (.74)</td>
<td>2.38 (.64)</td>
<td>2.49 (.47)</td>
<td>2.78 (.34)</td>
<td>2.45 (.73)</td>
<td>2.59 (.67)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>1.86 (.66)</td>
<td>2.16 (.70)</td>
<td>2.51 (.54)</td>
<td>2.69 (.66)</td>
<td>2.28 (.73)</td>
<td>2.69 (.70)</td>
</tr>
<tr>
<td>/r/</td>
<td>Pre</td>
<td>1.74 (.53)</td>
<td>1.8 (.42)</td>
<td>2.03 (.52)</td>
<td>2.39 (.52)</td>
<td>2.03 (.70)</td>
<td>2.25 (.23)</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>1.94 (.51)</td>
<td>1.95 (.37)</td>
<td>2.15 (.56)</td>
<td>2.37 (.54)</td>
<td>2.15 (.66)</td>
<td>2.38 (.27)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>1.82 (.56)</td>
<td>2.01 (.55)</td>
<td>2.15 (.61)</td>
<td>2.09 (.45)</td>
<td>2.1 (.65)</td>
<td>2.32 (.54)</td>
</tr>
</tbody>
</table>

The differences in N between Table 2.5 and Table 2.6 reflect the fact that some learners’ data were removed from the analysis because the learners did not complete the posttest or took the posttest at the wrong time. These types of problems occurred most often with second year learners during the third session, because there was no lab assistant present that day to help keep all learners following instructions.
2.3.4 Distribution of Scores

Production pretest, posttest, and delayed posttest scores were analyzed for normality of distribution across each of the three matriculation levels. A Shapiro Wilk test found all production tests scores to be normally distributed ($p > .05$) with just one exception: first year learners’ pretest. The data exhibited no significant skew or kurtosis. For each matriculation level at each test administration time, the skewness statistic was divided by the skew standard error. All were between the values of -2 and 2 with one exception: first year learners’ pretest scores were slightly skewed (2.27). The kurtosis statistics were similarly divided by the kurtosis standard errors and were between the values of -2 and 2 for all production tests.

2.3.5 RM ANOVA and Subsequent Contrasts

Repeated measure analyses of variance (RMANOVAs) were used to compare the effects and interactions of test time, instructional condition, and matriculation level. The RMANOVAs were used to analyze scores on the full test (all phones) as well as individual phones. The within-groups factor was time of test (pretest, immediate posttest, and delayed posttest) and the between-groups factors were instructional condition (+PI and -PI) and matriculation level (first, second, and third year). On the full production test (aggregate of all phones), there was a main effect for time $F(1.64, 125)^{14} = 4.34, p = .02, \eta_p^2 = .05$, but no interaction reached significance (all $F \leq .82$, all $p \geq .05$). The results of

---

14 Note that for this main effect, Mauchly’s test indicated that the assumption of sphericity was violated, and so degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity. This procedure was followed each time sphericity could not be assumed.
the RMANOVA are presented in Table 2.7. The pairwise comparisons\textsuperscript{15}, presented in Table 2.8, indicated that first year students’ scores were significantly lower than those of second and third year students but that second and third year students’ scores were not significantly different. This finding merely reiterated the findings of the T-tests reported in §2.3.3 and because the level by time interaction did not reach significance, there was no evidence that matriculation level affected change across time, after instruction. More interestingly, the pairwise comparisons indicated that learners’ posttest scores (\(\bar{x} = 15.29, s = 3.80\)) were significantly greater than their pretest scores (\(\bar{x} = 14.89, s = 3.80\)), but their delayed posttest scores (\(\bar{x} = 15.29, s = 3.78\)) were not. Thus for the production test overall (aggregate of 8 target phones) learners improved slightly immediately after instruction, with time of test accounting for 5\% of the variation in scores, but instructional condition and matriculation level did not have a significant effect. The aggregate production test scores are plotted in Figure 2.17.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|c|c|}
\hline
Source & SS & df & MS & F & \(\eta^2\) & Power \\
\hline
\hline
Between Subjects & & & & & & \\
Condition & 27.26 & 1.00 & 27.26 & 1.34 & 0.02 & 0.21 \\
Level & 521.95 & 2.00 & 260.97 & 12.83**** & 0.25 & 1.00 \\
Condition X Level & 99.10 & 2.00 & 49.55 & 2.44* & 0.06 & 0.48 \\
Error & 1546.04 & 76.00 & 20.34 & & & \\
\hline
Within Subjects & & & & & & \\
Time & 6.02 & 1.64 & 3.66 & 4.34** & 0.05 & 0.69 \\
Time X Condition & 0.16 & 1.64 & 0.10 & 0.12 & 0.00 & 0.07 \\
Time X Level & 2.27 & 3.29 & 0.69 & 0.82 & 0.02 & 0.23 \\
Time X Condition X Level & 1.14 & 3.29 & 0.35 & 0.41 & 0.01 & 0.13 \\
Error (Time) & 105.33 & 124.94 & 0.84 & & & \\
\hline
\end{tabular}
\caption{RMANOVA of Production Test: Aggregate of 8 Phones}
\end{table}

\textsuperscript{15} These and all subsequent pairwise comparisons were adjusted for multiple comparisons with the Bonferroni procedure.
A RMANOVA was also used to compare scores across individual phones. The within-groups factors were time of test (pretest, immediate posttest, and delayed posttest) and phone ([p, t, k, β, δ, γ, r, r]) and the between-groups factors were instructional condition (+PI and -PI) and matriculation level (first, second and third year). The results of the RMANOVA are presented in Table 2.9. There was a main effect for time of test $F(2, 144) = 3.92, p = .02, \eta_p^2 = .05$ and a main effect for phone type $F(3.76, 270) = 98.42, p < .001, \eta_p^2 = .58$. Note that instructional condition approached but did not reach statistical significance in interaction with phone ($p = .06$). The interactions that reached
significance were phone type by matriculation level $F(7.51, 270) = 3.21, p = .002, \eta^2_p = .08$, and phone type by time of test $F(10.38, 747) = 3.40, p < .001, \eta^2_p = .05$. The latter interaction (phone type by time of test) is more of interest to the analysis here, both because differences between matriculation levels are to be expected and have already been discussed here, and also because it is the change in scores over time following instruction that is the main concern of the present study. Figure 2.18 plots the phone by time interaction.
This phone type by time of test interaction was further investigated by comparing gain scores on the target phones. The target phones could be ordered by size of gains from pretest to posttest as follows: /p/ (\(\bar{x} = 0.18, s = 0.42\)) > /t/ (\(\bar{x} = 0.17, s = 0.45\)) > /k/ (\(\bar{x} = 0.14, s = 0.34\)) > /ɾ/ (\(\bar{x} = 0.12, s = 0.43\)) = /r/ (\(\bar{x} = 0.12, s = 0.35\)) > [β] (\(\bar{x} = -0.06, s = 0.43\)) = [ɣ] (\(\bar{x} = -0.06, s = 0.41\)) > [ð] (\(\bar{x} = -0.09, s = 0.29\)). Few differences reached significance (\(p < .05\)) however. The average pretest to posttest gain score for [ð] was significantly lower than all other phones, and the gain score for [ɣ] was significantly lower than the phone with the highest gain score, /p/, but none of the other differences reached significance. As for the gain scores from pretest to delayed posttests, none of the differences between phones reached significance. The pattern that emerged, then, in terms of a phone by time interaction, was that leaners improved their production of stops and rhotics significantly more than approximants, particularly [ð], but only immediately following instruction. In the delayed posttest, there were not significant differences between phones.
To determine the effects of time, condition, and level on individual phones, RMANOVAs were used to analyze scores on each phone, with the within-groups factor of test time (pretest, immediate posttest, and delayed posttest) and the between-groups factors of instructional condition (+PI and -PI) and matriculation level (first, second and third year). As will be discussed below, there was a paucity of significant differences other than the main effect of time. Therefore, the RMANOVA tables will not be presented. Furthermore, only the significant main effects and interaction effects related to the variable of time will be discussed, as it is the change in scores over time following instruction that is the main concern of the present study.

The results pertaining to VOT of the stop consonants /p, t, k/ will be presented first. While there was no significant main effect or interaction when all four items of each phone were included in the analysis, significant effects were found when the analysis was limited to the items in which the target phones were in stressed syllables (i.e., para, perro, pace, tal, tú, talle, que, como, and cace, but not pintar, tocar, or cubano). Both types of items were included in the test because it was expected that the learners would aspirate /p, t, k/ both word-initially and in stressed syllables (as reported in Hualde, 2005). However, since a clear pattern emerged with /p, t, k/ in stressed syllables, and the stimuli of similar studies (e.g., González-Bueno, 1997) have analyzed /p, t, k/ in stressed syllables only, the results of those analyses will be presented below.

For all stop consonants in initial, stressed position there was a significant effect for time (all $F > 4.23, p < .02$), with time of test accounting for 7 – 13% of variance in scores. Pairwise comparisons confirmed that posttest VOTs (in milliseconds) for all stop consonant phones were shorter than pretest VOTs: /p/ (pretest $\bar{x} = 51, s = 26$; posttest $\bar{x}$
= 44, s = 22), /t/ (pretest \( \bar{x} = 61, s = 30 \); posttest \( \bar{x} = 56, s = 28 \)), and /k/ (pretest \( \bar{x} = 65, s = 23 \); posttest \( \bar{x} = 58, s = 22 \)). However, only for /k/ was the delayed posttest VOT significantly shorter than the pretest (pretest \( \bar{x} = 65, s = 23 \); delayed posttest \( \bar{x} = 61, s = 21 \)). One interaction reached significance, which was the interaction of time by condition by level for the phone /t/ \((F(3.69, 144) = 2.63, p = .04, \eta^2_p = .06)\). This interaction was analyzed with corrected paired samples T-tests, which indicated that there were no significant differences in the VOTs that sub groups of learners produced for /t/\(^16\). In sum, then, learners in both instructional conditions reduced VOTs for the stop consonants /p, t, k/ in initial, stressed position immediately following instruction. This reduction in VOT was retained 3 weeks later in the case of /k/.

The results of the RM ANOVAs for the scores of the approximant target phones [\(\beta, \delta, \gamma\)] were quite disparate. For [\(\beta\)], there were no significant main effects or interactions (all \(F < 1.56, p > 0.20\)), suggesting that neither instructional condition was useful for improving learners’ production of [\(\beta\)], at least during the time period of this study. For [\(\delta\)], there was a main effect of time \((F(2, 154) = 6.02, p < .01, \eta^2_p = .07)\), but no significant interaction was found (all \(F < 2.36, p > 0.08\)), suggesting that instruction of either type was associated with a small but significant change in the ratings learners’ received on their production of [\(\delta\)]. Pairwise comparisons indicated that ratings of learners’ production of [\(\delta\)] on the pretest (\(\bar{x} = 1.43, s = 0.58\)) were not significantly different from their posttest ratings (\(\bar{x} = 1.35, s = 0.57\)) or their delayed posttest ratings.

\(^{16}\) The uncorrected paired samples T-tests indicated that the only significant change in VOT of /t/ occurred with third year learners in the –PI group, whose posttest VOT for /t/ (\(\bar{x} = 35 \text{ ms}, s = 19 \text{ ms}\)) was significantly shorter than the pretest (\(\bar{x} = 43 \text{ ms}, s = 21 \text{ ms}\)) \((t(8) = 2.68, p = 0.03, d = 0.37)\). As for the other sub groups, their VOT of /t/ did not change significantly (all \(t < 1.90, p > .09\)). However, once the alpha level was submitted to a Bonferroni correction to control for Type 1 error, the difference found for third year learners did not reach statistical significance.
but rather that the significant effect was the increase in ratings from their posttest to delayed posttest. It should be noted that there was no qualitative difference in learners’ production between pretest and posttest productions of [ð]. Rather, learners tended to incorrectly realize [ð] as [d] on both tests, but they did so more often on the posttest, particularly for the test items adiós (28% decrease in accuracy compared to pretest) computadora (15%), and avenida (11%). As for [ɣ], there was a significant main effect for time $F(2, 154) = 3.27, p = .04, n_p^2 = .04$, but pairwise comparisons indicated that neither the posttest nor delayed posttest ratings were significantly different from the pretest (all $p > .05$). There was also a significant interaction between time, condition, and level $F(4, 154) = 3.37, p = .01, n_p^2 = .08$. This interaction was analyzed with corrected paired samples T-tests, which indicated that there were no significant differences in the ratings sub groups of learners received for their production of [ɣ].

In sum, then, neither instructional condition was associated with an increase in the ratings of learners’ productions of [β, δ, γ] over time. There was no change over time for [β], a small decrease in ratings post-instruction for [ð], and little change over time for [ɣ].

The change in learners’ productions of the rhotic phones over time was more similar to that of the stop consonants than the approximants. The RM ANOVAs on the

---

17 The uncorrected paired samples T-tests indicated that, while there was no significant difference between pretest and posttest or delayed posttest ratings for [ɣ] (all $t < 1.97, p > .06$), two comparisons of scores were significant: the posttest ratings of first year learners in the –PI group ($\bar{x} = 1.41, s = 0.40$) were significantly lower than their pretest ratings ($\bar{x} = 1.61, s = 0.52$) ($t(22) = 2.20, p = 0.04, d = 0.07$), and the delayed posttest ratings of second year learners in the –PI group ($\bar{x} = 1.46, s = 0.31$) were significantly higher than their pretest ratings ($\bar{x} = 1.23, s = 0.26$) ($t(10) = 2.47, p = 0.03, d = 0.79$). The ratings of learners in the +PI group did not vary significantly over time, though the improvement in posttest ratings of first year learners in the +PI group ($\bar{x} = 1.34, s = 0.33$) neared statistical significance compared to the pretest (mean = 1.26, $s = 0.33$) ($t(18) = 1.97, p = 0.06$). However, once the alpha level was submitted to a Bonferroni correction to control for Type I error, none of these differences was deemed statistically significant.
rhotic data indicated that there was significant main effect of time for both /ɾ/ \(F(1.85, 143) = 3.63, p = .03, \eta^2_p = 0.05\) and /r/ \(F(2, 146) = 4.19, p = .02, \eta^2_p = .05\). There were no significant interactions for either phone (all \(F < 2.02, p > 0.10\)). Pairwise comparisons indicated that ratings of learners’ posttest productions of /ɾ/ \((\bar{x} = 2.40, s = 0.64)\) were significantly higher than their pretest \((\bar{x} = 2.24, s = 0.72)\) but that their delayed posttest ratings \((\bar{x} = 2.29, s = 0.71)\) were not significantly higher than the pretest. Likewise, ratings of learners’ posttest productions of /r/ \((\bar{x} = 2.09, s = 0.50)\) were significantly higher than their pretest \((\bar{x} = 1.95, s = 0.53)\), but their delayed posttest ratings (mean = 2.07, \(s = 0.57\)) were not significantly higher than their pretest. In sum, learners in both conditions made some small but significant improvement in their production of the rhotic phones after instruction, though this improvement was not retained in the 3-week delayed posttest.

Overall, then, the RM ANOVAs completed for each target phone indicated that after receiving either type of instructional intervention, learners slightly improved their production of rhotics and stops in stressed syllables but did not improve their production of approximants. The improvement was maintained until the 3-week delayed posttest only for /k/.

2.4 Discussion

2.4.1 Summary of Results

The first research question asked whether instruction in L2 phonetics and phonology would improve learners’ ability to produce more native-like L2 phones.
Based on the generally positive effects found in the literature, it was hypothesized that instruction would prove beneficial for learners’ production of the target phones. However, the data did not suggest that the phonetics instruction provided any advantage in the production test, either for individual phones or for all the phones analyzed together. The only effect that reached significance for almost all phones and for the aggregate test was time. The main effect of time indicated that learners in both instructional conditions improved their pronunciation of most phones, at least immediately following instruction.

The second research question asked whether the effectiveness of the instruction would depend on learners’ experience level, operationalized here as their current matriculation level. It was hypothesized that the effectiveness of instruction would vary by target phone and would interact with learners’ curricular level in possibly complex ways. There was an interaction with matriculation level for just two phones: /t/ and [ɣ], but for both phones, the only differences found between matriculation levels involved learners in the -PI, and these differences did not reach statistical significance once they were corrected for multiple comparisons. Therefore the data did not suggest that learning from the phonetics instruction was influenced by matriculation level.

Much like previous studies, these data suggested that instruction did not affect all L2 phones in equal measure. Elliott (1997) provided an extensive discussion of why some of these target phones might respond to instruction differently, basing his claims variously on known contrasts between Spanish phonology and English phonology, notions of universal markedness, and general theories of phonological development. Of particular interest in the current data were the approximants ([β, ð, γ]), which did not seem to improve with experience across matriculation levels prior to instruction and did
not improve following instruction as did the other phones. Empirical data suggests that these approximants are resistant to instruction and are late acquired (Díaz-Campos, 2004; Zampini, 1993). It may be that the spectral cues differentiating the approximants from their analogous stops /b, d, g/ are less well perceived by English speaking learners than the differences between the other target phones and their analogous L1 phones, which has been claimed to predict learnability (e.g., Flege, 1995). Rather, it may have to do with the class or approximants being more universally marked than stops and therefore later acquired (Jakobson, 1941; Eckman, 1977). González-Bueno and Quintana-Lara (2010) suggested that learners start to recognize the spirantization rule (the rule for when stops should be realized as approximants) for [ð] and [ɣ] around the intermediate proficiency level, whereas [β] is not acquired until more advanced levels. Learners in the current study did not improve their pronunciation of [β] over time in response to either instructional condition, and they may not have been developmentally ready to do so because of their intermediate level of Spanish proficiency.

The present data supported much of what was reported in previous pronunciation research. Learners have improved their pronunciation of these consonantal phones as they gained L2 experience yet have not typically reached native speaker norms even after achieving advanced levels of proficiency (e.g., Face & Menke, 2010; Reeder, 1998). Pronunciation instruction has lead to modest improvement in learners’ pronunciation overall and for some phones in particular, namely the stops /p, t, k/ in stressed syllables and the rhotic phones /ɹ/ and /ɾ/ (Elliott, 1995, 1997; González-Bueno, 1997; Lord, 2005). The approximant phones [β, ð, ɣ] did not improve post-instruction in the present study, which concurs with what Elliott (1997) found but stands in opposition to Lord’s (2005)
study with more advanced learners in a full semester phonetics course. One could argue that the length of the instructional intervention was crucial, that is, that a longer instructional treatment would have been required to produce a significant effect on learners’ pronunciation the approximant phones. However, it is unclear that this should be the case, since learners’ pronunciation of these phones actually worsened immediately following instruction.

2.4.2 Implications

The most important contribution made by the present study was its demonstration that, for most phones, the control condition was just as effective for improving learners’ pronunciation as the explicit pronunciation instruction. Pronunciation instruction has typically been assessed as a holistic approach, including several methodological tools, such as production practice and feedback, along with the explicit teaching of phonetics. The design of the present study allowed the effect of explicit phonetics teaching to be separated from several other factors that could potentially affect learners’ pronunciation. The control group in this study participated in a focused listening activity and took dictation from native speakers’ speech. Their learning condition exposed them to native renditions of the target phones in roughly equal amounts as the instructed group and also required them to focus on the form (sound) of the input in order to take dictation. The control group practiced pronouncing the target phones in amounts roughly equal to the instructed group. The feedback both groups received was identical; they simply compared their speech to that of a native speaker. Therefore, the main difference between the groups’ instruction was the explicit lessons on Spanish phonetics and phonology. They
were instructed as to the point, place, and manner of articulation; grapheme-phoneme correspondences in Spanish; and relevant contrasts between Spanish and English phonology. The instructed group did not perform better than the control group in subsequent tests of their pronunciation.

The data indicated that the explicit teaching of phonetics was not as helpful as previously thought. It has been suggested that the explicitness of the phonetics instruction, above other factors, is the most necessary component of pronunciation instruction (Derwing & Munro, 2005, p. 388; Fullana, 2006; Venkatagiri & Levis, 2007), and pronunciation practice with feedback alone has not been shown to improve learners’ pronunciation of these target Spanish phones (Ducate & Lomicka, 2009). However, the current study took a first step towards isolating the explicit phonetics teaching from other factors such as exposure and attention to the target phones, which recent research indicates may be relevant for acquisition of L2 speech (e.g., Chung, 2008). With the other factors held relatively constant, this study found no added benefit from explicit phonetics instruction. Elliott (1997) suggested that instruction was likely to affect performance on discrete word-level tasks more than spontaneous speech. The present study utilized a word reading task and yet still found no significant effect for instructional condition, which strengthens the argument that explicit phonetics instruction may be less impactful than previously thought.

The second contribution of the present study was its recruitment of learners at multiple matriculation levels in order to address the issue of curricular sequencing for pronunciation instruction. Previous investigations have included only learners enrolled in an intermediate or advanced L2 class (e.g., Elliott, 1995, 1997; González -Bueno, 1997;
Lord, 2005). Yet those researchers (Elliott, 1995; Lord, 2005) expressed an interest in carrying out similar studies with more novice learners. Contrary to the notion that earlier is better for pronunciation instruction (Arteaga, 2000), matriculation level was not a significant factor in this study. The data suggested that first, second, and third year learners responded similarly to the phonetics instruction and the dictation exercises for most phones. Type of instruction and matriculation level interacted for just two phones, /t/ and [ɣ], but the (non significant) differences were found to be associated with the control group, not the experimental group.

The question of curricular sequencing did seem material, however, in the analysis of the post-instructional questionnaires. Learners rated their lessons overall and in terms of helpfulness, usefulness, and difficulty, on a scale of 0-10. Ratings reported by the +PI and -PI groups were compared for each matriculation level using independent samples T-tests. One difference was noteworthy. Only first year learners in the +PI rated their lessons as better than the -PI overall, $t(38) = 2.04, p = .05$. No such difference was found with second and third year learners. Thus if one believes that learners’ appraisals of instructional materials may influence their engagement and learning, one should note that first year learners appraised the phonetics lessons more highly than the dictation exercises.

It seems fair to conclude that both types of methodologies assessed here, pronunciation instruction with explicit phonetics instruction as well as focused listening with dictation, were beneficial for learners. Both instructional techniques led to significant improvement of pronunciation in the short term, at least for most phones. Only one phone ([β]) out of the eight target phones did not respond at all to instruction.
The effect of instruction was small, around 5%, and it was not retained 3 weeks post-instruction for most phones, so neither instructional treatment was a panacea for decreasing accent. The effect may have been greater had the instructional treatment been longer. However, it is fair to say that the time allotted was sufficient to impart explicit phonetics instruction, which was the focus of the study. In their post-module and post-study questionnaires, learners did not report that they wanted or needed more time to complete the modules or wished to have more in-depth information about phonetics. Frequent comprehension checks required that students understood the information presented in each section of the phonetics modules before moving on, and no student was unable to complete the modules in the time allotted. Although the instructional treatment was briefer than in some prior studies (e.g., Elliott 1995; Lord 2005), it was equivalent to others (e.g., González-Bueno, 1997). Furthermore, the type and length of the instructional modules assessed in the present study could be viewed as assets of the study design, as the modules are easily incorporable into contemporary Spanish L2 curricula. The modules are free and available online. They require neither an onerous time commitment on the part of the student nor expertise in phonetics on the part of the instructor. The modules could be recycled at multiple points during a course or across a curriculum, and the current study indicates that learners at different proficiency levels could benefit from completing either series of modules.

2.4.3 Limitations of the Production Test

The current study had many limitations, which will be discussed at length in Chapter 5. The present chapter is concerned only with production data, and so the
limitations of the production test will now be addressed. The production test required learners to read isolated words and phrases from a printed list. Word reading could have exaggerated learning effects from instruction or have had the opposite effect and incited more native language interference as compared with spontaneous, unmonitored speech (Tarone, 1979). The results of the present study cannot be directly compared with studies that utilized other types of tasks, and the results should not be extrapolated to infer the accuracy with which learners would have produced the same phones in spontaneous speech. Also, two items in the production test were Spanish-English cognates (agosto, cubano) that may have exacerbated transfer effects for the relevant target phones.

2.4.4 Conclusion

By means of conclusion, the author wishes to note three issues that are important to advancing research in phonetics/pronunciation instruction. The first issue was addressed in the current study, which is the need to tease apart the many elements of pronunciation instruction in order to better understand the relative contribution of each and thereby improve and tailor instructional techniques for teaching pronunciation to L2 and FL learners. The second issue was incorporated in the present study’s design but not addressed directly, which is the need to reconsider the native speaker ideal often assumed in L2 pronunciation research. For example, Lord (2005) recruited native speakers only to provide baseline data for the VOT of /p, t, k/, but assumed for all other phones and features that any native speaker would perform consistent with idealized native speaker norms. The present study, however, recruited college-educated, balanced bilingual speakers with native accents in Spanish to provide baseline data for all the target phones.
Their VOTs for stops were longer than expected, and their productions of the approximant phones did not receive perfect ratings. That is, their speech did not always reflect the “idealized” native speaker norms referenced in previous literature. Yet, in the author’s opinion, their speech is still an appropriate target for the learners in the current context. In fact, these bilingual native speakers of Spanish may represent a more suitable target for the FL learners recruited. Research on pronunciation instruction should begin to better reflect the “bilingual turn” being advocated for SLA research more generally (Ortega, 2009).

The third issue is whether accentedness is in fact worthy of future study. It has been argued here that accent is important both because learners are concerned with their accent and because accentedness can in some cases impact comprehensibility and intelligibility. However, it is fair to say that comprehensibility and intelligibility are more consequential to L2 speakers than accent (Derwing & Munro, 1997). Achieving a target-like accent may even be an unrealistic and de-motivating goal for learners (Levis, 2005). Thus researchers, teachers, and learners alike need to consider carefully what level of relative importance they are willing to assign to accentedness. In the author’s opinion, researchers should strive to balance measures of all three in future studies – accentedness, comprehensibility and intelligibility.

The present study attempted to assess the effectiveness of explicit teaching about phonetics and phonology as separate from other facets typically included in pronunciation instruction. The data suggest that the explicit lessons on phonetics and phonology added no measurable benefit for learners’ pronunciation of the target phones. Obviously, additional research must be carried out to verify this result. The study also
measured changes in learners’ perception of the target phones, which is the subject of the next chapter.
CHAPTER 3: PERCEPTION

3.1 Introduction

3.1.1 Levels of Speech Perception

Acquiring target-like perception of L2 phones is advantageous, not only because target-like perception aids in word identification and parsing of the speech stream (Sebastián-Gallés, 2005), but also because target-like perception purportedly facilitates and/or is necessary for acquiring target-like pronunciation. It was established in §1.2 that most models of L2 speech acquisition posit a connection between the perception and production of speech, though the precise nature of the relationship between perception and production is not yet entirely understood (Akerberg, 2005). Empirical evidence was cited as well demonstrating that perception is important for production (e.g., Munro, 2008). The Speech Learning Model (SLM) (Flege, 1988, 1992, 1995, 1999, 2002, 2003a) was discussed in detail. The SLM is the only model of L2 speech acquisition that makes explicit claims about the relationship between perception and production, and it is the model that motivated the methodology of the present study. The SLM posits that to fully acquire L2 sounds learners must first learn to discern subtle phonetic differences between L2 sounds and analogous sounds in the L1 as they occur in a variety of phonetic environments, at which point learners can create new phonetic categories for L2 sounds. Finally, they may be able to produce the L2 sounds in target-like ways.

The ability to “discern” L2 sounds, in the language of the SLM, is related to perception. It was noted in §1.2 that perception is a broad term, encompassing multiple
levels of processing. For example, Studdert-Kennedy (1974) delineated four independent stages of auditory perception: auditory, phonetic, phonological, and lexical/syntactic/semantic. Wode (1981; 1990) described two modes of processing, a continuous and a categorical. The latter was dependent on acquired L1-based phonemic categories, but the former was not. Gaskell and Marslen-Wilson’s phonological inferencing model drew a distinction between the levels of “phonetic” and “phonological” processing (1998). Strange (2002) referred to the same distinction as “categorial” and “categorical.” While contemporary models of auditory perception and word recognition maintain that auditory input is processed in multiple stages, these stages are no longer characterized as fully independent (see Akerberg, 2005, for a review).

Adults prefer to process speech phonemically/categorically under normal conditions, though their phonetic/continuous (also called “acoustic”) processing abilities remain intact throughout the life span (e.g., Dehaene-Lambertz et al., 2005; Dehaene-Lambertz & Gliga, 2004; Pisoni, 1973; Repp, 1983; Werker & Logan, 1985; Williams, 1979). Adults’ tendency to use their L1 knowledge and the phonemic/categorical processing mode for L2 speech hinders their ability to perceive L2 speech like a native speaker of that L2. The SLM does not characterize the acquisition process in terms of continuous or categorical processing and does not address whether or not the two perceptual stages of the acquisition it proposes are analogous to this processing distinction. According to the SLM, creation of new phonetic categories for L2 sounds is possible for an L2 learner as long as she can detect at least some of the phonetic

---

18 Flege (1995) only briefly mentions the existence multiple level of processing, stating that “a failure to discern cross-linguistic phonetic differences may arise at,” the point of “access to sensory properties,” “the on-line processing of speech” or “pre-attentive processes,” yet these levels are not treated in detail (p. 241).
differences between L2 sounds and their analogous L1 counterparts, at which point she can begin creating new phonetic categories for L2 sounds. Similarly, Strange and Shafer described the process of acquiring an L2 sound system as first detecting differences between native and non-native sounds and then developing the appropriate selective perceptual routines (SPRs) to “hear” them reliably (2008). However, it is not entirely clear which are the appropriate tasks and task conditions with which to measure learners’ abilities in these domains. Exactly how should one test a learner’s ability to “discern” or “detect” the difference between L2 and L1 sounds? How should one test whether a learner has indeed created new phonetic categories for L2 sounds? § 3.1.2 discusses the methodological challenges implicated in L2 speech perception research.

3.1.2 Methodological Issues in L2 Speech Perception Research

Attempts have been made to separately assess different levels of processing with different experimental designs (e.g., Polka, 1992). The two most commonly used experimental techniques are phoneme discrimination and identification tasks (see Boomershine, Hall, Hume & Johnson, 2008; Logan & Pruitt, 1995, for reviews). Neither task type is without limitations. Indeed, Flege (2003b) notes that, among other problems, discrimination tasks lack ecological validity, whereas identification tasks induce more realistic processing demands yet usually mediate phonological processing through written script. The limitations of these task designs have sometimes been overcome with task-specific training, yet providing training necessarily alters the object of investigation (Flege, 2003b). Discrimination tasks present learners with stimuli, usually in AX or ABX fashion, i.e., the two target stimuli are presented in succession, or there is a
distracter sound placed between them. Learners determine whether the speech stimuli are identical. Learners may simply choose one of two responses (i.e., “same,” “different”) or they may rate the stimuli in terms of likeness, for instance, on a Likert scale from 1 – “not similar” to 4 – “very similar” (Archila-Suerte, Zevin, Bunta & Hernandez, 2011). Learners may rate their confidence level in their own responses. Identification tasks require learners to identify which phone they hear. For instance, an experiment of the American English /ɹ/-/l/ contrast might present learners with the auditory stimulus “road” and ask learners to decide whether they heard “road” or “load.”

However, both discrimination and identification tasks could hypothetically tap either level of perception (phonetic/continuous/acoustic or phonemic/categorical) depending on the task conditions and, in fact, it is difficult, perhaps impossible, to completely separate the two types of processing in practice, because they occur automatically, rapidly, and in parallel fashion (Boomershine, Hall, Hume & Johnson, 2008). Thus it is important to specify the exact conditions used in an experimental perception task. For example, significant differences in processing have been found across different types of linguistic stimuli. Consonantal stimuli, which have temporally short cues, are perceived differently than the relatively steady-state vocalic stimuli, and consonants are perceived differently if presented in isolation (truncated stimuli) versus with surrounding vowels (Werker & Logan, 1985).

The timing of the presentation of stimuli is highly important as well. As a case in point, Werker and Logan manipulated a range of inter-stimulus intervals (ISIs) in a study of adult L2 consonantal contrast perception and found evidence for at least three independent levels of processing (1985) – “auditory,” “phonetic,” and “phonemic,” at
least in the initial stages of exposure to a new L2. Their participants (n=30) were adult
speakers of English with no knowledge of Hindi. They were trained to discriminate the
Hindi voiceless, unaspirated retroflex and /t/ dental /t/ contrast. Participants were
presented with the target phones in an AX discrimination task, with a two-choice button
press response (“same”/ “different”). The discrimination task included three different ISI
conditions: 250ms, 500ms, and 1500ms. The consonantal contrast was presented in three
different types of stimuli pairings: phonetically identical, phonetically different but
phonemically identical, and phonemically different.

Werker and Logan argued that high accuracy on the task was evidence of
“phonetic processing,” which they described as processing phonetic detail of the stimuli
that are relevant to the L2 (Hindi) but irrelevant to the L1 (English), whereas low
accuracy had two possible explanations. Low accuracy could be due to false negatives,
when participants did not detect the retroflex/dental contrast that was presented, engaging
rather in “phonemic processing,” which Werker and Logan described as L1-like
processing. Low accuracy could also be due to false positives, when participants
detected a contrast that was not actually presented, engaging rather in “auditory
processing” (also called “psychoacoustic” processing), which they described as
processing for fine-grained acoustic information that is irrelevant to both L1 and L2. The
false positives were elicited with the phonetically different but phonemically identical
stimuli pairings. Werker and Logan argued that the categorical/continuous processing
distinction should be further subdivided into at least three levels, because continuous
processing could be either phonetic or auditory. Auditory processing has been described
as the unanalyzed, fleeting sensory memory trace that is available for a very short time after a stimulus has been presented (Munro, 2008).

The participants in their study had no prior exposure to Hindi phonology, so it is unclear to what degree Werker and Logan’s (1985) methodological choices and interpretations of the results would be equally suitable for L2 learners at a later stage in L2 acquisition. What is more relevant for the present study is their conclusion that adults “under task conditions that are most similar to those used in everyday oral communication (long intervals between repetitions of the same exemplar; high memory demands) … rely on a language-specific phonemic processing strategy” (p. 43). Their results caution against employing ISIs as short as 500ms in AX discrimination tasks if the goal is to assess learners’ categorical processing rather than acoustic processing. An ISI of 1500ms is “the ISI condition most likely to tap linguistic categories rather than auditory sensitivities” and tap how “speakers naturally respond to these syllables” (p. 42).

Apart from varying ISIs and using consonants and vowels either alone or in isolation, another methodological complication in perception studies is the issue of how to assess perception of non-contrastive phones. The vast majority of experimental L2 speech perception research reported to date deals with linguistic targets that are contrastive in the L2 (e.g., /u/-/i/ or /i/-/u/). Such contrasts are readily incorporated into syllable-length and even word-length stimuli suitable for discrimination and identification tasks, and they are obviously important for acquiring an L2. However, the present study was concerned with accent in L2 and endeavored not to assess learners’ ability to detect Spanish (L2) contrasts but rather assess learners’ ability to discern subtle phonetic differences between L2 sounds and analogous sounds in the L1, which is the
first stage of phonological acquisition according to the SLM. The present study thus incorporated a discrimination task that paired Spanish-like stimuli with English-like stimuli and an identification task that required learners to respond (in writing) with which phone they had identified in the Spanish speech stimuli. While each task targeted a separate stage of L2 acquisition posited in the SLM, they were not designed to tap different processing levels. The task conditions were both relatively “offline,” as the identification task was untimed and the discrimination task had a long ISI (1500ms). Thus, both tasks likely tapped categorical / phonemic processing. These task conditions were chosen because they were most similar to the conditions used in real life communicative situations and were the conditions most likely to assess to what degree learners could and would use Spanish-specific (target-like) phonetic categories while in a phonemic/categorical processing mode (Werker & Logan, 1985).

3.1.3 Instruction and Spanish L2 Perception

§1.3 presented a review of studies investigating L2 perception. Most of those studies targeted English as an L2 and examined L2-relevant contrastive phones, with the English /ʌ/-/ʊ/ contrast being the most widely investigated. Many studies provided learners with intensive and extensive auditory training in a laboratory setting. It was argued that there is a need for more research in L2s that are not English, phones that are not contrastive, and instructional techniques that are classroom-based. The present study analyzed the extent to which instruction in Spanish phonetics and phonology improved the ability of English speaking Spanish learners to perceive Spanish phones in more target-like ways.
The experimental instruction in the present study was much more explicit than the type of training participants received in the laboratory-based studies. It is an empirical question whether more explicit or less explicit instruction is most beneficial for improving L2 perception. Explicit knowledge about L2 phonetics may serve as an attention-orienting device to help learners change their L1-informed automatic processing routines (Guion-Anderson, in press, p. 6). Archila-Suerte et al. (2011) argued that “while younger children use an implicit/procedural process to categorize L1 and L2 phonemes, older children and adult learners use explicit rules to learn L2 phonemes” (2011, p. 191), and so their acquisition of L2 phonology is necessarily mediated by explicit knowledge. Archila-Suerte et al.’s (2011) investigation of L2 perception recruited native speakers of Spanish (n=98) who began learning English either “early” (before age 5), “late” (after age 10) or somewhere in between. They found that the late learners and the learners with low English proficiency perceived a 4-way English vowel contrast less well than the early starters, particularly the English vowels that were located within Spanish vowel categories. Though their study was not an investigation of the role of instruction, their results suggested that even with a relatively early start and relatively high L2 proficiency attained, learners do not necessarily acquire native-like phonetic categories implicitly through consistent and extensive contact with the L2, so instruction may be necessary.

On the other hand, the SLM suggests that L2 input is of primary importance. The SLM does not address whether instruction could be helpful, and it certainly does not characterize instruction of any type (explicit or implicit) as strictly necessary. Rather, the process of acquiring L2 speech is contingent on 1) speech learning mechanisms that remain intact throughout the lifespan and 2) lots of exposure to native-like speech (Flege,
2003a). Presumably, as learners gain more experience hearing the L2, they acquire a denser network of phonetic exemplars and form more L2-like phonetic categories, thus relying less and less on L1 categories to filter L2 speech.

Munro (2008) suggested that one’s ability to learn is affected by one’s aptitude for speech sounds. Munro focused on memory, particularly phonological short term memory, which is beyond the scope of the present discussion. However, Munro provided a review of the numerous ways other researchers have characterized variability in learners’ processing of L2 speech (Lambert, 1977; Thompson, 1991; Markham, 1997). Lambert (1977) distinguished between “code users,” who prefer more top-down or categorical perception, and “code formers,” who tend to use more bottom-up or continuous perception and are thus predisposed to create better L2 speech representations. Thompson (1991) used the term “ear-mindedness,” or tendency to rely on L1 phonemic knowledge or raw phonetic/acoustic information when perceiving novel speech sounds. Similarly, Markham (1997) claimed that some individuals prefer “middle-up” processing, which is not truly bottom-up auditory perception but rather more focused on phonetic details and less reliant on L1 phonemic categories.

It has been suggested that differences in individuals’ processing tendencies might predict their ability to learn from particular types of instruction. Helmke and Wu (1980) found that pronunciation drills were beneficial for “high auditory discriminators,” but “low auditory discriminators” did not benefit from the pronunciation drills. Bradlow et al. (1997) found that “high performers” were able to use perceptual training exercises to their benefit and “fine-tune” the L2 contrasts that they already perceived relatively well.
In contrast, the “poor performers” did not seem to benefit from the perceptual training exercises at all in terms of creating new categories for the contrastive L2 phones.

Any demonstrated effect of instruction on L2 speech perception will most certainly depend on the particular target phones in question as well. As Munro (2008) went on to explain in his review, researchers have described some L2 phones as “fragile” because they required intensive, consistent, and repeated training to improve learners’ perception of the phones (e.g., Polka, 1991; Werker et al., 1981), whereas other L2 phones are more “robust,” requiring virtually no instruction (e.g., Logan et al., 1991; Tees & Werker, 1984), generally because of their inherent perceptual saliency. Some L2 phones are simply more difficult to perceive, with or without instruction.

Nearly all of the research published on traditional pronunciation or phonetics instruction in L2 Spanish, as reviewed in §2.1.1, has reported changes in Spanish learners’ production rather than perception. One notable exception was the study of Ausín and Sutton (2010), which presented Spanish learners with a variety of consonantal phones in CVC syllables. Ausín and Sutton (2010) reported that learners were better able to detect nonnative accents in Spanish after receiving instruction. It should be noted, however, that learners did not improve their perception of five out of eleven consonantal phones, so the results were actually quite mixed. Very little research within the instructed learning context has investigated both perception and production with the same learners. Zampini (1998) investigated English speaking Spanish learners’ perception and production of the Spanish /p/-/b/ contrast. Zampini reported that learners’ perceptual boundaries for the VOT of /p/ and /b/ did not correlate with the VOT they produced for those phones in most of the target items (e.g., pace / base; peso / beso). However, much
more research is clearly needed to understand the impact that phonetics instruction has (or does not have) on the perception of instructed L2 learners.

3.1.4 Research Questions and Hypotheses

The present study attempted to answer several of the questions that had not been addressed by previous studies regarding learners’ perception of the target phones. The research questions that motivated the present analysis were: 1. Does instruction in L2 phonetics and phonology improve learners’ ability to perceive the acoustic differences between L2 phones and their analogous L1 phones?, 2. Does instruction in L2 phonetics and phonology improve learners’ accuracy in identifying L2 phones in the speech stream?, and 3. Does the effectiveness of instruction depend on learners’ L2 curricular level? The first research question was addressed with an AX discrimination task. The second research question was addressed with an identification task.

The dearth of relevant prior research precluded forming strong hypotheses for the research questions. Though studies reported generally positive effects of training on perception, the conditions (intensive perceptual training) and linguistic targets (L2 contrastive phones) were different from those of the current study, which provided learners with explicit phonetics instruction in an attempt to improve learners’ ability to distinguish between Spanish phones and analogous English phones.

3.2 Methods
The context, participants, target phones, instructional treatments and experimental design were all the same as those described in §2.2. This section will provide methodological details only of the perception tests.

3.2.1 Discrimination Test

3.2.1.1 Discrimination Test Materials

The discrimination test was an AX discrimination test much like those used in other studies to evaluate perception of relevant L2 contrasts (e.g., Polka, 1992). However, following the prediction of the SLM that in order to acquire L2 sounds learners must first distinguish between L2 phones and their analogous L1 counterparts, this AX discrimination test paired the Spanish target phones with their analogous English phones, in mono- and disyllabic utterances with a variety of Spanish vowels. The stimuli consisted of sound recordings made by a male native Spanish speaker from Argentina who was trained in Spanish phonetics and phonology and had near-native proficiency in English. The recordings were made using the audio editing software SoundTrackPro, at a 16-bit sampling rate of 48 KHz in a sound-proof booth.

The test items of interest for the present analysis paired the Spanish target phones with English-accented renditions of those phones and included 5 items each of the voiceless stops /p, t, k/ in initial position followed by all 5 Spanish vowels, the approximants [β, δ, γ] in intervocalic position using five different vowel pairs, and the

---

19 Argentinean dialects exhibit several “non-standard” phonological variations, including elided /s/, the existence of a post-alveolar fricative /ʃ/ rather than a palatal lateral approximant /ʎ/, and assibilation of /ɾ/ (Alvar, 1996). However, this speaker’s dialect did not exhibit any non-standard realizations of the target phones [p, t, k, β, δ, γ, r, r].
rhotic phones in initial, medial, and final positions. So, for instance, an item might pair 
\[p^h e\] with \[pe\] or \[e\beta e\] with \[ebe\]. A pilot study showed that perceiving the Spanish

tap/trill distinction in some positions was also difficult for beginning learners, and so that
pair was also included. However, while 5 items were included contrasting \([r]\) with \([r]\),
only 3 items each contrasting \([r]\) with \([\text{]}\) and \([r]\) with \([\text{]}\) were included, because these
contrasts were found to be quite easy for participants to distinguish in pilot testing. Thus,
in total, the discrimination test included 41 items in which target phones were contrasted
with analogous English counterparts. Only these items will be discussed in the following
analysis. Table 3.1 lists the relevant pairs of phones tested and the phonological
environments in which they were presented.

<table>
<thead>
<tr>
<th>Phone Pairs</th>
<th>Phonological Environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>([p]/[p^h]), ([t]/[t^h]), ([k]/[k^h])</td>
<td>([_a], [_e], [_i], [_o], [_u])</td>
</tr>
<tr>
<td>([\beta]/[b], [\delta]/[d], [\gamma]/[g])</td>
<td>([_a], [e], [u_o], [i_e], [o_i])</td>
</tr>
<tr>
<td>([r]/[\text{]}])</td>
<td>([o_o], [i_], [_a])</td>
</tr>
<tr>
<td>([r]/[\text{]}])</td>
<td>([a_a], [e_], [u_], [e_])</td>
</tr>
<tr>
<td>([r]/[r])</td>
<td>([a_a], [i_i], [o_o], [e_], [_c])</td>
</tr>
</tbody>
</table>

Other test items used as distracters and not of interest in the following analysis
included target phones paired in identical pairs (e.g., \([e\delta e] [e\delta e]\)) and non-target Spanish
phones in identical pairs (e.g., \([fe] [fe]\)) and AX pairs (e.g., \([fe] [fe]\)). A full list of the
85 item pairs that were included in the discrimination test is provided in Appendix 3.1.
These same item pairs, scrambled into different inter-pair and intra-pair orders, were used
for the pretest, immediate posttest, and delayed posttest. Immediate posttests were shorter
versions that included only the target phones that were treated in the instructional module
immediately preceding the test, plus distracters.
A single acoustic token was used for each of the syllable types. All tokens representing [pa], for instance, were acoustically identical. All were the same recording. While recording the stimuli, both the researcher and the native Spanish speaker took great care that the only audible differences in the AX stimuli pairings were the target segments. The volume, pitch, duration, and quality of the surrounding vowels were kept as similar as possible between the two stimuli in the AX pairings. All these parameters sounded identical to the ear of both the researcher and the native Spanish speaker, who was trained in phonetics. An acoustic analysis of the stimuli was subsequently performed using Praat. The results of the analysis, presented in Appendix 3.2 and Appendix 3.3, suggested that the two tokens of each AX pairing were very similar in every way except the relevant target phones.

### 3.2.1.2 Discrimination Test Administration Procedures

The order in which learners took the three experimental tests (production, discrimination, and identification) was counterbalanced. While completing all instructional modules and testing, learners were seated at individual PC computer stations in a language laboratory, wearing noise-canceling headphones and listening to sound recordings at a volume that was comfortable for them. The instructions and sound recordings of test items for the discrimination and identification tests were in .WAV files, which the participants opened with QuickTime Player. The instructions stated, “For each item number, you will hear two recordings. If the recordings sound exactly the same to you, choose ‘same.’ If you hear any difference at all between the two recordings, regardless of how slight that difference is, choose ‘different.’” Learners highlighted or
underlined their chosen response of either “same” or “different” for each item on a paper answer sheet, after first completing two practice items with feedback. So, for instance, a learner heard a pair such as [pʰe], [pe] and was instructed to mark either “same” (incorrect response) or “different” (correct response). Each item pairing was heard only once. Learners were instructed not to rewind. After half of the items were presented, the test paused for a short break and repeated the instructions. Additionally, for each item, learners rated how confident they were in their choice, on a scale of 1 – 4.

The average duration of a stimulus exemplar was 1 second, with a 1500ms inter-stimulus interval (ISI). As explained in §3.1.2, this long ISI should have precluded participants from using only their acoustic store to compare the stimuli, and therefore acoustically identical tokens could be used without compromising the task objective. Acoustically identical tokens were ideal for this study because they allowed even the phonetically untrained participants to behave in a task-appropriate way under the discrimination test conditions. A pilot study indicated that participants misinterpreted instructions such as “If you hear the same sound” and consequently underreported the number of phonetic differences they were able to discern. On the other hand, participants were unlikely to misinterpret the instructions “If you hear any difference at all between the two recordings, choose ‘different.’” This choice of language in the task instructions necessitated the use of acoustically identical tokens for each target syllable.

3.2.1.3 Discrimination Test Scoring Procedures

For the discrimination test, responses were either correct or incorrect and so were assigned 1 point or 0 points.
3.2.2 Identification Test

3.2.2.1 Identification Test Materials

The identification test was designed to assess learners’ ability to identify the target Spanish phones in natural speech. Identification tests typically present learners with minimal pairs (e.g., English road/load, Spanish paca/baca) and ask learners to identify which of the lexical items they perceived. Those tasks work well in studies investigating sounds that contrast in the L2 but not in, or not in the same way as, the L1. However, the target phones in this study arise in phonologically determined, predictable environments and cannot be used to construct minimal pairs, with the exception of Spanish /p/ vs. /b/ and /t/ vs. /r/ in word-medial, intervocalic position. Therefore this identification test presented learners with Spanish-like non-words (e.g., huirrin) and instructed them to type what they heard in standard Spanish orthography. The use of standard orthography is not ideal for a perception study, as it introduces the intervening variable of learners’ knowledge of Spanish grapheme-phoneme correspondences, but phonetic symbols could not be used with the control group, which received no instruction in phonetics (-PI).

The test items were recorded using the same native Spanish speaker and same recording procedures as the discrimination test. The test items included five tokens of each target phone [p, t, k, β, δ, ñ, r, r], in the same phonological environments used for the discrimination test. A full set of the test items is presented in Appendix 3.4, with the
target phones bolded. Similar, but not identical, items were used in each version of the test, in an attempt to avoid test learning effects.

3.2.2.2 Identification Test Administration Procedures

For the identification test, the stimuli consisted of recordings of bi-syllabic and tri-syllabic nonce words. The instructions stated, “You are going to hear some nonsense words in Spanish. Type exactly what you hear, using Spanish spelling. Sometimes there is more than one way to spell the nonsense word, so just do the best you can. Don’t worry about accent marks.” Each nonce word was repeated once. Learners were instructed not to rewind. Learners typed their responses in a .DOC file, after first completing two practice items. Throughout all sessions of the study, the researcher and multiple lab assistants monitored the learners to ensure that they were following instructions.

3.2.2.3 Identification Test Scoring Procedures

For the identification test, learners’ entire responses were recorded, but only the graphemes corresponding to target phones were scored and analyzed. In an attempt to separate spelling errors from perception errors, 1 point was assigned to spellings that were non-standard yet indicated that the learner had likely perceived the target phone correctly. So, for example, \( \tilde{k} \) was assigned one point as a non-standard orthographic representation of the target phone \([k]\). Table 3.2 illustrates the scoring scheme, with sample items and examples of correct and incorrect responses provided by learners.
### Table 3.2 Scoring of Identification Test Items

<table>
<thead>
<tr>
<th>Target Phone</th>
<th>Standard Orthography</th>
<th>Accepted Grapheme(s) (Assigned 1 point)</th>
<th>Sample Test Item</th>
<th>Examples of Correct Responses</th>
<th>Examples of Incorrect Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>[p]</td>
<td>p</td>
<td>[p]</td>
<td>[pe.'su.do]</td>
<td>pesudo</td>
<td>besudo</td>
</tr>
<tr>
<td>[t]</td>
<td>t</td>
<td>[t]</td>
<td>['te.ti]</td>
<td>teti</td>
<td>deti</td>
</tr>
<tr>
<td>[k]</td>
<td>c (before a, o, u), gu (before e, i)</td>
<td>k, c, a, gu</td>
<td>[ke.'γem.bo]</td>
<td>queguembo, ceguembo, keguembo, qeguembo</td>
<td>teguembo</td>
</tr>
<tr>
<td>[β]</td>
<td>b, v</td>
<td>b, v</td>
<td>[o.'βi.'to]</td>
<td>obitó, ovitó</td>
<td>olbitó, olyitó</td>
</tr>
<tr>
<td>[ð]</td>
<td>d</td>
<td>d, th</td>
<td>[o.'ði.ta]</td>
<td>odita, othita</td>
<td>ovida, orvida</td>
</tr>
<tr>
<td>[ɣ]</td>
<td>g (before a, o, u), gu (before e, i)</td>
<td>g, gu</td>
<td>['tre. γe]</td>
<td>trege, tregue</td>
<td>trele, trede</td>
</tr>
<tr>
<td>[r]</td>
<td>r</td>
<td>r, rr</td>
<td>['sa.'βa.ra]</td>
<td>sábara</td>
<td>sabada</td>
</tr>
<tr>
<td>[r]</td>
<td>r (intervocalic)</td>
<td>r, rr</td>
<td>[wi.'rin]</td>
<td>huirrin</td>
<td>wirin</td>
</tr>
</tbody>
</table>

For both the discrimination and the identification tests, scores on the pretest, immediate posttest, and delayed posttest were calculated by adding the points received by each learner for all the items relevant to each target phone on each test. Finally, overall scores were calculated for each test time by adding the scores on all target phones. In the case of the immediate posttest, this overall score encompassed scores from sessions 2 and 3, since learners completed half of the instructional modules, followed by module-specific discrimination and identification posttests, during sessions 2 and 3.
3.3 Results

3.3.1 Discrimination Test

3.3.1.1 Reliability

Discrimination test scores were reliable. The Kuder-Richardson 20 coefficient for the 41 pretest items of interest here (target phones in AX item pairings) was 0.85, indicating that the items were internally consistent and that it was reasonable to report their aggregate score as an index of discrimination ability. A paired sample T-test compared the first half of the 85-item pretest with the second half and indicated that there was no significant difference ($t(79)=.27, p=.79$), suggesting that the test was homogeneous. Finally, 3 items, chosen at random, were repeated twice on each test as a test/retest measure of stability. The Kuder-Richardson 20 coefficients found for these sets of three items each were reasonably high (.49 for 3 [ko]/[kʰo] items, .59 for 3 [iːb]/[ibe] items, and .85 for 3 [iɾi]/[iri] items) and their paired correlations were .30 or higher, suggesting that participants’ responses were relatively stable.

3.3.1.2 Descriptive Statistics

Learners’ scores for each of the target phones are presented in Table 3.3. The discrimination test was a forced-choice test with two possible responses, and thus a 50% accuracy rate represented a score that was at chance. On the pretest, learners’ accuracy was near chance on [k]/[kʰ] (43% accuracy), [β]/[b] (53%), and [ɣ]/[g] (49%). According to García-Lecumberri’s (1999) criteria for assessing perceptual difficulty, [k]/[kʰ], [β]/[b]
and [ɣ]/[ɡ] presented a great deal of difficulty for these learners (below 65% accuracy).

According to the same criteria, the pairs [t]/[tʰ] (76%), [ɾ]/[r] (72%), and [ð]/[d] (66%) presented some difficulty for these learners (65-80% accuracy). The pairs [p]/[pʰ] (87%), [ɾ]/[ɾ̚] (89%), and [r]/[ɾ̚] (99%) presented little difficulty for these learners (80-99% accuracy).

<table>
<thead>
<tr>
<th>Table 3.3 Average Score Per Phone on Discrimination Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Year</td>
</tr>
<tr>
<td>+PI (n=18)</td>
</tr>
<tr>
<td>-PI (n=20)</td>
</tr>
<tr>
<td>+PI (n=17)</td>
</tr>
<tr>
<td>-PI (n=16)</td>
</tr>
<tr>
<td>+PI (n=10)</td>
</tr>
<tr>
<td>-PI (n=9)</td>
</tr>
<tr>
<td>NS (n=10)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[p] - [pʰ]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(5 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[ɾ] - [ɾ̚]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(5 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[β] - [b]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(5 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[ɤ] - [d]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(5 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[γ] - [a]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(5 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[ɾ] - [ɾ̚]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(3 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>[ɾ̚] - [ɾ]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>(3 items)</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
</tr>
</tbody>
</table>

For sake of comparison, learners were generally more accurate at discriminating vocalic phones, which were included as distractor items (88%), and they detected identical pairs (in the form XX) with near perfect accuracy (97%). The pretest scores suggested that learners’ discrimination of [k]/[kʰ], [β]/[b], and [ɣ]/[ɡ] was quite poor, and they also had
room for improvement on their discrimination of [t]/[tʰ], [ɾ]/[ɾ], and [ð]/[d], yet they could discriminate [p]/[pʰ], [ɾ]/[ɾ], and [r]/[ɾ] with ease, even before treatment. English [ɾ] differs from Spanish [ɾ] and [r] across several articulatory and acoustic parameters. Consequently, there are more perceivable differences in the acoustic signals of these phones than the other pairs tested, so learners’ ease in discriminating [ɾ]/[ɾ] and [r]/[ɾ] is not surprising. Learners’ high scores on the [p]/[pʰ] contrast likely has to do with their phonological distribution in English. The short voice onset timing of Spanish [p] is comparable to the voice onset timing of English [b], so learners may have perceived the /p/-/pʰ/ contrast as a /b/-/p/ contrast (see Zampini, 1998, for discussion of perception of /b/-/p/). More evidence for this interpretation is presented in discussion of the identification test.

The discrimination test scores of the Spanish native speakers (NSs) are included in Table 3.3 for reference, but they should not be used as a benchmark against which to compare Spanish learners’ performance. If the discrimination test had been a test of Spanish contrasts (e.g., Spanish /p/-/b/) then one would expect the NSs to perform at ceiling levels. However, the discrimination test was a bilingual test that contrasted Spanish phones with analogous English phones. Therefore one would not expect NS of Spanish necessarily to be accurate on this discrimination test. The SLM would predict that a NS of Spanish could perform well on this discrimination test if she had enough exposure to English to begin forming target-like phonetic categories for English. Anecdotally, the two NS participants who had the most English exposure and the most target-like English accents, in the researcher’s opinion, achieved the highest scores on the
discrimination test (100% and 88%, respectively). The NSs were not tested for proficiency or accent, so this is merely anecdotal evidence.

3.3.1.3 Distribution of Scores and Elimination of /p, r, r/ Items

A Shapiro-Wilk test indicated that the discrimination post and delayed posttest scores were not normally distributed ($p < .01$). The scores were also negatively skewed as a result of the near ceiling scores of particular items. An item analysis was performed to better understand these data. For each item at each test time, a point-biserial correlation was calculated, which is the correlation between the (dichotomous, correct/incorrect) score of the item and the (continuous) aggregate test score. Point-biserial correlations range from -1 to 1. Correlations of less than .25 are generally considered non-discriminatory (Larson-Hall, 2010). The correlations for each item at each of the three test times were added together, and those items whose aggregate correlation was below .75 were deemed non-discriminatory. With only four exceptions, all the items found to be non-discriminatory pertained to the target phones /p, r, r/. All the items for these non-discriminating phones were subsequently eliminated from the analysis, thereby culling the discrimination test to 25 items, or 5 items per target phone [t, k, β, δ, γ].

The data from the culled test were then reanalyzed for normality. A Shapiro-Wilk test indicated that the discrimination scores for [t, k, β, δ, γ] were normally distributed ($p > .05$) as a whole and also when grouped by test time and by learners’ matriculation level. The only exception was that of the first year learners’ discrimination posttest, which did not pass the Shapiro-Wilk test of normality ($p = .01$). The discrimination scores for [t, k, β, δ, γ] generally did not display skew or kurtosis, as measured by dividing each statistic
by its standard error. There was just one exception: the first year learners’ discrimination pretest scores were slightly negatively skewed (skew of -2.21). It was decided that the distribution of the data was normal enough to satisfy the assumptions of an ANOVA.

3.3.1.4 RMANOVA and Subsequent Contrasts

Learners’ scores on the 25-item discrimination test were submitted to a RM ANOVA. The within-groups factors were time of test (pretest, immediate posttest, and delayed posttest) and the between-groups factors were instructional condition (+PI and -PI) and matriculation level (first, second and third year). The results of the RMANOVA are presented in Table 3.4. There was a main effect of time, $F(1.75, 138) = 21.54, p < .001, \eta^2_p = .21$, indicating that learners’ scores on average changed over time, after instruction. There was also a significant time by condition interaction effect, $F(1.75, 138) = 6.17, p < .001, \eta^2_p = .07$, indicating that the change in score over time differed depending on which instructional condition (+PI/-PI) the learner was in and that 7% of the variance in scores across time could be explained by instructional treatment condition. The interaction of time by level was not significant, nor was the interaction of time by condition by level, suggesting that prior language experience did not influence learners’ changes in scores, though the observed power was below 80% for these last two tests and so they were not conclusive. Figure 3.1 plots the significant time by condition interaction.
Table 3.4 RM ANOVA of Discrimination Test: Aggregate of \([t, k, \beta, \delta, \gamma]\)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>(r)</th>
<th>(\eta^2)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>80.36</td>
<td>2.00</td>
<td>40.18</td>
<td>0.87</td>
<td>0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>Level</td>
<td>1.47</td>
<td>1.00</td>
<td>1.47</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>332.97</td>
<td>2.00</td>
<td>166.49</td>
<td>3.61**</td>
<td>0.08</td>
<td>0.65</td>
</tr>
<tr>
<td>Error</td>
<td>3645.42</td>
<td>79.00</td>
<td>46.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>429.07</td>
<td>1.75</td>
<td>244.89</td>
<td>21.54****</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>122.96</td>
<td>1.75</td>
<td>70.18</td>
<td>6.17****</td>
<td>0.07</td>
<td>0.85</td>
</tr>
<tr>
<td>Time X Level</td>
<td>47.78</td>
<td>3.50</td>
<td>13.63</td>
<td>1.20</td>
<td>0.03</td>
<td>0.34</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>52.63</td>
<td>3.50</td>
<td>15.02</td>
<td>1.32</td>
<td>0.03</td>
<td>0.38</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>1573.70</td>
<td>138.41</td>
<td>11.37</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at \(p < .10, **p < .05, ***p < .01, ****p < .001\)

With a Greenhouse-Geisser Correction

Table 3.5 Pairwise Comparisons of Discrimination Test: \([t, k, \beta, \delta, \gamma]\)

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>MD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower</td>
<td>upper</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest - Posttest</td>
<td>-3.21****</td>
<td>0.41</td>
<td>-4.21 - 2.21</td>
</tr>
<tr>
<td>Pretest - Delayed Posttest</td>
<td>-0.75</td>
<td>0.58</td>
<td>-2.16 - 0.67</td>
</tr>
<tr>
<td>Posttest - Delayed Posttest</td>
<td>2.46****</td>
<td>0.53</td>
<td>1.16 - 3.77</td>
</tr>
</tbody>
</table>

Significant at \(p < .10, **p < .05, ***p < .01, ****p < .001\)

Adjustment for Multiple Comparisons: Bonferroni

Figure 3.1 suggests that the +PI group had an advantage in retention of learning from posttest to delayed posttest and in overall gains from pretest to delayed posttest, as
compared with the -PI. Independent samples T-tests controlled for Type 1 error with the Bonferroni correction confirmed that while the gain from pretest to posttest was not significantly higher for the +PI than the -PI, \( t(92) = -0.23, p = .82 \), the retention from posttest to delayed posttest was significantly higher for the +PI than the -PI, \( t(85) = 2.64, p = .01, d = 0.56 \), and the gain from pretest to delayed posttest was significantly higher for the +PI than the -PI, \( t(84) = 2.51, p = .01, d = 0.54 \). In sum, then, for the 25-item discrimination test, instructional condition was a significant factor, whereas matriculation level was not. Both instructional condition groups made significant improvement immediately after instruction, but the +PI retained significantly more three weeks later at the delayed posttest.

A RMANOVA was also used to compare scores across individual phones. The within-groups factors were time of test (pretest, immediate posttest, and delayed posttest) and phone ([t, k, β, δ, γ]) and the between-groups factors were instructional condition (+PI and -PI) and matriculation level (first, second and third year). The results of the RMANOVA are presented in Table 3.6. There was a significant main effect for time, a significant main effect for phone, a significant condition by level interaction and a significant time by phone interaction. The latter interaction (phone type by time of test) was of greatest interest to the present analysis, because it was the \textit{change} in particular phones over time following instruction that was the main concern of the present study. Figure 3.2 plots the phone by time interaction.
This phone type by time of test interaction was further investigated by comparing gain scores on the target phones. The target phones could be ordered in size of gains from pretest to posttest as follows: \(/k/ (x = 0.94, s = 1.21) > [\delta] (x = 0.83, s = 1.20) > [\beta] (x = 0.72, s = 1.42) > [\gamma] (x = 0.51, s = 1.32) > /t/ (x = 0.40, s = 1.18). Only one
difference reached significance ($p < .05$) however, which was that the gain from
discrimination pretest to posttest for the phone /k/ was significantly greater than that of
/t/. The target phones could be ordered in size of gains from pretest to delayed posttest as
follows: /k/ ($\bar{x} = 0.39, s = 1.20$) > [ð] ($\bar{x} = 0.31, s = 1.21$) > [ɣ] ($\bar{x} = 0.24, s = 1.33$) >
[β] ($\bar{x} = -0.06, s = 1.33$) > /t/ ($\bar{x} = -0.15, s = 1.19$). The only difference that reached
significance however was that the “gain” from discrimination pretest to posttest for the
phone /t/ (a negative value, and so represents a loss) was significantly lower than that of
/k/, [ð], and [ɣ].

To determine the effects of time, condition, and level on individual phones,
RMANOVAs were used to analyze scores for each individual phone, with a within-
groups factor of test time (pretest, immediate posttest, and delayed posttest) and between-
groups factors of instructional condition (+PI and -PI) and matriculation level (first,
second and third year). As will be discussed below, there was a paucity of significant
differences other than the main effect of time. Therefore, the full RMANOVA tables will
not be presented. Furthermore, only the significant main effects and interaction effects
related to the variable of time will be discussed, as it was the change in scores over time
following instruction that was the main concern of the present study.

The only effects that reached significance were: a main effect of time for [k]/[kʰ],
[ð]/[d], and [ɣ]/[g] (all $F \geq 4.67, p \leq .01, \eta_p^2 \geq .06$) and time by instructional condition
interaction for [β]/[b] ($F(2, 160) = 15.53, p \leq .001, \eta_p^2 = .16$). The +PI improved its
discrimination of [β]/[b] from pretest to posttest and from pretest to delayed posttest more
than the -PI. The time by condition interaction did not reach significance for any other
phone (all $F \leq 1.61, p \geq .10$). The time by matriculation level interaction was not
significant for any phone (all $F \leq 1.25, p \geq .17$), nor was the time by instructional condition by matriculation level interaction (all $F \leq 2.55, p \geq .09$). The observed power was too low to draw conclusions about interaction effects for most phones. The raw scores suggested that learners’ performance on individual phones was similar to their performance on the test overall (the +PI had an advantage over the -PI, regardless of matriculation level), but the groups difference did not reach significance for any phones pairs other than $[\beta]/[b]$.

For each item of the discrimination test, participants also rated how confident they were in their choice, on a scale of 1 – 4. The confidence report data from the 25-item test were submitted to a RM ANOVA, and the only significant effect found was the main effect for time, $F(1.72, 138) = 6.50, p = .002, \eta_p^2 = .08$. Both groups grew more confident over time. Paired samples T-tests confirmed that +PI group’s average confidence rating was significantly higher on the delayed posttest (mean 3.27, SD 0.56) than the pretest (mean 3.10, SD 0.54), $t(45) = 3.60, p = .001, d = .31$. The –PI group’s average confidence rating was significantly higher on the delayed posttest (mean 3.29, SD 0.48) than the pretest (mean 3.10, SD 0.49), $t(38) = 2.36, p = .02, d = .39$, and its average confidence rating was significantly higher on the immediate posttest (mean 3.28, SD 0.43) than the pretest, $t(44) = 2.81, p = .007, d = .35$. Thus, the learners in the +PI group gained no more confidence in their ability to discriminate items than the -PI. Both groups simply grew more confident in their discrimination abilities over time, possibly because they got more used to the testing format.

It is interesting to note that learners in the +PI felt significantly more confident in the delayed posttest than the pretest, and their accuracy did indeed improve significantly
from the pretest to the delayed posttest. In contrast, -PI learners felt significantly more confident in the delayed posttest than the pretest, yet their accuracy did not improve significantly from the pretest to the delayed posttest. However, a series of correlation analyses did not find any significant correlations between average confidence reports and raw scores or gain scores on the 25-item discrimination test. Therefore, learners in this study did not demonstrate accurate self-assessment of their ability to discriminate between L2 and analogous L1 phones.

3.3.2 Identification Test

3.3.2.1 Reliability

Identification test scores were reliable. The Kuder-Richardson 20 coefficient for the pretest items was 0.82, indicating that the items were internally consistent and that it was reasonable to report their aggregate score as an index of identification ability. As a test/retest measure of stability, 3 items chosen at random were repeated twice on each test. The Kuder-Richardson 20 coefficients found for these sets of three items each were all at or above .70 with one exception (‘cubor’, coefficient of .64), suggesting that participants’ responses were stable.

3.3.2.2 Descriptive Statistics

As was the case with the discrimination test, learners’ accuracy on the identification test varied across the target phones. On the pretest, learners identified [t, k, \(\beta\)] with relatively high accuracy (84%, 97%, and 93%, respectively). They identified [\(\delta\)]
with lower accuracy (59%), and their accuracy on the other phones was somewhere in between (77% for [p], 81 % for [y], 72 % for [ɾ], and 70% for [r]). Table 3.7 presents learners’ identification test scores and standard deviations for each target phone, separated by instructional condition and matriculation level.

The identification test tapped the ability to identify (and transcribe) phones in Spanish-like words. Unlike the discrimination test, it did not involve both Spanish and English phonology, and therefore NSs’ performance on the identification test could be used as meaningful benchmarks with which to compare learners’ scores. NSs performed with near perfect accuracy on most phones. Learners at all matriculation levels were significantly less accurate than NSs on the aggregate (all 8 phones) identification test.

<table>
<thead>
<tr>
<th>Table 3.7 Average Score Per Phone on Identification Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Year</td>
</tr>
<tr>
<td>+PI (n=19)</td>
</tr>
<tr>
<td>[p]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[t]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[k]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[β]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[β]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[η]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[r]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
<tr>
<td>[r]</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Delayed</td>
</tr>
</tbody>
</table>
both at the pretest and the delayed posttest. Learners at all matriculation levels were also significantly less accurate than NSs on the rhotic phones and [ð]. First year learners were significantly less accurate than NSs on all phones except /k/ and [ɣ]. Second year learners were less accurate than NSs on several phones, but not /k/, [β], or [ɣ]. Third year learners did not perform significantly different than NSs on any individual phones except [ð], /r/, and /ɾ/. Thus the data suggested that learners improved their identification of most phones with experience; and were able to approximate NSs’ performance by the third year for the target phones /p/, /t/, /k/, [β], and [ɣ]; yet even third year learners identified [ð], [ɾ], and [r] less accurately than NSs.

3.3.2.3 Distribution of Scores

A Shapiro-Wilk test indicated that the identification test scores for were normally distributed \((p > .05)\) when analyzed separately at each test time and matriculation level. The only exception was first year learners’ identification posttest, which did not pass the Shapiro-Wilk test of normality \((p = .04)\). The identification scores generally did not display skew or kurtosis, as measured by dividing each statistic by its standard error, with three exceptions: first year learners’ identification posttest and delayed posttest scores were slightly negatively skewed (-2.23 and -2.20, respectively), as were second year learners’ identification posttest scores (-2.33). It was decided that the distribution of the data was normal enough to satisfy the assumptions of an ANOVA.
3.3.2.4 RM ANOVA and Subsequent Contrasts

Learners’ scores on the 40-item identification test were submitted to a RM ANOVA. The results are displayed in Table 3.8. There was a main effect of time, $F(1.83, 126) = 64.95, p < .001, \eta_p^2 = .49$, indicating that learners’ scores changed over time. There was also a significant time by condition interaction effect, $F(1.83, 126) = 13.81, p < .001, \eta_p^2 = .17$, indicating that the change over time differed depending on which instructional condition (+PI/-PI) the learner was in and that 17% of the variance in scores could be explained by instructional condition. The interaction of time by level was not significant, nor was the interaction of time by condition by level, suggesting that matriculation level did not influence learners’ scores, though the power was below 80% for these last two tests and so they were not conclusive. Pairwise comparisons indicated that learners’ scores were significantly different at all three test administration times. Figure 3.3 plots the significant time by instructional condition interaction.

![Table 3.8 RM ANOVA of Identification Test: Aggregate of 8 Phonemes](image-url)
The +PI appeared to have an advantage on the identification test in terms of gains from their pretest to posttest. Paired samples T-tests with a Bonferroni correction indicated that the +PI group’s scores were significantly higher on the posttest ($\bar{x} = 35.56, s = 2.93$) than the pretest ($\bar{x} = 30.78, s = 3.89$), $t(40) = 10.55, p < .001, d = 1.4$, and significantly higher on the delayed posttest ($\bar{x} = 33.63, s = 3.00$) than the pretest, $t(40) = 7.57, p < .001, d = 0.83$. The –PI group’s scores were significantly higher on the posttest ($\bar{x} = 34.74, s = 3.41$) than the pretest ($\bar{x} = 32.85, s = 3.32$), $t(34) = 3.98, p < .001, d = 0.56$, and significantly higher on the delayed posttest (mean 34.97, SD 2.60) than the pretest, $t(34) = 5.21, p < .001, d = 0.72$. Independent samples T-tests confirmed that the gains from pretest to posttest was significantly higher for the +PI group than the –PI group,
\( t(73) = 4.40, p < .001, d = 1.02 \). However, the gains from pretest to delayed posttest were not significantly higher for the +PI group than the -PI group, \( t(73) = 1.33, p = .19 \). In sum, then, for the identification test, instructional condition was a significant factor, whereas matriculation level was not. Both instructional condition groups made significant improvement immediately after instruction and retained significant gains three weeks later in the delayed posttest. The +PI group made greater improvement than the -PI group immediately after instruction, but not three weeks later.

A RMANOVA was also used to compare scores across individual phones. The within-groups factors were time of test (pretest, immediate posttest, and delayed posttest) and target phone and the between-groups factors were instructional condition (+PI and -PI) and matriculation level (first, second and third year). The results of the RMANOVA are presented in Table 3.10. There was a significant main effect for time, a significant main effect for phone, and several significant interactions: condition by level, time by condition, phone by level, and time by phone. The phone type by time of test was of greatest interest to the present analysis, because it was the change in particular phones’ scores over time following instruction that was the main concern of the present study. Figure 3.4 plots the phone by time interaction.
This phone type by time of test interaction was further investigated by comparing gain scores on the target phones. The target phones could be ordered in size of gains from pretest to posttest as follows: [δ] (\( \bar{x} = 1.17, s = 1.19 \)) > /p/ (\( \bar{x} = 0.61, s = 1.13 \)) > [γ] (\( \bar{x} = 0.46, s = 0.81 \)) > /r/ (\( \bar{x} = 0.32, s = 1.01 \)) > /r/ (\( \bar{x} = 0.23, s = 0.97 \)) > [β] (\( \bar{x} = 0.21, s = 0.64 \)) > /ɾ/ (\( \bar{x} = 0.17, s = 0.97 \)) > /k/ (\( \bar{x} = 0.13, s = 0.34 \)). Several differences reached significance (\( p < .05 \)). Learners’ gains in identification from pre to posttest were
significantly greater for [ð] than all other phones except /p/, and their gains for /k/ were lower than for [ð], /p/, and [ɣ]. The target phones could be ordered in size of gains from pretest to delayed posttest as follows: /t/ (\( \bar{x} = 0.72, s = 0.93 \)) > [ð] (\( \bar{x} = 0.48, s = 0.91 \)) > /r/ (\( \bar{x} = 0.39, s = 0.64 \)) > [ɣ] (\( \bar{x} = 0.38, s = 0.64 \)) > /p/ (\( \bar{x} = 0.30, s = 0.80 \)) > /t/ (\( \bar{x} = 0.15, s = 0.90 \)) > /k/ (\( \bar{x} = 0.02, s = 0.41 \)) > [β] (\( \bar{x} = 0.01, s = 0.57 \)). Learners’ gains in identification from pre to delayed posttest were significantly greater for /t/ than /k/, [β] and /r/. Their gains for /k/ were significantly lower than [ð], [ɣ] and /r/. Their gains for [β] were significantly lower than [ð], [ɣ] and /r/. Though the fact that accuracy was already quite high in the pretest for several phones made this comparison of gain scores perhaps less illuminating, the analysis suggested that after instruction, learners improved their identification of [ð] more than other phones, while their accuracy in identifying /k/ changed very little. Three weeks after instruction, learners improved their identification of /t/ more than most phones, whereas their accuracy in identifying /k/ and [β] remained virtually unchanged.

To determine the effects of time, condition, and level on individual phones, RMANOVAs were used to analyze scores each individual phone, with a within-groups factor of test time (pretest, immediate posttest, and delayed posttest) and between-groups factors of instructional condition (+PI and -PI) and matriculation level (first, second and third year). There was a main effect for time for all phones (all \( F \geq 4.53, p \leq .01, \eta_p^2 \geq .06 \)). There was a significant time by instructional condition interaction effect for three target phones: /p/ (\( F(1.80, 135) = 3.81, p = .03, \eta_p^2 = .05 \)); /t/ (\( F(2, 148) = 3.16, p = .05, \eta_p^2 = .04 \)) and /r/ (\( F(1.70, 124) = 4.08, p = .03, \eta_p^2 = .05 \)). Learners’ scores on these three phones (/p, t, r/) mirrored the trend witnessed on the aggregate test score. That is, the +PI
improved more than the -PI from pretest to posttest, but group differences evened out by three weeks later at the delayed posttest. Effect sizes were small (4-5%). The time by condition interaction did not reach significance for the other target phones (all $F \leq 2.03, p \geq .14$). The time by level interaction was not significant for any phone (all $F \leq 2.00, p \geq .09$) except [ɣ] ($F(4,156)= 2.50, p = .05, \eta^2_p = .06$), nor was the time by level by condition by level interaction (all $F \leq 1.80, p \geq .13$), suggesting that participants’ scores on the identification test were not explained by matriculation level.

The identification test also provided evidence about the nature of learners’ phonetic categories for Spanish L2 phones. Table 3.11 reports the types of errors elicited, i.e., the grapheme used erroneously with highest frequency to represent each target phone. Table 3.11 displays the percentage of incorrect responses attributable to this most common erroneous grapheme (for brevity, the proportion is displayed as an average across groups, levels, and test times), categorization of the error as likely due to misidentification of the phone or rather due to spelling, and the percentage of total responses that included the erroneous grapheme. The common errors elicited with /k/, [ɣ], word-final /r, r/, and intervocalic /r/ are likely to be due to incomplete acquisition of Spanish grapheme-phoneme correspondences, or simple spelling errors, rather than errors in identifying the target phone. Recall that such errors did not lower learners’ scores (refer to the rating scheme in Table 3.2).
However, participants made errors with other phones that were suggestive of particular stages in their interlanguage phonological development. The strongest case could be made with the phone /ɾ/, which learners frequently identified using the grapheme \(d\). This error suggested that learners associated intervocalic taps with alveolar stop consonants, which is their phonological relationship in English. Though participants were generally very accurate in their identification of /p, t/, whenever they made errors, they represented /p/ orthographically as \(b\) and they represented /t/ orthographically as \(d\). These errors suggested that the short VOT associated with Spanish /p, t/ had not been fully acquired by participants, and they were still in the process of aligning the VOT boundaries of their phonetic categories to Spanish targets. Though learners were also

<table>
<thead>
<tr>
<th>Target Phone</th>
<th>Most Frequent Erroneous Grapheme(s)</th>
<th>Interpretation of Error Type</th>
<th>As Proportion of Total Errors (Incorrect Responses Only)</th>
<th>Most Frequent Erroneous Grapheme(s) as % of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>[p]</td>
<td>(b)</td>
<td>phoneme ID</td>
<td>91.89%</td>
<td>14.93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.65%</td>
<td></td>
</tr>
<tr>
<td>[t]</td>
<td>(d)</td>
<td>phoneme ID</td>
<td>94.51</td>
<td>6.22%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.21%</td>
<td></td>
</tr>
<tr>
<td>[k]</td>
<td>(k, g) (before (o, a)); (q, qu) (before (e, i))</td>
<td>spelling</td>
<td>92.10%</td>
<td>14.29%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.23%</td>
<td></td>
</tr>
<tr>
<td>[β]</td>
<td>(lbs, lv)</td>
<td>phoneme ID</td>
<td>63.69%</td>
<td>1.63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.82%</td>
<td>1.05%</td>
</tr>
<tr>
<td>[ð]</td>
<td>(v)</td>
<td>phoneme ID</td>
<td>69.27%</td>
<td>24.63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.59%</td>
<td>11.56%</td>
</tr>
<tr>
<td>[ɣ]</td>
<td>(g) (before (e, i))</td>
<td>spelling</td>
<td>96.72%</td>
<td>47.67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>41.49%</td>
<td>35.09%</td>
</tr>
<tr>
<td>[ɾ]</td>
<td>(d) (intervocalic)</td>
<td>phoneme ID</td>
<td>83.13%</td>
<td>57.90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>34.99%</td>
<td>31.58%</td>
</tr>
<tr>
<td>[ɾ]</td>
<td>(rr) (word-final)</td>
<td>spelling</td>
<td>56.39%</td>
<td>14.18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.56%</td>
<td>4.38%</td>
</tr>
<tr>
<td>[ɾ]</td>
<td>(rr) (intervocalic)</td>
<td>spelling</td>
<td>87.14%</td>
<td>41.42%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>48.04%</td>
<td>55.26%</td>
</tr>
<tr>
<td>[ɾ]</td>
<td>(rr) (word-final)</td>
<td>spelling</td>
<td>65.49%</td>
<td>19.01%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.83%</td>
<td>4.38%</td>
</tr>
</tbody>
</table>
generally accurate in their identification of $[\beta]$, when they made errors, they often inserted a liquid ($l$), suggesting that at times they attributed the approximant acoustic cues of $[\beta]$ to another phone and perceived an illusory liquid before the target phone. Finally, learners were relatively less accurate at identifying $[\delta]$, and their use of the grapheme $v$ suggested that they identified a voiced fricative but did not yet associate the Spanish allophone $[\delta]$ with the phoneme /d/ and/or with the grapheme $d$. It was quite encouraging to note that for almost all the error types mentioned above, the proportion of the errors to participants’ responses steadily decreased across the three matriculation levels, suggesting that with increased exposure to Spanish, learners acquired more target-like phonetic categories and grapheme-phoneme correspondences. However, it should be noted that there was room for improvement, particularly with /ɾ/, which 32% of third year learners still identified as $d$.  

3.4 Discussion

3.4.1 Summary of Results

The first research question asked if instruction in L2 phonetics and phonology would improve learners’ ability to perceive the acoustic differences between L2 phones and their analogous L1 phones. An AX discrimination test paired each of the target Spanish (L2) phones with their analogous (L1) English phones in a variety of phonetic contexts. Learners’ discrimination of /p, r, r/ items was near ceiling levels on the pretest. Once these items were removed from the analysis, the +PI was found to have an
advantage over the -PI. The advantage was small, about half a standard deviation, interaction effect size of 7%, and the advantage was apparent not immediately following instruction but rather three weeks later in the delayed posttest. Though the instructional effect could not be confirmed for most phones when analyzed separately, the significant result for the aggregate test suggested that phonetics instruction did indeed improve learners’ ability to perceive the acoustic differences between some L2 phones and their analogous L1 phones.

The second research question asked if instruction in L2 phonetics and phonology would improve learners’ accuracy in identifying L2 phones in the speech stream. An identification test presented learners with nonce words in Spanish and asked them to type what they understood, using standard Spanish orthography. Both the +PI and the -PI made significant improvement immediately after instruction and retained significant gains three weeks later in the delayed posttest. The +PI, however, increased their scores from pretest to posttest significantly more than the -PI, with instructional condition accounting for 17% of the variance in scores. The same pattern was found for several target phones when analyzed separately ([p, t, r]). Thus instruction in L2 phonetics and phonology improved learners’ accuracy in identifying L2 phones immediately following instruction but was not necessarily more beneficial than exposure to Spanish phones through dictation exercises (the -PI condition) in the longer term.

The third research question asked if the effectiveness of instruction would depend on learners’ L2 matriculation level. For both the discrimination test and the identification test, using whole test scores as well as scores for individual phones, the time by condition by level interaction was not significant, suggesting that the effectiveness of instruction
did not depend on learners’ L2 curricular level. These tests of interaction suffered from low statistical power and could not be considered conclusive. In fact, the descriptive statistics indicated that there might have been some differences between matriculation levels. However, none of these differences was statistically significant.

3.4.2 Implications

The goal of the present study was to assess the effectiveness of L2 phonetics instruction in terms of improving learners’ ability to perceive L2 phones, specifically 8 consonantal Spanish phones that are problematic for L1 English speakers. It was argued that target-like perception is important for acquiring target-like pronunciation, yet most prior studies only reported on post-instructional improvement in learners’ production of these phones. The main finding was that even brief instruction in Spanish phonetics and phonology did afford learners a small but significant advantage, compared to a control, for discriminating L2 phones from their analogous L1 counterparts and also for identifying the L2 phones in the speech stream.

Two perception tests were employed in this study to target two types of perceptive skills. Though both tasks were “offline” and thus supposedly tapped categorical/phonemic processing as opposed to continuous/auditory processing, still one would not expect learners’ performance on the tasks to be identical. For instance, a learner might distinguish the acoustic differences between (Spanish) /p/ and (English) [pʰ] in a discrimination task, yet the same learner might identify Spanish /p/ as being a realization of the phoneme /b/, mistaking for example pollo (chicken) for bollo (bun). Being able to both discriminate [p] / [pʰ] and identify /p/ as /p/ is important for acquiring
target-like pronunciation according to the SLM. Learners who received phonetics instruction (the +PI) in this study improved their performance on both tasks relative to a control group.

Previously, researchers working with similar populations of learners reported that instruction had little effect on reducing learners’ VOT in their production of /p, t, k/ (González-Bueno, 1997; Lord, 2005), that instruction improved learners’ pronunciation of [ɣ] but not [β, δ] (González-Bueno, 1997), and that the Spanish trill (/ɾ/) continued to be problematic for learners to pronounce well through the most advanced levels (Face & Menke, 2010), though it responded better than other phones to explicit articulation instruction (Elliott, 1997). The present study complimented those studies by demonstrating that learners’ poor discrimination of L2 and analogous L1 phones, particularly [k]/[kʰ], [β]/[b], and [ɣ]/[g], may have been one factor limiting learners’ target-like production. The present study also indicated that phonetics instruction, even when in the form of brief instructional interventions added to an existing FL curriculum, could be beneficial for learners at all experience levels, helping them to better discriminate and identify problematic L2 phones.

Explicitly drawing learners’ attention to particular acoustic features of the L2 system appeared to be more expedient than merely exposing them to L2 sounds in the hopes that they would discover those relevant acoustic features. As was stated in the introduction, Guion-Anderson describes the learning process in this way: “classroom teachers provide explicit instruction about what to ‘listen for’ in order to allow students to orient their attention toward relevant aspects of the L2 speech and develop their ‘hooks’ into the L2. Once this is accomplished, learners can break out of their automatic
processing routines and take in the acoustic information relevant to the L2 for further processing” (in press, p. 6). The phonetics instruction in this study oriented learners to attend to relevant acoustic features, and they were able to use this knowledge to their benefit in the discrimination and identification tasks.

Importantly, the increased scores of the +PI were not likely attributable to mere task-specific practice. The reader will recall from the description of the instructional modules in §2.2.4 (and Appendix 2.8) that each module in the phonetics instruction (+PI) contained an activity labeled “discrimination exercise” that was actually an identification exercise. The identification exercise in the /p, t, k/ module instructed learners to listen to 10 syllables and indicate whether the speaker pronounced an English or Spanish consonant and reminded learners that “the primary difference you’ll here is whether a great deal of air is expelled or not.” Similarly, in the /b, d, g/, [β, ð, ɣ] module, learners listened to 10 syllables and indicated whether the speaker “pronounced a fricative consonant or an occlusive consonant” after being reminded that “that the primary difference you’ll hear is whether there is a complete stop to the airflow or not.” In the /ɾ, r/ module, learners listened to 10 syllables and indicated if the speaker “pronounced a Spanish tap [ɾ], a Spanish trill [r], or the English [ɹ].” Such identification activities are standard exercises included in most phonetics instruction.

Learners had to complete the exercises to 100% accuracy before proceeding to the next section. After submitting their answers, they saw a screen that displayed their incorrect responses. They were returned to the assessment, with their original selections still highlighted. They could listen again to the items they had answered incorrectly, or they could simply change their responses and re-submit. The system did not have the
capability to record learners’ activity during this phase, so the number of attempts was not recorded. However, the researcher and lab assistants did not witness any participant spending more than a few seconds on each item. It would be reasonable to estimate that learners in the +PI group spent on average 3-5 minutes on this “discrimination exercise.”

The “discrimination exercise” included in the +PI modules was unlikely to provide true task-specific practice for the experimental discrimination test, because the task conditions were quite different. In the practice exercise, learners identified one syllable at a time either as Spanish-like or English-like. They did not discriminate between two syllables. The practice exercise was exceedingly explicit, directing learners to pay attention to particular features and ordering the stimuli in a predictable sequence, whereas the discrimination test gave learners no indication about the phones being targeted in a particular item pairing or the features that were important for any particular item pairing. Thus while the exercise may have represented one of the more useful components of phonetics instruction for attuning learners to relevant phonetic features of L2 phones, it was not considered a task-specific practice advantage of the +PI condition.

Furthermore, the results of the identification test could not be explained in terms of practice. The learners in the +PI got no transcription practice during instruction, whereas the -PI instruction was transcription based. Yet the +PI still fared better than the -PI in the identification test, which required them to transcribe nonce words. It seems that the +PI learned something about the L2 phones during instruction that they were able to transfer to other task types.

The advantage conferred by phonetics instruction to the +PI was linked to time of test. Both groups improved immediately after instruction on both perception tests,
suggesting that both phonetics instruction and dictation exercises were beneficial in the short term. While both groups improved significantly from pretest to delayed posttest on the identification test, only the +PI improved significantly from pretest to delayed posttest on the discrimination test. Said another way, the knowledge or skill built through the phonetics instruction was retained longer than that of the dictation exercises, and the learners in the -PI retained learning only on the test that was similar in its processing demands to the instructional condition they had received. A possible interpretation for these findings is that learners in the -PI gained implicit knowledge and/or task-specific practice that benefited them only in the identification test, whereas learners in the +PI gained more explicit knowledge about the target phones which they were able to employ for both task types and also retain and utilize in the longer term. Neither perception task was speeded to the degree that would preclude learners from consciously employing explicit knowledge. These interpretations must be tested in future research, as the present study did not include measures of implicit or explicit knowledge or online measures of learners’ orientation towards the tests.

The relative advantage of the +PI should not be overstated. The effect sizes (5%-17%) were small enough to make it clear that brief lessons in phonetics and phonology is no panacea for learners’ problems in perceiving L2 phones, much as previous studies have found that even semester-long courses in phonetics do not result in learners’ production becoming fully target-like. Though alternative types of instruction were not the focus of this study, the data indicated that focused listening with dictation also served to improve learners’ perception of L2 phones, and so certainly a case could be made for including many types of potentially beneficial instruction in the FL curriculum.
Furthermore, the present study provided indications about how one might tailor instruction in Spanish phonetics and phonology so as to maximize efficient use of instructional resources. For instance, learners in this study were quite accurate in their discrimination of \([p]/[p^b], [ɾ]/[ɾ],\) and \([r]/[ɹ]\) even at the pretest, indicating that more practice in discriminating these phone pairs may not be necessary. However, for these same phones, the data suggested that learners could benefit from more practice in perceiving Spanish \([p]\) as \(/p/\) rather than \(/b/\), discriminating \([ɾ]/[ɾ]\) when they occur word-initially and word-finally, and identifying intervocalic \([ɾ]\) as \(/ɾ/\) rather than an allophone of \(/d/\). The data further suggested that learners were only at chance in their discrimination of \([k]/[k^h], [β]/[b],\) and \([ɣ]/[g]\) at the pretest. If, as the SLM suggests, the first step toward achieving target-like pronunciation is to learn to discriminate subtle phonetic differences between L2 phones and their analogous L1 phones, then particular attention should be paid to attuning learners’ discrimination of these problematic phone pairs. Finally, since the effectiveness of instruction generally did not depend on learners’ L2 experience level in the present study, the data suggested that there was no optimal time to insert phonetics instruction into the Spanish FL curriculum, but rather that phonetics instruction could be appropriate and beneficial at several points along the introductory and intermediate stages.

### 3.4.3 Limitations of the Perception Tests

The limitations of the study overall will be discussed in the concluding chapter. This section will deal specifically with limitations in the design of the perception tests, of which there were several. The AX discrimination test was restricted to mono-
disyllabic stimulus items lacking semantic content, in an attempt to eliminate confounding variables, and so the results cannot be extrapolated to infer the accuracy with which learners could have discriminated the same phone pairs in more extended or meaningful language. The discrimination task was a two-option forced choice test, limiting the range of possible scores. A more nuanced scale for rating discrimination, such as Likert scale similarity judgments (see Boomershine et al., 2008, for an example), could have provided richer data about learners’ discriminatory abilities, as could reaction time measurements or a range of additional discrimination tasks with varying conditions.

The choice to use acoustically identical stimuli was a justified as a means to ensure that learners interpreted the task instructions appropriately, given that the target phones were not all contrastive and could not form minimal pairs in Spanish. The long ISI presumably ensured that learners were drawing on their phonetic and phonemic knowledge, not just their acoustic memory store, to complete the task. However, it should be noted that in most prior research (on contrastive L2 phones), discrimination tasks employed phonetically identical but acoustically different stimuli. The extent to which using acoustically identical stimuli could have unduly influenced learners’ responses remains an empirical question.

The identification test asked learners to type their responses using Spanish orthography in order to provide identical conditions to the +PI and the -PI, which received no instruction in phonetic symbols. As such, the identification test could not strictly isolate phonetic perception from knowledge of Spanish grapheme-phoneme correspondences. An effort was made to isolate the two types of knowledge by accepting non-standard spellings if they seemed to indicate that the learner had correctly identified
the target phone. All the abovementioned limitations highlight the need to develop better
tasks for measuring perception of L2 sounds, particularly when the sounds are not
contrastive in the L2 (so minimal pair tasks cannot be employed) and when a control
group will be incorporated in the design (precluding the use of phonetic symbols in
assessment tasks). Future studies should attempt to address these task limitations. Much
work is needed in the area of developing experimental tasks that can investigate learners’
perception of non-contrastive L2 phones at multiple levels of processing. The current
study should be understood as a first step in that direction and as a suggestion that the
impact of phonetics instruction is to attune learners’ perception to the L2.
4.1 Introduction

4.1.1 The Perception-Production Link

Several of the models of L2 speech acquisition presented in §1.2 suggested that perception leads production. The Speech Learning Model (SLM) and its claims that target-like perception precedes target-like production were discussed in detail in §1.2. These theoretical claims are supported by a good deal of data suggesting that there is a strong link between perception and production of L2 speech sounds (see Akerberg, 2005; Llisterrri, 1995, for reviews). For instance, Flege and colleagues found moderate correlations between the perception and production of L2 (usually English) segmental contrasts (e.g., Flege, MacKay & Meador, 1999; Flege, 1993). More recently, Hwang (2011) reported both behavioral and neurological data demonstrating that Korean speakers who mispronounced stop-nasal sequences in English by inserting a vowel also misperceived those sequences in the same way. They reported hearing an illusory vowel after voiced stops. Evans and Iverson (2007) even found a correlation between perception and production in adult L1 accent change within the United Kingdom. The young adults (n=27) in their study whose accent had changed and adopted London norms after attending university also demonstrated a change in perception of local accents.

Goto (1971) disputed the putative link between perception and production based on evidence that some Japanese learners of English could produce the /ʃ/-/l/ contrast well even though they had trouble perceiving the contrast in recordings of their own voices,
but Goto’s (1971) results have not been replicated (Munro, 2008). Several researchers have contended that perception and production skills for L2 contrasts develop differently, citing evidence that learners acquire L2 phonemic contrasts (Eckman, Iverson, Fox, Jacewicz & Lee, 2011) and particular features of L2 segments (de Jong, Silber & Park, 2009; de Jong, Hao & Park, 2009) with different acquisitional patterns in the perceptive and productive modes. While it is obviously true that perception and production are separate behavioral actions, and while there is evidence that they are dissociated neurologically as well, since brain-lesioned patients can be selectively impaired in one or the other (e.g., Praamstra, Hagoort, Maassen & Crul, 1991; Dronkers, 1996), it is also clear that there is an important, if complex, relationship between perception and production.

Llistterri (1995) cautioned that it could be problematic to directly compare perception tasks with production tasks. Llistterri (1995) argued that the contradictory results sometimes obtained in the research regarding a putative perception-production correlation might have more to do with the nature of the experimental tasks used than their underlying constructs. Production tasks may range from isolated word reading to spontaneous speech samples drawn from real conversations, but even in their narrowest form they present participants with a much wider range of response choices than the typical forced-choice, two-alternative perception test. Perception tests are speeded, whereas production tests generally are not. For these and other reasons, comparisons between perception and production tasks must be undertaken with caution, and researchers should note the extent to which the conditions of the experimental tasks are comparable.
Specifying the exact conditions of the tasks is particularly crucial in perception tasks because different conditions favor different levels of processing, as was discussed in §3.1.2. In their study of French speaking learners of English (n=17), Peperkamp and Bouchon (2011) did not find a correlation between learners’ perception and production of the /i/-/ɪ/ contrast. Though previous studies had found a correlation, Peperkamp and Bouchon explained that their findings were related to task conditions. They used an ABX discrimination task with an inter-stimulus interval (ISI) of 500ms, whereas “all the previous studies that found a perception-production correlation used either an off-line task, such as identification or goodness rating, or discrimination with a much larger ISI (at least 1.2s)” (p. 164). They hypothesized that with offline tasks or longer ISI, participants could use their phonological loop, or subvocal rehearsal mechanism, which is also used in production. They speculated that the shorter ISI prevented participants from exploiting their phonological loop and thus participants’ performance on the perception task was disassociated from their performance on the production task.

The Werker and Logan (1985) study was discussed in detail in §3.1.2. They found evidence for at least three independent levels of processing in their Hindi /ʈ/-/t/ contrast discrimination test: auditory, phonetic, and phonemic. Both auditory and phonetic processing could be characterized as continuous, using Wode’s terminology (1981; 1990). Werker and Logan (1985) concluded that there was strong evidence for auditory processing at 250ms ISI, and some evidence for auditory processing after several blocks of many trials with 500ms ISI, but no evidence of auditory processing with ISIs as long as 1500ms. It follows that the discrimination tasks best suited to tapping linguistically
relevant phonological categories, and ones that are most likely to be correlated with production abilities, are those with ISIs at 1500ms or more.

In sum, it is important to note the difficulties inherent in directly comparing results of perception and production tasks (Llisterrri, 1995). It is also important to remember that identical discrimination task stimuli may elicit different responses from participants depending on the speed of their presentation. Yet even with identical stimuli and participants, different task conditions may result in one perception task correlating with a production task whereas another does not.

Having argued that perception and production of an L2 are closely linked abilities and that both contribute to the phenomenon of “foreign” accent, the question now at hand is whether instruction is equally necessary or equally helpful for improving both abilities. A number of laboratory-based training studies have suggested that intensive perception training helped participants to perceive L2 contrastive phones better as well as improve their production of those phones, and other studies have suggested that pronunciation training was associated with improvements in perception (all reviewed in §1.3). It was argued in §1.3, §1.4, and §3.1.3 that there is relatively little research carried out within the instructed learning context of a typical L2/FL classroom that addressed perception of L2 speech sounds, much less directly relates perception with production. One exception is Zampini (1998), who investigated English speaking Spanish learners’ perception and production of the Spanish /p/-/b/ contrast. Zampini reported that learners’ perceptual boundaries for the VOT of /p/ and /b/ did not correlate with the VOT they produced for those phones in most of the target items (e.g., pace / base; peso / beso). However, much
more research is clearly needed to understand the impact that phonetics instruction has (or does not have) on the perception and production of instructed L2/FL learners.

4.1.2 Research Questions and Hypotheses

The present analysis sought to fill this gap by probing the relative facilitative effect of phonetics instruction in L2 learners’ honing of their perception and production skills. The research questions motivating the current analysis were: 1. Is learners’ ability to perceive L2 phones related with their ability to produce them?, 2. How does the effect of L2 phonetics instruction on perceptive and productive skills compare?, and 3. How do learners’ perceptive abilities impact the effect of instruction on learners’ production? The hypothesis relating to research question one was that learners’ perceptive and productive abilities were correlated. Theoretically the abilities are interrelated, and empirically they have been found to correlate when the perception tasks employed are offline and the targets are contrastive L2 phones. The second hypothesis was that instruction would have a greater impact on perception than production for most phones. The results of the preceding chapters would suggest so, since the +PI group outperformed the –PI in the perception tasks but not in the production task. The third hypothesis was that learners’ perceptive abilities would influence to what extent learners improved their production of the target phones after instruction.

4.2 Methods

The methods used were the same as those described in previous chapters. The target phones, context, participants, experimental design, instruction, tests, and scoring
are the same as those described previously (in §2.2 and §3.2). Recall that the
discrimination and identification tests included five items per target phone, whereas the
production test included three items per target phone. In order to give equal weight to
each item across the tests, the production test scores were multiplied by a factor of
1.6667. The discrimination test included just three items for the /ɾ/-/ɹ/ pairing and the /r/-
/ɹ/ pairing, but scores on these items were similarly multiplied by a factor of 1.6667. In
this way, every target phone was associated with 5 possible points on every experimental
test. Also, in order to make comparisons across the three tests fair and transparent, all 40
points per test were included in the following analysis. No items or phones were
excluded.

4.3 Results

4.3.1 Correlation Analysis

In order to address the first research question (Is learners’ ability to perceive L2
phones related with their ability to produce them?), a correlation analysis was conducted
on learners’ pre-test scores on the discrimination, identification, and production tests.
The analysis examined pretest scores only and combined the +PI and –PI groups because
it was designed to determine the relationship between the three abilities before
instructional intervention. Each target phone was analyzed separately in addition to the
aggregate set of phones. Participants’ scores were grouped by matriculation level
because the correlations between the abilities could conceivably have been stronger at
one developmental stage than another. Table 4.1 presents the Kendall’s Tau ($\tau$)
coefficients that were significant at $\alpha = .05$. The distribution of significant correlations between these variables was nearly identical when the coefficients calculated were Pearson product-moment correlation coefficients. As shown in Table 4.1, the discrimination, identification, and perception scores correlated positively for several target phones and learner groups.

![Table 4.1 Kendall's Tau Correlations Between Pretest Scores](image)

<table>
<thead>
<tr>
<th></th>
<th>Discrimination &amp; Identification</th>
<th>Discrimination &amp; Production</th>
<th>Identification &amp; Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Year</td>
<td>Second Year</td>
<td>Third Year</td>
</tr>
<tr>
<td>aggregate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/t/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/k/</td>
<td>0.39</td>
<td>0.26</td>
<td>0.35</td>
</tr>
<tr>
<td>[β]</td>
<td></td>
<td>-0.37</td>
<td>0.53</td>
</tr>
<tr>
<td>[Ø]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[y]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɛ/</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>/ɪ/</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

All correlations listed are significant at $p < .05$

a Could not be computed because at least one of the variables was constant.

All tests were two tailed.

Discrimination scores did not vary with identification scores for the aggregate of phones or most individual phones, but there were two exceptions. Positive relationships\(^{20}\) between discrimination and identification were found for first year learners on /k/ ($\tau(44) = .39, p = .004$) and second year learners on /p/ ($\tau(30) = .35, p = .03$). Note that

\(^{20}\) Note that the magnitude of $\tau$ is not interpreted in the same way as Pearson’s $r$. Kendall’s $\tau$ is the difference between the probability that the observed data are concordant (i.e., in the same order for the two variables) and the probability that they are discordant (i.e., in different orders). A positive value for $\tau$ indicates concordance between two variables.
correlations for /ɾ/ and /ɾ/ could not be computed in several cases because discrimination of /ɾ/-/ɹ/ and /ɾ/-/ɹ/ was at ceiling level already in the pretest. So while the task abilities were not totally unrelated, learners’ performance on one was not related to performance on the other, for most phones. In particular, there was no evidence of a relationship between discrimination and identification of approximants ([β, δ, γ]) and there was no evidence of a relationship between these two perceptive abilities for third year learners.

Discrimination scores, however, did vary with production scores for some phones. Positive relationships were found for first year learners on /k/ (τ(43) = .26, p = .03), second year learners on /k/ (τ(30) = .35, p = .016), third year learners on [β] (τ(17) = .53, p = .011) and third year learners on [γ] (τ(17) = .41, p = .04). A negative relationship was found for second year learners on [β] (τ(30) = -.37, p = .02). This result was unexpected, and since it was the only negative correlation found, it might best be regarded as an anomaly. There was no evidence of a relationship between discrimination and production for the rhotic phones, though correlations could not be computed because discrimination of /ɾ/-/ɹ/ and /ɾ/-/ɹ/ was at ceiling level already in the pretest.

Identification scores varied with production scores, both for the aggregate scores as well as for two individual phones. The 8-phone aggregate identification scores were positively correlated with production scores for first year learners (τ(44) = .44, p < .001) and second year learners (τ(30) = .26, p = .04), and their correlation neared statistical significance for third year learners (τ(17) = .32, p = .06). There was a significant correlation found for two out of eight phones. A positive correlation between identification and production of /t/ was found for first year (τ(43) = .37, p = .007) and
second year ($\tau(30) = .36, p = .014$) learners. In addition, there was a positive relationship between first year learners’ identification and production of [δ] ($\tau(43) = .28, p = .04$).

In sum, while there was less evidence of a concordant relationship between the two perception tasks, the data indicated that the perception abilities (particularly identification) were positively correlated with production abilities, before beginning instruction. This relationship was found with various phones. The degree of interrelatedness among experimental tests varied across the phones and matriculation levels. There was more evidence of a positive relationship between perception and production of the stop consonants /t, k/ than for other phones, as the tests were positively correlated for at least two of the three matriculation groups for /t, k/. Learners at each matriculation evidenced at least two significant, positive relationships between a perception and a production test, though the phones in question differed, and the number of relationships decreased steadily across matriculation levels (4 for first year learners, 3 for second year learners, and 2 for third year learners).

4.3.2 Descriptive Statistics

Gain scores were calculated for each test, from pretest to posttest as well as pretest to delayed posttest. Descriptive statistics of the gain scores are presented in Tables 4.2 and 4.3, separated by learner matriculation level and instructional condition. Figure 4.1 graphs the gain scores of the +PI and -PI groups for all six tests.
### Table 4.2 Gains Pretest to Posttest

<table>
<thead>
<tr>
<th></th>
<th>First Year</th>
<th>Second Year</th>
<th>Third Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+PI (N=23)</td>
<td>-PI (N=23)</td>
<td>+PI (N=18)</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}$ (s)</td>
<td>$\bar{x}$ (s)</td>
<td>$\bar{x}$ (s)</td>
</tr>
<tr>
<td>Grand Total (out of 120)</td>
<td>9.36 (6.75)</td>
<td>7.84 (6.78)</td>
<td>7.53 (5.56)</td>
</tr>
<tr>
<td>Experimental Tests (out of 40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td>2.79 (4.08)</td>
<td>3.93 (4.90)</td>
<td>3.43 (3.71)</td>
</tr>
<tr>
<td>Identification</td>
<td>5.05 (3.51)</td>
<td>3.16 (3.49)</td>
<td>3.58 (1.87)</td>
</tr>
<tr>
<td>Production</td>
<td>1.52 (2.53)</td>
<td>0.75 (3.14)</td>
<td>0.52 (2.75)</td>
</tr>
<tr>
<td>Individual Phones (3 tests) (out of 15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>1.39 (1.63)</td>
<td>0.97 (1.34)</td>
<td>0.65 (1.88)</td>
</tr>
<tr>
<td>/t/</td>
<td>1.13 (1.79)</td>
<td>1.20 (2.04)</td>
<td>0.69 (1.57)</td>
</tr>
<tr>
<td>/k/</td>
<td>0.76 (1.24)</td>
<td>1.94 (1.43)</td>
<td>1.35 (1.67)</td>
</tr>
<tr>
<td>[β]</td>
<td>1.51 (1.43)</td>
<td>0.51 (1.97)</td>
<td>0.66 (2.02)</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>2.16 (1.44)</td>
<td>1.90 (1.80)</td>
<td>2.15 (2.14)</td>
</tr>
<tr>
<td>[ɣ]</td>
<td>0.99 (1.91)</td>
<td>0.71 (1.93)</td>
<td>1.08 (1.44)</td>
</tr>
<tr>
<td>/s/</td>
<td>0.49 (2.27)</td>
<td>0.76 (1.56)</td>
<td>0.18 (2.04)</td>
</tr>
<tr>
<td>/z/</td>
<td>0.28 (1.95)</td>
<td>0.26 (1.77)</td>
<td>0.77 (1.36)</td>
</tr>
</tbody>
</table>

### Table 4.3 Gains Pretest to Delayed Posttest

<table>
<thead>
<tr>
<th></th>
<th>First Year</th>
<th>Second Year</th>
<th>Third Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+PI (N=20)</td>
<td>-PI (N=20)</td>
<td>+PI (N=18)</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}$ (s)</td>
<td>$\bar{x}$ (s)</td>
<td>$\bar{x}$ (s)</td>
</tr>
<tr>
<td>Grand Total (out of 120)</td>
<td>4.25 (5.61)</td>
<td>2.62 (5.50)</td>
<td>3.82 (3.46)</td>
</tr>
<tr>
<td>Experimental Tests (out of 40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination</td>
<td>0.86 (4.17)</td>
<td>0.35 (4.59)</td>
<td>1.83 (4.11)</td>
</tr>
<tr>
<td>Identification</td>
<td>2.67 (2.27)</td>
<td>2.24 (2.79)</td>
<td>1.62 (2.21)</td>
</tr>
<tr>
<td>Production</td>
<td>0.97 (3.15)</td>
<td>0.34 (2.90)</td>
<td>0.61 (2.29)</td>
</tr>
<tr>
<td>Individual Phones (3 tests) (out of 15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>0.58 (1.40)</td>
<td>0.27 (1.10)</td>
<td>0.55 (1.44)</td>
</tr>
<tr>
<td>/t/</td>
<td>0.91 (1.18)</td>
<td>1.05 (2.14)</td>
<td>0.76 (1.50)</td>
</tr>
<tr>
<td>/k/</td>
<td>0.74 (1.44)</td>
<td>0.62 (1.17)</td>
<td>0.39 (1.79)</td>
</tr>
<tr>
<td>[β]</td>
<td>0.10 (1.49)</td>
<td>-0.89 (1.42)</td>
<td>0.15 (1.82)</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>0.67 (1.69)</td>
<td>0.63 (1.51)</td>
<td>0.87 (1.87)</td>
</tr>
<tr>
<td>[ɣ]</td>
<td>0.74 (1.29)</td>
<td>0.51 (1.98)</td>
<td>0.69 (1.45)</td>
</tr>
<tr>
<td>/s/</td>
<td>0.68 (1.43)</td>
<td>0.57 (1.32)</td>
<td>0.39 (1.29)</td>
</tr>
<tr>
<td>/z/</td>
<td>-0.17 (1.80)</td>
<td>-0.14 (1.54)</td>
<td>0.03 (1.24)</td>
</tr>
</tbody>
</table>
Figure 4.1 shows that gains on all tests were rather small (5 or fewer points) and were relatively smaller for the production tests than the perception tests at any given test time. Yet all were positive values, indicating that both instructional conditions were beneficial for all three abilities in the short term and three weeks later at the delayed posttest, to varying degrees. Figure 4.1 suggests that the +PI demonstrated more gains than the –PI group for most tests at most test administration times and that an interaction between instructional condition and test type occurred in the immediate posttest but not in the delayed.

4.3.3 Distribution of Gain Scores

Subsets of gain scores were analyzed for normality. In all, 48 subsets were analyzed: first, second, and third year +PI and –PI groups’ discrimination test gains, identification test gains, production test gains, and grand total gains, in each case both
from pretest to posttest as well as pretest to delayed posttest test (3x2x4x2=48). Most of these subsets of gain scores were normally distributed. Only 6 of the subsets did not pass a Shapiro-Wilk test of normality (p < .05), just 5 subsets demonstrated slight skew and/or kurtosis, and these subsets were distributed seemingly at random across matriculation levels and instructional condition groups; no one test or test administration time appeared to have a particularly abnormal distribution. It was decided that the distribution of the data was normal enough to satisfy the assumptions of an ANOVA.

4.3.4 RMANOVA and Subsequent Contrasts

In order to address the second research question (How does the effect of L2 phonetics instruction on perceptive and productive skills compare?), gain scores were analyzed with a RMANOVA with instructional condition and matriculation level as a between-subjects factors and test type and time (i.e., pretest-posttest gain or pretest-delayed posttest gain) as within-subject factors. This analysis was performed with gain scores from each of the eight phones as well as the aggregate scores. Tables 4.4 – 4.20 present the results and follow-up pairwise comparisons performed (with corrected p values) for significant main effects and interactions. Note that the effect of time was not submitted to pairwise comparisons because it had only two levels here. To an extent, these RMANOVAs repeated the analyses presented in chapters 2 and 3. Yet the relationship between test types became clearer when all three tests were entered into one model, and it was easier to visualize the impact of instruction on each ability and across target phones when gain scores were plotted together (in Figures 4.2 – 4.10).
### Table 4.4 RMANOVA of Gain Scores on Three Test Types: Aggregate of 8 Phones

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>( \eta^2 )</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>19.98</td>
<td>1.00</td>
<td>19.98</td>
<td>1.19</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Level</td>
<td>31.73</td>
<td>2.00</td>
<td>15.86</td>
<td>0.95</td>
<td>0.02</td>
<td>0.21</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>19.36</td>
<td>2.00</td>
<td>9.68</td>
<td>0.58</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Error</td>
<td>1377.09</td>
<td>82.00</td>
<td>16.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>200.79</td>
<td>1.00</td>
<td>200.79</td>
<td>39.69***</td>
<td>0.33</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>1.12</td>
<td>1.00</td>
<td>1.12</td>
<td>0.22</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Time X Level</td>
<td>8.45</td>
<td>2.00</td>
<td>4.22</td>
<td>0.84</td>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>1.01</td>
<td>2.00</td>
<td>0.50</td>
<td>0.10</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>414.86</td>
<td>82.00</td>
<td>5.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>308.37</td>
<td>1.72</td>
<td>179.00</td>
<td>7.85***</td>
<td>0.09</td>
<td>0.92</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>90.76</td>
<td>1.72</td>
<td>52.69</td>
<td>2.31</td>
<td>0.03</td>
<td>0.43</td>
</tr>
<tr>
<td>Type X Level</td>
<td>57.34</td>
<td>3.45</td>
<td>16.64</td>
<td>0.73</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>46.84</td>
<td>3.45</td>
<td>13.60</td>
<td>0.60</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>3222.70</td>
<td>141.26</td>
<td>22.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>70.01</td>
<td>1.48</td>
<td>47.32</td>
<td>6.20***</td>
<td>0.07</td>
<td>0.81</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>36.26</td>
<td>1.48</td>
<td>24.51</td>
<td>3.21*</td>
<td>0.04</td>
<td>0.52</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>24.35</td>
<td>2.96</td>
<td>8.23</td>
<td>1.08</td>
<td>0.03</td>
<td>0.28</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>42.34</td>
<td>2.96</td>
<td>14.31</td>
<td>1.87</td>
<td>0.04</td>
<td>0.47</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>926.60</td>
<td>121.31</td>
<td>7.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

With a Greenhouse-Geisser Correction

### Table 4.5 Pairwise Comparisons of Three Test Types: Aggregate of 8 Phones

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>( \eta^2 )</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>5.64</td>
<td>1</td>
<td>5.64</td>
<td>0.25</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>214.47</td>
<td>1</td>
<td>214.47</td>
<td>7.95***</td>
<td>0.09</td>
<td>0.80</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>289.70</td>
<td>1</td>
<td>289.70</td>
<td>24.00***</td>
<td>0.23</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Time X Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>265.85</td>
<td>1</td>
<td>265.85</td>
<td>8.59***</td>
<td>0.10</td>
<td>0.83</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>443.40</td>
<td>1</td>
<td>443.40</td>
<td>16.99****</td>
<td>0.17</td>
<td>0.98</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>22.58</td>
<td>1</td>
<td>22.58</td>
<td>2.37</td>
<td>0.03</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

### Figure 4.2 Gain Scores (all target phones)

- **Test Type**
  - Discrimination
  - Identification
  - Production

- **Axes**
  - X-axis: Time
  - Y-axis: Mean Gain Scores (Pretest to Posttest, Pretest to Delayed Posttest)
What is most interesting to note from the results relating to the aggregate of target phones is that 1) learners improved all skills following instruction, 2) learners improved in their discrimination and identification of the target phones more than their production, and 3) learners retained more of their gains in identification and production than in discrimination. Furthermore, the time by type by condition interaction neared statistical significance for the eight-phone aggregate ($F(1.48,121) = 3.21, p = .06$), and had the conservative Greenhouse-Geisser correction not been applied for a violation of sphericity, this interaction would have been significant ($p = .05$). As was suggested by Figure 4.1, instructional condition may have played a more important role than could be elucidated by the present design.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.20</td>
<td>1</td>
<td>0.20</td>
<td>1.07</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Level</td>
<td>0.23</td>
<td>2</td>
<td>0.11</td>
<td>0.61</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>0.06</td>
<td>2</td>
<td>0.03</td>
<td>0.16</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>6.31</td>
<td>1</td>
<td>6.31</td>
<td>17.65**</td>
<td>0.18</td>
<td>0.99</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.08</td>
<td>1</td>
<td>0.08</td>
<td>0.22</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Time X Level</td>
<td>1.08</td>
<td>2</td>
<td>0.54</td>
<td>1.51</td>
<td>0.04</td>
<td>0.31</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>0.62</td>
<td>2</td>
<td>0.31</td>
<td>0.87</td>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>29.31</td>
<td>82</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>20.09</td>
<td>2</td>
<td>10.32</td>
<td>8.07***</td>
<td>0.09</td>
<td>0.95</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>9.13</td>
<td>2</td>
<td>4.69</td>
<td>3.67**</td>
<td>0.04</td>
<td>0.66</td>
</tr>
<tr>
<td>Type X Level</td>
<td>3.65</td>
<td>4</td>
<td>0.94</td>
<td>0.73</td>
<td>0.02</td>
<td>0.23</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>3.58</td>
<td>4</td>
<td>0.92</td>
<td>0.72</td>
<td>0.02</td>
<td>0.23</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>204.15</td>
<td>160</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>0.67</td>
<td>2</td>
<td>0.35</td>
<td>0.88</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>1.06</td>
<td>2</td>
<td>0.56</td>
<td>1.39</td>
<td>0.02</td>
<td>0.29</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>1.93</td>
<td>4</td>
<td>0.51</td>
<td>1.26</td>
<td>0.03</td>
<td>0.38</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>1.20</td>
<td>4</td>
<td>0.32</td>
<td>0.79</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>62.63</td>
<td>155</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *$p<.10$, **$p<.05$, ***$p<.01$, ****$p<.001$
The results relating to the target phone /p/ indicated that learners in both groups improved their identification of /p/ more than their discrimination and production of /p/.

It was perhaps not surprising that there would be small gains in discrimination, considering that the pretest discrimination scores were so high (87% accurate). However, the significant type by condition interaction suggested that for the +PI group, these differences in gain scores across test types were greater than those in the –PI. That is, learners in the implicit dictation condition (-PI) improved their scores about equally as well on the three experimental tests. On the other hand, learners receiving phonetics instruction (+PI) improved their ability to identify and produce /p/ more than their discrimination. In fact, their “gain” scores were negative, indicating that after receiving phonetics instruction, their ability to discriminate [p]-[pʰ] actually decreased.
Table 4.8 RMANOVA of Gain Scores on Three Test Types: /t/

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.17</td>
<td>1</td>
<td>0.17</td>
<td>0.65</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Level</td>
<td>1.12</td>
<td>2</td>
<td>0.56</td>
<td>2.22</td>
<td>0.05</td>
<td>0.44</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>0.42</td>
<td>2</td>
<td>0.21</td>
<td>0.83</td>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1.47</td>
<td>1</td>
<td>1.47</td>
<td>4.14**</td>
<td>0.05</td>
<td>0.52</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.33</td>
<td>1</td>
<td>0.33</td>
<td>0.93</td>
<td>0.01</td>
<td>0.16</td>
</tr>
<tr>
<td>Time X Level</td>
<td>1.69</td>
<td>2</td>
<td>0.85</td>
<td>2.39*</td>
<td>0.06</td>
<td>0.47</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>0.36</td>
<td>2</td>
<td>0.18</td>
<td>0.50</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>29.04</td>
<td>82</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>16.50</td>
<td>2</td>
<td>9.05</td>
<td>5.29***</td>
<td>0.06</td>
<td>0.80</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>3.99</td>
<td>2</td>
<td>2.19</td>
<td>1.28</td>
<td>0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>Type X Level</td>
<td>2.23</td>
<td>4</td>
<td>0.61</td>
<td>0.36</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>6.45</td>
<td>4</td>
<td>1.77</td>
<td>1.04</td>
<td>0.03</td>
<td>0.31</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>255.73</td>
<td>150</td>
<td>1.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>21.28</td>
<td>2</td>
<td>12.35</td>
<td>24.23****</td>
<td>0.23</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>0.39</td>
<td>2</td>
<td>0.23</td>
<td>0.45</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>0.28</td>
<td>3</td>
<td>0.08</td>
<td>0.16</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>2.27</td>
<td>3</td>
<td>0.66</td>
<td>1.29</td>
<td>0.03</td>
<td>0.37</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>72.02</td>
<td>141</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *p < .10, **p < .05, ***p < .01, ****p < .001

With a Greenhouse-Geisser Correction

Table 4.9 Pairwise Comparisons of Three Test Types: /t/

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>13.13</td>
<td>1</td>
<td>13.13</td>
<td>6.54**</td>
<td>0.07</td>
<td>0.72</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>0.05</td>
<td>1</td>
<td>0.05</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>11.58</td>
<td>1</td>
<td>11.58</td>
<td>9.85***</td>
<td>0.11</td>
<td>0.87</td>
</tr>
<tr>
<td>Time X Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>84.98</td>
<td>1</td>
<td>84.98</td>
<td>36.76****</td>
<td>0.31</td>
<td>1.00</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>18.48</td>
<td>1</td>
<td>18.48</td>
<td>9.97***</td>
<td>0.11</td>
<td>0.88</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>24.20</td>
<td>1</td>
<td>24.20</td>
<td>21.93****</td>
<td>0.21</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Significant at *p < .10, **p < .05, ***p < .01, ****p < .001

Figure 4.4 Gain Scores for /t/

- Test Type
  - Discrimination
  - Identification
  - Production

Mean Gain Scores

Pretest to Posttest  Pretest to Delayed Posttest

Time
The results relating to /t/ indicated that both groups improved their identification of /t/ more so than their discrimination and production of /t/. In fact, the gains witnessed in discrimination of /t/ immediately following instruction disappeared by the 3-week delayed posttest. This was an interesting result, particularly because learners’ pretest accuracy in identifying /t/ was high (84%). Just the opposite occurred with /k/. Learners in both groups improved their discrimination of /k/ more than their identification and production of /k/, in the posttest and delayed posttest. This result was not surprising, since learners had little opportunity to improve on their pretest accuracy in identifying /t/ (97%) when compared with their poor pretest discrimination of [k]-[kʰ] (43%).

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.14</td>
<td>1</td>
<td>0.14</td>
<td>0.78</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Level</td>
<td>0.04</td>
<td>2</td>
<td>0.02</td>
<td>0.12</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>0.15</td>
<td>2</td>
<td>0.08</td>
<td>0.42</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>Error</td>
<td>15.00</td>
<td>82</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>8.81</td>
<td>1</td>
<td>8.81</td>
<td>32.52</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.95</td>
<td>1</td>
<td>0.95</td>
<td>3.49</td>
<td>0.04</td>
<td>0.45</td>
</tr>
<tr>
<td>Time X Level</td>
<td>1.18</td>
<td>2</td>
<td>0.59</td>
<td>2.18</td>
<td>0.05</td>
<td>0.43</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>0.72</td>
<td>2</td>
<td>0.36</td>
<td>1.34</td>
<td>0.03</td>
<td>0.28</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>22.22</td>
<td>82</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>36.79</td>
<td>2</td>
<td>19.37</td>
<td>19.67</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>5.64</td>
<td>2</td>
<td>3.02</td>
<td>3.02</td>
<td>0.04</td>
<td>0.51</td>
</tr>
<tr>
<td>Type X Level</td>
<td>2.73</td>
<td>3</td>
<td>0.87</td>
<td>0.73</td>
<td>0.02</td>
<td>0.21</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>0.97</td>
<td>3</td>
<td>0.31</td>
<td>0.26</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>155.38</td>
<td>129</td>
<td>1.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>5.91</td>
<td>1</td>
<td>4.15</td>
<td>9.48</td>
<td>0.10</td>
<td>0.94</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>1.10</td>
<td>1</td>
<td>0.77</td>
<td>1.76</td>
<td>0.02</td>
<td>0.31</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>0.56</td>
<td>3</td>
<td>0.20</td>
<td>0.45</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>0.69</td>
<td>3</td>
<td>0.24</td>
<td>0.56</td>
<td>0.01</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001
With a Greenhouse-Geisser Correction
Table 4.11 Pairwise Comparisons of Three Test Types: /k/

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>29.30</td>
<td>1</td>
<td>29.30</td>
<td>24.60***</td>
<td>0.23</td>
<td>1.00</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>25.77</td>
<td>1</td>
<td>25.77</td>
<td>22.02***</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>0.11</td>
<td>1</td>
<td>0.11</td>
<td>0.26</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Time X Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>17.95</td>
<td>1</td>
<td>17.95</td>
<td>11.32***</td>
<td>0.12</td>
<td>0.91</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>17.48</td>
<td>1</td>
<td>17.48</td>
<td>10.29***</td>
<td>0.11</td>
<td>0.89</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Significant at *$p<.10$, **$p<.05$, ***$p<.01$, ****$p<.001$.

Figure 4.5 Gain Scores for /k/

Table 4.12 RMANOVA of Gain Scores on Three Test Types: [b]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.42</td>
<td>1</td>
<td>0.42</td>
<td>1.50</td>
<td>0.02</td>
<td>0.23</td>
</tr>
<tr>
<td>Level</td>
<td>0.14</td>
<td>2</td>
<td>0.07</td>
<td>0.25</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>0.80</td>
<td>2</td>
<td>0.40</td>
<td>1.44</td>
<td>0.03</td>
<td>0.30</td>
</tr>
<tr>
<td>Error</td>
<td>22.73</td>
<td>82</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>13.78</td>
<td>1</td>
<td>13.78</td>
<td>33.22***</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.09</td>
<td>2</td>
<td>0.09</td>
<td>0.21</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Time X Level</td>
<td>2.04</td>
<td>4</td>
<td>0.50</td>
<td>1.02</td>
<td>0.06</td>
<td>0.48</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>0.31</td>
<td>4</td>
<td>0.08</td>
<td>0.37</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>34.01</td>
<td>82</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>7.39</td>
<td>2</td>
<td>4.20</td>
<td>2.39*</td>
<td>0.03</td>
<td>0.44</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>0.94</td>
<td>4</td>
<td>0.23</td>
<td>0.30</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Type X Level</td>
<td>10.83</td>
<td>8</td>
<td>1.35</td>
<td>1.75</td>
<td>0.04</td>
<td>0.49</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>9.45</td>
<td>16</td>
<td>0.59</td>
<td>1.53</td>
<td>0.04</td>
<td>0.43</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>252.95</td>
<td>144</td>
<td>1.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>8.62</td>
<td>2</td>
<td>4.31</td>
<td>8.63***</td>
<td>0.10</td>
<td>0.93</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>0.93</td>
<td>3</td>
<td>0.31</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>6.21</td>
<td>6</td>
<td>1.04</td>
<td>2.01</td>
<td>0.07</td>
<td>0.72</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>1.12</td>
<td>3</td>
<td>0.37</td>
<td>0.56</td>
<td>0.01</td>
<td>0.16</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>81.90</td>
<td>127</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *$p<.10$, **$p<.05$, ***$p<.01$, ****$p<.001$.

With a Greenhouse-Geisser Correction.
The significant time by type by level interaction for the target phone [β] suggested that after instruction (either +PI or –PI) first and second year learners improved their discrimination of [β]-[b] most, though only second year learners retained (meager) gains in the delayed posttest. The third year learners, on the other hand, improved their production of [β] more than the other skills. Pretest identification scores for [β] were quite high (93% accurate), which may have limited opportunities for growth.

As for [ð], instruction (either +PI or –PI) was clearly more beneficial for the perceptive skills than productive skills, though retention in discrimination and identification gains from posttest to delayed posttest declined sharply.
Table 4.14 RMANOVA of Gain Scores on Three Test Types: [0]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.17</td>
<td>1</td>
<td>0.17</td>
<td>0.73</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Level</td>
<td>0.01</td>
<td>2</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>1.06</td>
<td>2</td>
<td>0.53</td>
<td>2.27</td>
<td>0.05</td>
<td>0.45</td>
</tr>
<tr>
<td>Error</td>
<td>19.05</td>
<td>82</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>9.52</td>
<td>1</td>
<td>9.52</td>
<td>15.98</td>
<td>0.16</td>
<td>0.98</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.84</td>
<td>1</td>
<td>0.84</td>
<td>1.40</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>Time X Level</td>
<td>3.69</td>
<td>2</td>
<td>1.84</td>
<td>3.09</td>
<td>0.07</td>
<td>0.58</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>0.27</td>
<td>2</td>
<td>0.14</td>
<td>0.23</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>48.86</td>
<td>82</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>60.09</td>
<td>2</td>
<td>33.05</td>
<td>16.93*</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>9.16</td>
<td>2</td>
<td>5.04</td>
<td>2.58*</td>
<td>0.03</td>
<td>0.48</td>
</tr>
<tr>
<td>Type X Level</td>
<td>7.84</td>
<td>4</td>
<td>2.16</td>
<td>1.10</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>0.57</td>
<td>4</td>
<td>0.16</td>
<td>0.08</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>291.07</td>
<td>149</td>
<td>1.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>11.78</td>
<td>2</td>
<td>6.13</td>
<td>10.69***</td>
<td>0.12</td>
<td>0.99</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>1.14</td>
<td>2</td>
<td>0.59</td>
<td>1.05</td>
<td>0.01</td>
<td>0.22</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>2.16</td>
<td>4</td>
<td>0.56</td>
<td>0.98</td>
<td>0.02</td>
<td>0.30</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>2.23</td>
<td>4</td>
<td>0.58</td>
<td>1.01</td>
<td>0.02</td>
<td>0.31</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>90.32</td>
<td>158</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

With a Greenhouse-Geisser Correction

Table 4.15 Pairwise Comparisons of Three Test Types: [8]

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>1.88</td>
<td>1</td>
<td>1.88</td>
<td>0.83</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>35.08</td>
<td>1</td>
<td>35.08</td>
<td>20.00***</td>
<td>0.20</td>
<td>0.99</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>53.18</td>
<td>1</td>
<td>53.18</td>
<td>40.99***</td>
<td>0.33</td>
<td>1.00</td>
</tr>
<tr>
<td>Time X Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>35.47</td>
<td>1</td>
<td>35.47</td>
<td>15.07***</td>
<td>0.16</td>
<td>0.97</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>35.20</td>
<td>1</td>
<td>35.20</td>
<td>19.95***</td>
<td>0.20</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

Figure 4.7 Gain Scores for [8]

![Figure 4.7 Gain Scores for [8]](image-url)
The type by level interaction for the target phone [ɣ] suggested that only second year leaners benefited from instruction in terms of improving their production of [ɣ].
whereas first and third year learners actually got worse following instruction. As for the perceptive skills, first year learners gained and retained more in identification of [ɣ] after instruction as compared to second and third year learners, who improved their discrimination of [ɣ]-[g] more than their identification of [ɣ].

Table 4.18 RMANOVA of Gain Scores on Three Test Types: /ɣ/

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.43</td>
<td>1.00</td>
<td>0.43</td>
<td>0.35</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Level</td>
<td>3.33</td>
<td>2.00</td>
<td>1.66</td>
<td>1.36</td>
<td>0.03</td>
<td>0.29</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>0.60</td>
<td>2.00</td>
<td>0.30</td>
<td>0.25</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>Error</td>
<td>100.61</td>
<td>82.00</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.11</td>
<td>1.00</td>
<td>0.11</td>
<td>0.19</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Time X Level</td>
<td>0.21</td>
<td>2.00</td>
<td>0.10</td>
<td>0.18</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>1.05</td>
<td>2.00</td>
<td>0.52</td>
<td>0.92</td>
<td>0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>46.66</td>
<td>82.00</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>22.09</td>
<td>1.89</td>
<td>11.72</td>
<td>9.04***</td>
<td>0.10</td>
<td>0.97</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>5.45</td>
<td>1.89</td>
<td>2.89</td>
<td>2.23</td>
<td>0.03</td>
<td>0.44</td>
</tr>
<tr>
<td>Type X Level</td>
<td>1.18</td>
<td>3.77</td>
<td>0.31</td>
<td>0.24</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>3.86</td>
<td>3.77</td>
<td>1.02</td>
<td>0.79</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>Error (Type)</td>
<td>200.50</td>
<td>154.55</td>
<td>1.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>4.36</td>
<td>1.90</td>
<td>2.30</td>
<td>4.77**</td>
<td>0.06</td>
<td>0.77</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>2.23</td>
<td>1.90</td>
<td>1.18</td>
<td>2.44*</td>
<td>0.03</td>
<td>0.47</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>2.04</td>
<td>3.79</td>
<td>0.54</td>
<td>1.12</td>
<td>0.03</td>
<td>0.34</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>0.67</td>
<td>3.79</td>
<td>0.18</td>
<td>0.36</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Error (Time X Type)</td>
<td>74.95</td>
<td>155.41</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001
With a Greenhouse-Geisser Correction

Table 4.19 Pairwise Comparisons of Three Test Types: /ɣ/

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type: Discrimination - Identification</td>
<td>9.91</td>
<td>1</td>
<td>9.91</td>
<td>16.32***</td>
<td>0.17</td>
<td>0.98</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>2.19</td>
<td>1</td>
<td>2.19</td>
<td>1.90</td>
<td>0.02</td>
<td>0.28</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>0.04</td>
<td>1</td>
<td>0.04</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Time X Type

| Discrimination - Identification | 2.41| 1 | 2.41| 1.97| 0.02 | 0.28 |
| Discrimination - Production    | 6.51| 1 | 6.51| 5.11***| 0.06 | 0.61 |
| Identification - Production    | 16.83| 1 | 16.83| 11.95****| 0.13 | 0.93 |

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001
The results for /ɾ/ suggested that instruction was less beneficial for discrimination than the other skills, though learners’ high pretest accuracy in discriminating [ɾ]-[ɾ] (89%) limited opportunities for growth. Learners furthermore improved their identification of /ɾ/ steadily over time, with higher gains from pre to delayed posttest than pre to immediate posttest. Similarly, learners were highly accurate at discriminating [ɾ]/[ɾ] in the pretest (99%), so it is not surprising that they did not improve in this regard. The time by type interaction suggested that learners improved their identification of /ɾ/ immediately following instruction but did not retain any of those gains three weeks later.
Table 4.20 RMANOVA of Gain Scores on Three Test Types: /r/

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>2.38</td>
<td>1.00</td>
<td>2.38</td>
<td>2.33</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>Level</td>
<td>2.13</td>
<td>2.00</td>
<td>1.06</td>
<td>1.04</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>Condition X Level</td>
<td>2.98</td>
<td>2.00</td>
<td>1.49</td>
<td>1.46</td>
<td>0.03</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>83.93</td>
<td>82.00</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>3.68</td>
<td>1.00</td>
<td>3.68</td>
<td>7.82***</td>
<td>0.09</td>
<td>0.79</td>
</tr>
<tr>
<td>Time X Condition</td>
<td>0.01</td>
<td>1.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Time X Level</td>
<td>0.27</td>
<td>2.00</td>
<td>0.14</td>
<td>0.29</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>Time X Condition X Level</td>
<td>1.03</td>
<td>2.00</td>
<td>0.51</td>
<td>1.09</td>
<td>0.03</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Error (Time)</strong></td>
<td>38.63</td>
<td>82.00</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>1.92</td>
<td>1.57</td>
<td>1.22</td>
<td>0.96</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Type X Condition</td>
<td>2.13</td>
<td>1.57</td>
<td>1.36</td>
<td>1.07</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td>Type X Level</td>
<td>4.35</td>
<td>3.15</td>
<td>1.38</td>
<td>1.09</td>
<td>0.03</td>
<td>0.30</td>
</tr>
<tr>
<td>Type X Condition X Level</td>
<td>1.60</td>
<td>3.15</td>
<td>0.51</td>
<td>0.40</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Error (Type)</strong></td>
<td>163.26</td>
<td>128.96</td>
<td>1.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time X Type</td>
<td>8.44</td>
<td>1.80</td>
<td>4.69</td>
<td>9.83****</td>
<td>0.11</td>
<td>0.97</td>
</tr>
<tr>
<td>Time X Type X Condition</td>
<td>0.98</td>
<td>1.80</td>
<td>0.55</td>
<td>1.15</td>
<td>0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>Time X Type X Level</td>
<td>1.26</td>
<td>3.60</td>
<td>0.35</td>
<td>0.74</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>Time X Type X Condition X Level</td>
<td>3.55</td>
<td>3.60</td>
<td>0.98</td>
<td>2.06*</td>
<td>0.05</td>
<td>0.57</td>
</tr>
<tr>
<td><strong>Error (Time X Type)</strong></td>
<td>70.41</td>
<td>147.64</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

With a Greenhouse-Geisser Correction

Table 4.20 Pairwise Comparisons of Three Test Types: /r/

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time X Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination - Identification</td>
<td>14.04</td>
<td>1</td>
<td>14.04</td>
<td>4.21**</td>
<td>0.05</td>
<td>0.53</td>
</tr>
<tr>
<td>Discrimination - Production</td>
<td>0.04</td>
<td>1</td>
<td>0.04</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Identification - Production</td>
<td>15.53</td>
<td>1</td>
<td>15.53</td>
<td>6.87***</td>
<td>0.08</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Significant at *p<.10, **p<.05, ***p<.01, ****p<.001

Figure 4.10 Gain Scores for /r/
Several patterns emerged in this analysis. As displayed in Tables 4.4-4.20, there were significant main effects of time and test type as well as an interaction between time and type for nearly all target phones, whereas other main effects and interactions were infrequent or absent. The main effect of time, not surprisingly, reflected the fact that gain scores tended to be higher immediately following instruction and lower three weeks after instruction in the delayed posttest. The main effect of test type most often (with the aggregate as well as with [p, t, k, δ and υ]) reflected the fact that learners’ improvement in production was smaller than their improvement in one or both perceptive tasks. The time by test type interaction most often (with the aggregate as well as with [t, k, β δ and υ]) related to the fact that learners’ improved discrimination abilities declined more sharply from posttest to delayed posttest than the other skills. The advantage learners gained post-instruction in terms of discrimination appeared to be relatively more fleeting.

In sum, then, instruction led to more improvement in learners’ perception of the target phones than their production. Effect sizes for the main effect of test type ranged from 9-33%. However, the improvement occurred in both instructional conditions (+PI and –PI). Though there was some evidence that type of instruction may have played a role, condition did not represent a significant main effect for any phone, and the only significant interaction with condition was the type by condition interaction found for /p/. Thus the effect of both the +PI and –PI instructional conditions was similar in terms of their capacity for changing learners’ performance on a particular experimental task (discrimination, identification, or production) more or less than other tasks. Matriculation level was likewise not a significant factor.
4.3.5 RM ANCOVA Analysis of Production Scores

Given that the SLM and other theories of L2 speech acquisition posit that perception leads production (as discussed in §1.2) and in light of the meager improvements in production of target phones after instruction (as reported in §2.3) relative to the improvements in perception (as reported in §3.3), one last analysis of the data was performed in order to address the third research question of the present chapter, which asked how learners’ perceptive abilities impacted the effect of instruction on learners’ production. In this analysis, the production data for individual target phones as well as the aggregate were again submitted to RM ANCOVA, but this time with perception scores as covariates. As before, the within-groups factor was time of test and the between-groups factors were instructional condition and matriculation level.

This RM ANCOVA analysis was done in six different parts. First, learners’ discrimination pretest scores were treated as a single covariate. Then learners’ identification pretest scores were treated as a single covariate. Then both the discrimination and the identification pretest scores were used as covariates in the same model. The pretest scores were used to approximate learners’ preexisting perceptive abilities, prior to instructional influence. In the final three parts of the analysis, the discrimination gain scores (from pretest to posttest) were treated as a covariate, followed by identification gain scores, and finally both perception gain scores together. The gain scores were used as a measure of the improvement in learners’ perception after receiving instruction. In each of these six analyses, only the perception scores relevant to a particular target phone were considered. So, for instance, in the RM ANCOVAs on
learners’ production of /p/, only the discrimination and/or identification pretest and gain scores relating to items targeting /p/ were considered as covariates.

The results of the analysis suggested a clear and consistent pattern. For the sake of brevity, the full results will not be reported for each phone. Recall that the RMANCOVAs without covariates (as reported in §2.3.5) indicated a significant main effect of time for all target phones except [β]. In contrast, when submitted to RMANCOVAs with perception scores as covariates, the data did not exhibit any main effect of time. The main effect for time was lost for most phones in all six RMANCOVAs, i.e., when discrimination and/or identification pretest scores or gain scores were treated as covariates. The main effect of time was still significant even with covariates only in three cases: the aggregate of eight phones with discrimination gain scores as covariate and /p/ and /ɾ/ with discrimination and/or identification gain scores as covariates.

The results of the RMANCOVAs suggested that learners’ perception was related to their change in production of the target phones after instruction of either type (+PI or –PI), in at least two ways. First, when perception pretest scores were employed as covariates, thereby controlling for the differences in learners’ preexisting perceptive abilities, instruction did not appear to have an effect on learners’ production of the target phones. Second, when perception gain scores were employed as covariates, thereby controlling for the increases in perceptive capacities over time, instruction did not appear to have an effect on learners’ production of the target phones. The results suggested that learners’ perception of the target phones, before instruction as well as after instruction, influenced the capacity of learners to apply what they learned from either type of
instruction to their production of the target phones. What had appeared on the surface as improvement in productive abilities after instruction was perhaps related to underlying changes in perception.

4.4 Discussion

4.4.1 Summary of Results

The first research question asked if learners’ ability to perceive L2 phones was related with their ability to produce them. The results partially confirmed the hypothesis that learners’ perceptive and productive abilities would be correlated, because correlations were found for some but certainly not all phones. The second research question asked if L2 phonetics instruction was equally effective for improving learners’ perception and production. The results partially confirmed the hypothesis that instruction would have a greater impact on perception than production. Post-instruction, learners made more gains on the perception tests than the production test, the effect size of test type being 9-33%. However, this result was found for learners in both instructional conditions, so the result was not specific only to instruction in phonetics and phonology but rather was an effect of instruction more generally. The third research question asked how learners’ perceptive abilities impacted the effect of instruction on learners’ production. The results confirmed the hypothesis that perceptive abilities had an impact, both prior to instruction as well as after instruction. When perceptive abilities were held constant, the significant main effect for time on learners’ production scores disappeared.
4.4.2 Implications

The hypothesis that there would be a link between perception and production was based on the theoretical link between the two abilities and the empirical evidence of a correlation between the two, at least when the perception tasks employed were “offline” and the targets were contrastive L2 phones. Indeed, the correlation analysis of the present study suggested that while the link between the two perception tests was more tenuous, learners’ (pre-instruction) performance on the production test was positively correlated with their performance on the perception tests, particularly the identification test. This positive relationship was found in relation to various target phones, and at least one phone for learners at each matriculation level, though the positive correlation was not exhibited uniformly across the phones. The data thus supported the findings of previous research indicating that a correlation between perception and production is expected when offline perception tasks are employed, such as identification or discrimination with long ISIs (Peperkamp & Bouchon, 2011). The underlying relationship may be that offline tasks allow learners to rely on the phonological loop, specifically subvocal rehearsal of the stimuli in the phonological loop, as Peperkamp and Bouchon (2011) suggested. Ostensibly the subvocal rehearsal mechanism relies on the same long-term phonological knowledge that supports vocalization, thereby explaining the link between perception and production.

These results begged the question of why significant correlations were not found across all tasks and all phones, even if non-significant correlations should not be not interpreted as evidence that the link between perception and production is broken (Munro, 2008). It is likely that a variety of intervening factors obfuscated the perception-
production relationship for particular target phones. For instance, articulatory constraints likely impinged upon learners’ production of /r/ and may have affected learners’ performance on the production test in different ways for other phones (Colantoni & Steele, 2008). Obviously, the very high discrimination pretest scores for rhotic phones impeded calculation of some correlations for those phones.

Interestingly, the only significant correlations found for the aggregate eight-phone test was between identification and production. This correlation obtained for first and second year learners but was strongest for first year learners. The strong correlation between identification and production may be due to the nature of the task. Identification is more “offline” than discrimination generally, according to Peperkamp and Bouchon (2011). However, positive correlations were not found between identification and production of all phones. Differences in the nature of grapheme-phoneme correspondence across the phones, and between learners’ L1 (English) and L2 (Spanish) may have impinged upon learners’ performance on the identification test. Use of orthography may have made the identification task particularly problematic for judging phoneme identification of [β] and /ɾ/, which exist in English but are represented with different graphemes in English and Spanish. The scoring method developed for the identification test attempted to avoid conflating knowledge of Spanish spelling with the correct identification of target phones. Nonetheless, it may have been the case that an incomplete or unautomatized knowledge of Spanish grapheme-phoneme correspondences negatively affected the ability of some learners to provide a correct response on the identification test even when they had correctly identified the target phone in the speech stimuli.
It was noted that there were major differences in task demands and conditions between these three tasks. The perception tests were timed, whereas the production test was not. The identification and production tests, but not the discrimination test, were mediated through written text. The discrimination was a two-alternative, forced-choice test. In comparison, the identification and production tests had no limit to the number of possible responses, but scoring of the identification test was on a scale of 0-1, whereas the production test scoring scale allowed for more options (0-3). The tests were designed to measure leaners’ abilities in three separate behaviors (discriminating the sounds in two syllables, identifying a particular sound in a nonce word, and producing a particular sound in a real word) and, to the extent possible, isolate each ability and provide rich data. However, because the tasks were necessarily so dissimilar in terms of processing demands and conditions, direct comparisons across tasks were difficult (Llistertri, 1995).

One might ask if the differences in gain scores were related to task practice effects. Instructed learners were probably familiar with reading words aloud and perhaps taking dictation, though perhaps less familiar with taking dictation of unfamiliar words. Ostensibly learners had no experience with AX discrimination tasks. One might contemplate whether learners’ perception pretest scores were lower due to this relative lack of familiarity and, consequently, whether their perception posttest scores were higher due to the practice they received via the pretest, in contrast to their production scores, which remained more stable due to their preexisting familiarity with the task. This interpretation was discarded in light of the results reported in chapters 2 and 3, in which there were significant interactions between instructional group and improvement
over time. These group differences suggested that practice effects alone could not explain gains in the perception tests.

All the previous research on L2 perception reviewed in the introductory chapter focused on L2 phonemic contrasts. Segmental contrasts, especially the English /x/-/l/ contrast, are particularly well studied, and suprasegmental contrasts, such as Chinese lexical tone, are also well documented. To the author’s knowledge, no prior study has compared multiple perception and production tasks in an examination of multiple L2 segments that are non-contrastive in the L2. Therefore the results of the current study are relatively novel, and the interpretations put forth here are necessarily speculative at this point. The stimuli of the discrimination test were chosen based on the prediction of the Speech Learning Model (SLM) that learners must first learn to discern the subtle phonetic differences between L2 phones and analogous L1 phones in order to acquire categories suited to the L2 and, eventually, produce the L2 phones without noticeable accent.

The present study included a variety of phones and a range of tasks in order to assess learners’ abilities at each of the levels of phonological acquisition laid out in the SLM in an effort to present a broad, comprehensive view of the effect of instruction. This all-encompassing approach was considered optimal for a study that purported to make pedagogical recommendations. On the other hand, the experimental design was unfortunately not well suited to provide rich data about any particular phone or task. For that matter, the study was not designed to provide nuanced data about the performance of any individual learner’s L2 developmental stage. To some extent the particulars were
lost in an effort to detect larger patterns, even though a complete understanding of the
data is contingent on understanding the particulars of individual phones.

As a case in point, there was no significant correlation between tasks for the
alveolar tap /ɾ/. Learners’ discrimination of the (L2) /ɾ/ - (L1) /ɹ/ distinction was at
ceiling level in the pretest. In fact, second and third year learners discriminated the
phones so well that their scores were constant, and a correlation between discrimination
and the other tasks could not be performed. Conceivably, a learner could discriminate the
/ɾ/ - /ɹ/ distinction perfectly yet perform poorly on the other tasks. The learner could have
moderate difficulty identifying /ɾ/ in the speech stream, perhaps identifying it correctly
except when in intervocalic and unstressed position, as many learners in the current study
did. The same learner could produce /ɾ/ with a noticeable accent, scoring very low on the
production test. For this learner, then, there would be no evidence of a significant
perception-production correlation (or even a discrimination-identification correlation).
Yet absence of evidence would not be evidence of absence. There could have been a
correlation present at the initial stage of acquisition that decreased as the learner
developed knowledge of /ɾ/, proceeding through the developmental stages postulated in
the SLM. Thus high perception scores could co-occur with low production scores.

The only result that would contradict the predictions of the SLM would be high
production scores co-occurring with low perception scores. The present study was not
designed to test such theoretical predictions, because it was concerned more with overall
trends and assessing the effectiveness of pedagogical techniques. Nonetheless, it is
reassuring that positive correlations did obtain for many phones and learners. The
correlations lent credence to the claim that the experimental tasks indeed tapped the
abilities they were supposed to tap. The more important result of the correlation analysis, however, was to bolster the assertion that improving either perception or production through instruction could potentially benefit both skills, because they were related. The degree of a possible transfer effect from one skill to another and the timing of the transfer, whether simultaneous or delayed, remained an empirical question. Thus a series of analyses was subsequently undertaken to assess the relative response of the skills to the instructional treatment.

The results suggested that after instruction learners improved their perception of the target phones more than their production of those same phones, generally regardless of instructional condition (+PI and –PI). Prior studies carried out in L2/FL classrooms suggested that phonetics instruction lead to small improvements in learners’ accent (e.g., Pardo, 2004). The few studies that addressed English-speaking learners of Spanish reported generally positive effects for phonetics instruction, though the effects varied widely across target phones (e.g., Elliott, 1995, 1997). If the current study’s results are any indication, those researchers may have found greater instructional effects had they measured learners’ perception. The data suggested that learners, at least relatively novice learners in their first, second and third year of college Spanish study, were more malleable in terms of perception than production for most phones. The data thus complimented the findings of previous laboratory-based research indicating that training in production lead to improvement in perception and vice versa but that training was most effective for perception (e.g., Lacabex, García Lecumberri & Cooke, 2009). The data also concurred with previous studies in that some L2 phones responded to treatment more
than others. Specifically, approximants were more resistant to instruction than other phones (e.g., Elliott, 1995, 1997). However, the lack of significant interactions and main effects related to the factor of instructional condition for most phones indicated that the assertions made above also applied to the more implicit instructional condition. The –PI group completed a focused listening with dictation activity. They heard native repetitions of the target phones and engaged in production practice with feedback in amounts roughly equivalent to that of the +PI group. Under those conditions, the –PI group performed similarly to the +PI in terms of gain scores over time across the three experimental tasks. The type by condition interaction reached significance for just one phone, suggesting that the +PI group improved identification of /p/ more than discrimination of /p/. The type by condition by time approached significance for the aggregate of phones, which represents some unconfirmed evidence that type of instruction may have been related to the relative gain scores across the three tasks, yet the provisional conclusion drawn from these data was that either type of instruction resulted in more improvement in perception than production. Whether learners were exposed to the target phones through a relatively more implicit task or rather engaged in explicit learning of Spanish phonetics and phonology, they tended to experience the same changes in the various perception and production tests.

Though both instructional conditions were associated with more improvement in perception than production, it should also be noted that for several phones [t, k, β δ and γ] and the aggregate, learners’ large posttest gain in discrimination was followed by a relatively steep decline three weeks later. Instruction seemed to have an immediate
impact that was not well retained. Arguably the discrimination test, even with its long ISI, favored more automatic processing than the other tests. Learners may revert to their automatized perceptual routines (influenced by the L1) without more sustained instruction.

Furthermore, the lack of significant interactions and main effects related to the factor of matriculation level for most phones indicated that the assertions made above were not limited to learners of any one particular experience level. The data suggested that matriculation level played a significant role for just two phones ([β] and [ɣ]) but overall was not a significant factor.

The final set of analyses, which employed learners’ scores on the perception tests as covariates, suggested that preexisting perceptive abilities as well as changes in perception over time were related to changes in learners’ production. While the SLM and other models suggest that perception precedes production and empirical data from laboratory studies suggests that intensive training leads to cross over effects in perception and production, no study to the author’s knowledge has analyzed the impact of perception on changes in learners’ production after explicit phonetics instruction. These data suggest that perception abilities are important to consider when measuring effectiveness of pronunciation instruction. What on the surface appeared to be improvement in pronunciation following instruction when pronunciation was analyzed in isolation later seemed rather an effect that reflected underlying changes in learners’ perception.
4.4.3 Conclusion: A Role for Individual Differences?

Researchers in the field of SLA have long questioned to what extent various individual differences between learners best predict their rate of learning, ultimate attainment, and response to particular instructional conditions. Age of learning has figured prominently in the literature on L2 phonological acquisition, both for accent generally (see Flege, 2003a, for a review) and for particular late acquired segments (e.g., Shively, 2008). Other individual differences are relatively less studied in the area of L2 phonology, but a few studies have suggested that attitude (Elliott, 1995), short-term memory (Munro, 2008), and phonological awareness (Kennedy & Trofimovich, 2010) might be influential. Thus no examination of instructional effectiveness would be complete without at least some discussion of how learner differences interacted with the instruction in question.

As a means of concluding this final analytic chapter, individual differences will be considered as variable that could predict the outcomes of the two instructional interventions (+PI and –PI). The results of the analysis were not particularly illuminating and are merely summarized here. The factors considered were: age of onset of Spanish learning, current age, gender, phonetic encoding ability, phonological short term memory, attitude towards pronunciation, number of high school and university courses taken, number of high school and university Spanish teachers who were native speakers of Spanish, time spent speaking Spanish outside of class, and time spent abroad in Spanish speaking regions. Phonetic encoding ability was measure with the phonetic script learning subset of the Modern Language Aptitude Test (Carroll, 1962). Phonological short-term memory was measured with two non-word repetition tasks, one
in Spanish and one in English (taken from Lado, 2008). Attitude was measured with Elliott’s Pronunciation Attitude Inventory (Elliott, 1995). All other variables were calculated from the information collected on the background questionnaire.

These individual difference variables were entered into a series of simple multiple regression analyses with each of the following dependent variables: gains (from pretest to posttest and from pretest to delayed posttest) made by learners in the +PI and –PI groups, in total as well as on each experimental test and for each phone separately. The results of the 48 regression analyses suggested that individual difference factors were not good predictors of gains, separately or in combination. Most of the models were not significant ($p > .05$) and had adjusted $r^2$ values near zero. Within the four models that were significant, the only individual difference variable that was significant and appeared to have a substantive effect was current age (coefficient -.96), which was negatively related to –PI learners’ total gains from pretest to posttest. That is, the older a participant was at the time of the study, the less likely he was to improve overall following exposure to the target phones through the focused listening with dictation activities. It was not surprising to find that age was related to rate of implicit learning.

Based on previous research (see for example Robinson, 2002), a full discussion of which is beyond the scope of this study, it had been assumed that individual differences would interact with the instructional conditions in more complex and meaningful ways, particularly the more implicit condition (dictation). However, each individual learner controlled the pacing of the dictation, and this instructional model may have made some individual differences irrelevant. For instance, differences in phonological short-term memory would become meaningful only if the treatment actually taxed learners’ memory
store. Yet learners in this study could pause the speech stream when they desired. Learners likely broke the speech in the dictation into chunks that were of a length that they could easily hold in their phonological store. Other possible interpretations are that the instructional treatments may simply have been too brief for learner differences to make a difference, the wrong set of factors may have been considered, or the measurements of those differences may have been too coarse. Other individual difference factors not considered but potentially predictive of learning are phonological awareness (see Venkatagiri & Levis, 2007) and pre-existing knowledge of Spanish phonetics and phonology.

Nonetheless, it was encouraging to conclude that the results and implications of the preceding analyses could likely be extended equally to learners of various aptitude, attitude, and language exposure backgrounds. It was not the case that only learners with particularly high phonological short-term memory capacity, for instance, could learn from either instructional condition. The lack of significant results regarding individual difference factors was indeed considered a positive outcome for this investigation of instructional effectiveness.
CHAPTER 5: CONCLUSION

5.1 Summary of Results

The first research question asked whether explicit instruction in L2 phonetics and phonology would improve Spanish FL learners’ ability to pronounce particularly problematic Spanish segments. The data (presented in Chapter 2) suggested that the effect of this brief instructional intervention was generally positive but quite small, and it was no greater than an instructional alternative (focused listening with dictation) for any target phone individually or for all the target phones analyzed together.

The second research question asked if instruction in L2 phonetics and phonology would improve learners’ ability to discriminate L2 phones and their analogous L1 phones. Though the instructional effect could not be confirmed for most phones when analyzed individually, the data (presented in Chapter 3) suggested that for the aggregate of phones, explicit instruction improved learners’ discrimination ability more than the instructional alternative.

The third research question asked if instruction in L2 phonetics and phonology would improve learners’ accuracy in identifying L2 phones in the speech stream. The data (presented in Chapter 3) suggested that it improved learners’ identification accuracy immediately following instruction but was not more beneficial than the instructional alternative in the longer term. The role of learners’ matriculation level was addressed in all the preceding analyses, and the data suggested that for all tests and most phones, matriculation level was not predictive of learning from the phonetics instruction.
In Chapter 4 it was reported that learners’ perception and production were correlated for several of the target phones. Theoretical arguments and empirical evidence were presented to support the claim that perception likely precedes production and, at a minimum, is closely linked to production. Chapter 4 also presented analyses demonstrating that phonetics instruction (as well as the instructional alternative) honed perception more than it honed production and that learners’ gains in perception may have been underlying their gains in production all along.

5.2 Pedagogical Implications

The central claim of this study, that phonetics instruction honed perception more so than production, should not be interpreted as a dismissal of the putative benefits of phonetics instruction. To the contrary, it was encouraging that the brief instructional interventions were shown to improve learners’ perception of the target phones. It is likely that these learners were developmentally poised to attune precisely their perceptive skills, and that this attunement would facilitate later development in production, at least for some phones. Likewise, the lack of a main effect for instructional condition in production scores did not support a rejection of explicit phonetics instruction wholesale but rather indicated that several features of phonetics instruction, including exposure to native speech, pronunciation practice and feedback, were helpful for developing learners’ productive skills. These features could be incorporated into other instructional methods.

In the typical undergraduate FL curricular sequence of American universities, study of phonetics and phonology is relegated to a third or fourth year course for language majors and minors (Correa, 2011). Those courses typically aim to improve
learners’ pronunciation while they teach students about theories of sound production and perception as well as Spanish dialectology (Correa, 2011). Celce Murcia, Brinton & Goodwin (1996) characterized phonetics instruction as typically having two pedagogical foci. The first, intuitive-imitative in nature and focused on making learners’ pronunciation more target-like, might be best suited for beginning learners (Brown, 1995; Celce Murcia et al., 1996). The second, more analytic and explicit in nature and focused on analyzing linguistic features of the language in detail, might be best suited for advanced learners (Brown, 1995; Celce Murcia et al., 1996).

The present study provided first, second and third year FL learners with instruction that included a highly explicit but brief introduction to Spanish phonetics and phonology and emphasized imitation of target-like pronunciation, which presumably was well suited to most of the learners recruited. The study compared the learning effects of those explicit lessons with an alternative that could be considered more “intuitive” but was otherwise comparable in terms of time on task, exposure to the target phones, pronunciation practice and feedback. Table 5.1 summarizes the findings for each target phone. Check marks denote evidence of learning after instruction. It is important to note that “instructional effect” here means simply an effect that reached statistical significance. Effect sizes were actually quite small, indicating that these brief instructional interventions were certainly no panacea for learners’ problems in perceiving and producing the target phones.
Table 5.1 Summary of Findings

<table>
<thead>
<tr>
<th></th>
<th>Discrimination</th>
<th>Identification</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate (6)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>/p/</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>(77%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/t/</td>
<td>✓ (76%)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>/k/</td>
<td>✓ (43%)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>[β]</td>
<td>✓ (53%)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>[θ]</td>
<td>✓ (66%)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>[υ]</td>
<td>✓ (49%)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>/c/</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(72%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/γ/</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(70%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Pretest accuracy below 80% on perception tests.
(2) Significant main effect for time coupled with increase in accuracy for both +PI and -PI conditions across time.
(3) Significant difference between pretest and posttest.
(4) Significant difference between pretest and delayed posttest (3 weeks after instruction).
(5) Significant time by instructional condition interaction coupled with the +PI group experiencing greater gain in accuracy.
(6) Includes all 8 target phones for identification and production tests, but only [t, k, β, θ, υ] for discrimination test.
(7) Pretest production significantly different from NSs.
Several pedagogical implications can be drawn from these results. It would appear that intermediate Spanish learners are not likely to benefit from practice in discriminating between target-like and English-accented pronunciations of Spanish /p/ and rhotic phones, as the learners in the current study were already quite good at discriminating these pairs in the pretest. They were more likely to improve their pronunciation of these phones after instruction. On the other hand, the learners in the present study improved their perception of the approximants after instruction but did not improve their pronunciation of approximants, a result that concurred with previous research on the acquisitional patterns of perception and production of Spanish approximants (González-Bueno & Quintana-Lara, 2010; Hurtado & Estrada, 2010; Shively, 2008). These data suggest that the developmentally appropriate instructional focus for intermediate learners of Spanish might be to attune their perception (rather than production) of the approximant phones, through explicit phonetics instruction and/or more implicit instructional alternatives.

Finally, practitioners should note that the explicit component of pronunciation instruction, i.e., the teaching of Spanish phonetics and phonology proper, may be less effective than previously thought for bringing about changes in learners’ pronunciation, whereas it may be more effective than previously realized in in terms of changing learners’ perception, at least in the short term. FL teachers are perpetually beleaguered with insufficient time and resources to provide all the instruction that could be potentially beneficial for learners’ language development. Empirical studies such as this one can help teachers and learners chose more efficient techniques for reaching their goals. If the goal is to achieve more target-like pronunciation of Spanish consonantal segments, then
exposure to native speech with a focus on form (particular phones) and pronunciation practice with feedback are likely to help, whereas brief phonetics lessons may provide very little added benefit.

5.3 Limitations

The present study had several limitations, perhaps the most consequential of which was its limited inventory of phonological targets. The eight target phones were selected because they were extensively researched in the Spanish SLA literature, but they were all consonantal segments and therefore represented just one small part of learners’ phonological competence. The range of phonological environments included was also limited. For instance, learners likely would have performed differently when pronouncing the segments in consonant clusters (Colontoni & Steele, 2006). Future studies should include vowels and suprasegmentals. It has been rightly argued that pronunciation teaching must expand beyond the segmental level (Pardo, 2004; Pennington & Richards, 1986) and that segmentals are not as important as suprasegmental features and even more “macroscopic things” like volume (Derwing & Munro, 2009). It is possible that shifting the focus of the explicit phonetics instruction from the segment to suprasegmental features, like prosody, could have been even more beneficial to learners’ segmental production (Missaglia, 1999).

Conversely, while the inclusion of multiple target phones in the present study was seen as an attribute, given the concern of pedagogical relevance, so many phones were included in the study that no more than five tokens of each phone could be included in any given experimental test, because class time was limited. The restricted number of
test items reduced the power of the statistical analysis to the point that definitive claims could not be made about the effect of instruction on particular phones. Thus the present study could not evaluate claims such as those presented in Lively et al. (1994, p. 2076) that contrasts that depend on spectral cues (e.g., level of closure in approximants) are more difficult to learn than temporal cues like voice onset timing. There were also clearly many limitations relating to the experimental tasks themselves, which have been discussed in previous chapters (§2.4.3 and §3.4.3).

In order to begin to address the issue of curricular sequencing, classes at three curricular levels were included in the study, roughly beginners (first year), low intermediate (second year), and high intermediate (third year). L2 experience was operationalized as course matriculation. However, course matriculation is clearly not a perfect measure of L2 proficiency or exposure to L2 phonology. The study did not include very advanced learners or true beginners, as almost all students reported having studied Spanish in high school.

The delayed posttest in the present study took place only 3 weeks after the last instructional module was completed, which was the maximum length of time available given the limitations of a short college semester and the schedules of multiple participating classes. Thus the “delayed” posttest in the current study was, in terms of post-treatment timing, more akin to the posttest of studies that administered their posttest at the end of a semester of instruction. A better measure of the effects of instruction over time would be given six months or more after treatment (Elliott, 2003).
5.4 Contribution and Future studies

To the author’s knowledge, the present study represented the first attempt to document the effect of classroom-based explicit phonetics instruction on Spanish learners’ production and perception. Given that theories of L2 speech acquisition posit an intimate link between perception and production, more research should endeavor to understand how instructional interventions might improve learners’ perception in addition to their production. Such future research should also endeavor to improve on the methods of previous studies that have been unduly biased, in the author’s opinion, towards finding positive effects of instruction. The present study was a first attempt to include a more balanced control group so as to assess the effectiveness of explicit lessons in phonetics and phonology while controlling for other aspects of pronunciation instruction that are putatively beneficial, including exposure to target sounds and practice. In this regard it contributed to the growing body of SLA that explores the relative advantages of more explicit versus more implicit instructional conditions, a body of research which to date has paid less attention to phonology than other aspects of language.

The experimental tasks utilized in the present study suffered from many limitations, making it clear that more work must be done to develop experimental tasks for measuring the perception of L2 phones that are not contrastive phonemes. Obviously L2 phonemic contrasts are of great importance to L2 learners, but the non-target like pronunciation of non-contrastive L2 phones also contributes to perceived accent and is of concern to L2 learners and teachers. Yet the experimental methods available for addressing their acquisition are less than ideal.
Other areas ripe for future research include the role of instruction for the phonological development of true L2 beginners, the role that individual difference factors might play in interacting with more extensive pronunciation instruction, and more longitudinal investigations. The current study provided learners with a relatively brief introduction to Spanish phonetics. The time allotted was sufficient to impart explicit phonetics instruction, which was the focus of the study, and learners reported that the time allotted was appropriate. However, it remains an empirical question whether more intensive, continuous and/or repeated exposure to explicit phonetics instruction would produce results that replicate or contradict the results reported here.

Finally, the present study proposed two broad questions for future studies in the area of L2 speech. The first was how to determine the ultimate goal for learners in terms of accent. Many studies have referred to native speaker norms that were extracted from fully or nearly monolingual populations. The current study recruited balanced bilingual speakers and found that their production of L1 phones did not always fit the monolingual “standard.” Yet it was argued that the bilingual speakers’ accent was a reasonable if not the most suitable target for L2 learners, which resonated with Lourdes Ortega’s entreaty that the field of SLA should take a “bilingual turn” (e.g., Ortega, 2009). The second question echoed Tracy Derwing and Miles Munro (e.g., 2009) in their enquiry as to whether future research should even be concerned with accent alone or rather focused on comprehensibility and intelligibility, which are arguably more important for communication. It remains an empirical question to what extent accuracy in production and perception of Spanish L2 consonantal segments impacts accent, comprehensibility and/or intelligibility and therefore how important these abilities truly are to learners.
Appendix 2.1 +PI Module for [b, d, g, β, δ, γ], Introduction

Welcome E!
If you are not E, please click here.

Introduction and Learning objectives

In this Spanish pronunciation module, we present the sounds /b, d, g/. These sounds are fairly similar to English /b, d, g/, but in Spanish they have variants that occur in certain contexts and are pronounced somewhat differently: [β, δ, γ]. Since these variants don’t exist in the same way in English, they are often difficult to pronounce for non-native speakers of Spanish. Failure to use them, though, can result in a very foreign-sounding accent, so it is important to learn how and when to use these sounds.

In this module, you will:
• Learn how the consonants /b, d, g/ are formed in Spanish
• Learn how their variants [β, δ, γ] are formed in Spanish
• Investigate in which contexts each of the sounds is used in Spanish
• Practice hearing and articulating the /b, d, g/ and [β, δ, γ] sounds in the correct contexts

If at any point in the module you want to return to previous pages to review the information, you can use the Index link on any page to navigate around the module.
Appendix 2.2 +PI Module for [b, d, g, β, δ, γ], Explanation

Explanation

The Spanish sounds /b, d, g/ are occlusive consonants. We explained in Module II that an occlusive consonant is articulated by creating an obstacle in the articulatory tract and letting air build up behind this obstacle; this air is then released suddenly.

Another group of consonants are fricative consonants. These sounds are created by forming only a partial obstruction in the articulatory tract, rather than a complete closure. In articulating fricative consonants, the air never stops completely but rather passes through this partial obstruction with friction. The English and Spanish sounds /s/ and /ʃ/, for example, are fricative sounds. You will notice that you slightly modify the airflow with your tongue or lips and teeth, but you never stop it completely, as with the occlusive consonants. (In fact, some people call occlusives consonants “stops” because the air stops, while they call fricatives consonants “continuants,” because the air continues without stopping.)

This module presents the /b, d, g/ sounds along with their fricative variants [β, δ, γ], which occur in certain required contexts, as will be explained further in the next sections.
Appendix 2.3 +PI Module for [b, d, g, β, δ, γ], Assessment of Explanation with Feedback
Appendix 2.4 +PI Module on \{b, d, g, β, δ, γ\}, Articulation

Articulation - 1

With the articulation of each of the occlusive sounds \(/b, d, g/\), there will be a complete closure or obstacle created in the mouth that keeps the air, momentarily, from escaping. In the pronunciation of \(/b/\), the obstacle is created by the top lip and bottom lip coming in contact with each other. In the pronunciation of \(/d/\), this obstacle is created through the tongue touching the back of the upper front teeth. In the pronunciation of \(/g/\), the obstacle occurs with what is known as the velum. Which is the very back of the mouth, almost your throat. Examples of these sounds can be found in words such as: voy, voy and qué. (Note that in most dialects of Spanish, there is no "v" sound. Accordingly, here we assume that the letters "β" and "iť" in Spanish represent the same sound, \(/b/\) or \(/β/\) according to the context.)

You may notice that these sounds are articulated in exactly the same places in your mouth as the consonants \(/p/; t/; k/;\) are; the only difference is that \(/b/; d/; g/\) are voiced consonants, meaning that the vocal chords vibrate when you say them, whereas \(/p/; t/; k/\) are voiceless.

The animated diagrams to the right side of the screen will help you visualize how and where each of these sounds is articulated.

Articulation - 2

In articulating the fricative sounds \(/[β, δ, γ]\) you will use the same points of articulation as with \(/b, d, g/\), but you will never completely block the airflow. Instead, you will narrow the space through which the air passes, creating friction, but never quite stopping it. Examples of these sounds can be found in words such as: haba, pido and agua.

The animated diagrams to the right side of the screen will help you visualize how and where each of these sounds is articulated.

Learning the articulation of these sounds, however, is only half of the challenge. You must also learn when to use which variant. The fricative sounds \(/[β, δ, γ]\) are actually much more common in Spanish than their occlusive counterparts and occur almost every time you will see the letters "β" (or "v"), "dβ" or "gβ". (Remember though that before the vowels "e" and "i" the letter "gβ" is pronounced somewhat like the English "h".) The occlusive sounds only occur after a pause (i.e., at the beginning of a sentence), or after the letter "n" or "m". Occlusive \(/dβ/\) also occurs after the letter "iβ". In all other cases, the fricatives are used.
Appendix 2.5 +PI Module on [b, d, g, β, δ, γ], Assessment of Articulation

**Articulation Assessment**

Before continuing, it is important that you understand what you have learned so far. Answer the questions below and then click Submit to check your answers. If you get an answer wrong you will be prompted to try again. When you have answered all the questions correctly, you will move on to the next section.

1. **The obstacle to the airflow in the pronunciation of /d/ in Spanish occurs with:**
   - [ ] The top and bottom lip
   - [x] The tongue and front teeth
   - [ ] The tongue and the roof of the mouth
   - [ ] The tongue and the velum

2. **The occlusive sounds /b, d, g/ occur:**
   - [ ] in all the same contexts as the fricatives in Spanish
   - [x] depending on the style of speech used
   - [ ] after a pause or after a nasal consonant
   - [ ] depending on the meaning of the word they are used in

3. **The fricative sounds [β, δ, γ] occur:**
   - [ ] in all the same contexts as the occlusives in Spanish
   - [ ] depending on the style of speech used
   - [ ] depending on the meaning of the word they are used in
   - [x] in all contexts other than after a pause or nasal (or /l/ in the case of [β])

[Submit] [Reset]
Appendix 2.6 +PI Module on [b, d, g, β, δ, γ], Comparison with English

Comparison with English

English and Spanish both have occlusive /b, d, g/ sounds, and they are very similar. The only real difference in the occlusive consonants is where the /d/ is articulated in each language. In English, /d/ is alveolar: it is formed with the tongue touching the roof of the mouth at the alveolar ridge. However, Spanish /d/ is dental: it is formed with the tongue closer to the front of the mouth, touching the back of the upper teeth. (This is the same distinction we saw between English /t/ and Spanish /d/.)

However, the greatest and most crucial difference is, as we mentioned earlier, that English /b, d, g/ don’t have fricative [β, ð, γ] counterparts. English does have a [ð] sound, but it is not a variant of /d/; in fact, it is its own sound and is often represented by the letters ‘th’ in English, such as in the words “the,” “they,” etc. No dialects of English have fricative [β] or [γ] as consistently occurring sounds, so these can be difficult for English-speakers to learn to pronounce correctly. You just have to remember that they have the same place of articulation as occlusive /b, g/ but the airflow never stops.

Listen to the following pairs of words, which illustrate the differences in English occlusives and Spanish fricatives. Click on a word to hear it pronounced by a native speaker.

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
<td>tabia</td>
</tr>
<tr>
<td>oode</td>
<td>oode</td>
</tr>
<tr>
<td>say</td>
<td>día</td>
</tr>
<tr>
<td>code</td>
<td>codo</td>
</tr>
<tr>
<td>agate</td>
<td>agua</td>
</tr>
<tr>
<td>ag</td>
<td>lago</td>
</tr>
</tbody>
</table>

(Remember that after a pause or an "n" (or "l") in the case of /d/), Spanish also uses occlusive sounds, which are similar to the English occlusive sounds. For example, in the phrase un día, the /d/ comes after an "n", so it is not pronounced [ðia] here, but rather [u ñ.dia].)
Appendix 2.7 +PI Module on \([b, d, g, \beta, \delta, \gamma]\), Assessment of Comparison with English

**Comparison with English Assessment**

Before continuing, it is important that you understand what you have learned so far. Answer the questions below and then click Submit to check your answers. If you get an answer wrong you will be prompted to try again. When you have answered all the questions correctly, you will move on to the next section.

1. **The greatest difference between the way in which the letters "b, d, g" are pronounced in English and Spanish is:**
   - in English, the sounds are always occlusive but in Spanish they can be fricative or occlusive
   - in Spanish, the sounds are always occlusive but in English they can be fricative or occlusive
   - in English the sounds are always fricative but in Spanish they are always occlusive
   - in Spanish the sounds are always fricative but in English they are always occlusive

2. **In addition to the above, there is also a difference in Spanish /d/ and English /d/ - this different articulation is due to:**
   - what part of the teeth the tongue touches
   - what part of the roof of the mouth the tongue touches
   - in Spanish the tongue touches the teeth but in English it touches the roof of the mouth
   - in Spanish the tongue touches the teeth but in English it touches the velum

Submit  Reset
Appendix 2.8 +PI Module on [b, d, g, β, δ, γ], Discrimination Exercise with Feedback
Appendix 2.9 +PI Module on [b, d, g, β, δ, γ], Pronunciation Practice

Now, it's your turn to practice pronouncing the sounds /b, d, g/ and [β, δ, γ] in Spanish, keeping in mind what you have learned so far. You will see a series of phrases presented in pairs - in the one on the left, the "b", "d" or "g" sounds is occlusive, but in the one on the right it is fricative. These sounds are highlighted for you so that you can be sure to pay attention to them as you read the words.

Click on the speaker icon after each phrase to hear the way I pronounce it. Then, read each of the phrases out loud, in a natural voice, paying as close attention as possible to the airflow and whether you are totally or partially obstructing it as you produce the highlighted sounds. This exercise is not recorded, just listen to yourself pronounce the phrases and think about whether you sound like the model you just heard. You can listen to and read each phrase as many times as you want until you are satisfied that you have the correct pronunciation.

<table>
<thead>
<tr>
<th>Occlusive</th>
<th>Fricative</th>
</tr>
</thead>
<tbody>
<tr>
<td>un beso</td>
<td>el beso</td>
</tr>
<tr>
<td>voy allí</td>
<td>yo voy allí</td>
</tr>
<tr>
<td>un día</td>
<td>cuatro días</td>
</tr>
<tr>
<td>dámelo</td>
<td>yo te doy</td>
</tr>
<tr>
<td>un guante</td>
<td>agua</td>
</tr>
<tr>
<td>un gato</td>
<td>el gato</td>
</tr>
</tbody>
</table>

Once you are satisfied that you are able to pronounce the different sounds in English and Spanish, go on to the next activity.
Appendix 2.10 – PI Instructions Example, for Second Year Learners

Today you are going to do watch some videos of native Spanish speakers talking about a particular topic. You will also perform dictation exercises, typing what you hear in the videos. There will be some parts of the videos that you won’t understand, but don’t stress out about it! Just relax and keep an open mind. This activity should take you no more than 30 minutes. It’s OK if you can’t finish the second dictation (parts 9-10).

Instructions:
1. Go to the page http://laits.utexas.edu/spe/, which should be open for you already.

2. Click on the Intermediate A level, and from the drop-down menu, choose Task 12: Talk about yourself

3. Click on the ‘Regina Ruiz: (Peru, Lima)’ video. IMMEDIATELY click on ‘N’ so that you can listen without reading. Listen and watch the video several times, without reading any transcript or translation. When you think you understand a good bit, start your dictation. That is, type exactly what the speaker says in Spanish. Pause the video after every few words if you need to. Go back and repeat the entire video as often as you need to produce what you think is a good dictation. (But stop after 15 minutes, regardless of how far you’ve gotten.)
Type your dictation here:

4. When you have finished your dictation, click ‘S’ to see the Spanish transcript. Compare it with your own. Where did you find difficulty? Can you find any patterns to your mistakes?
Type your thoughts here:

5. Do you think you understand the meaning? Click on ‘E’ to see the English translation of the transcript. Are there words or phrases that are new to you? Are there some words that you think would be helpful for you to remember and use in the future?
Type your thoughts here:

6. Pronunciation practice. Now you’ll listen carefully to some specific parts of the video, paying as close attention as possible to the speaker’s pronunciation. Then, read the same sections out loud, in a natural voice, trying to pronounce the words exactly as the native speaker does. You can listen to and read the sentences as many times as you want until you are satisfied that you have the correct pronunciation.

The sections you should focus on are:
The first clause (“Mi nombre es Regina Ruiz”)
The third sentence (“Me gusta mucho … Latino América”)
The last sentence (“Y el lugar … México”)

182
7. Now, click on the other videos related to this topic and listen to them. You may click on the Spanish transcript and/or English translation; do whatever you think is most helpful.

8. Now that you have watched all the other videos about this topic:
   a. do you notice anything interesting about the speakers’ pronunciations of certain sounds or words, or anything about their accents?
      **Type your thoughts here:**
   
   b. do you notice any phrases or words that are new to you and particularly interesting or useful? **Type your thoughts here:**

9. Finally, IF you have time, you are going to do one more dictation exercise. Click on Katherine Bermúdez: (Costa Rica, San José) then IMMEDIATELY click on ‘N’ so that the transcript disappears and you can listen without reading. Listen and watch the video several times, without reading any transcript. When you think you understand a good bit, start your dictation. That is, type exactly what the speaker says in Spanish. Pause the video after every few words if you need to. Go back and repeat the entire video as often as you need to produce what you think is good dictation.
   **Type your dictation here:**

10. When you have finished your dictation, click ‘S’ to see the Spanish transcript. Compare it with your own. Where did you find difficulty? Can you find any patterns to your mistakes?
    **Type your thoughts here:**

Thank you for your work today!
Appendix 2.11 Transcripts of Video Vignettes Used for First Year Learners in –PI

The clauses and sentences used for pronunciation practice are underlined.

Dictation A for First Year Learners
Topic: People’s physical characteristics
Vignette 1, “Simplified Example”:
Voy a describir a mi familia. Mi madre es de estatura mediana, rubia y de ojos claros. Mi padre, en cambio, es un tanto más alto, pero de tez morena y cabello oscuro. Mi hermano tiende a ser más gordito, pero de tez más blanca que la mía. Yo soy alto, delgado y de cabello y ojos oscuros.

Vignette 2, Speaker from San Juan, Puerto Rico:
Empiezo describiendo a mi hermano. Es de, mide aproximadamente 5 pies 10 pulgadas, tez blanca, el pelo marrón, tiene los ojos color marrón claros, pelo lacio no como el mío, que es, este... rizo. A ver qué más, un poquito gordito, parece que le va bien allá con la comida.

Dictation B for First Year Learners
Topic: What you like to do
Vignette 1, Speaker from Quito, Ecuador:
Hay cuatro cosas que me encantan... hacer. La una es leer. Me gusta mucho leer poesía o leer novelas. Y paso la mayor parte del tiempo con un libro cuando estoy haciendo la fila en el banco o la fila para comprar algo de comer; siempre tengo un libro conmigo. Pero también me gusta mucho cocinar porque uno de mis sueños es tener más adelante un restaurante de comida criolla ecuatoriana. Y siempre estoy experimentando con la comida y con nuevos ingredientes. También me gusta mucho bailar salsa. Y es algo que lo hago en los fines de semana. Me gusta salir y bailar como... y esto me ayuda a desestresarme y quitarme todo el peso que tengo de los estudios y del trabajo. Y finalmente me gusta hacer manualidades. Mi madre es artesana y por eso yo hago muchas cosas con con las manos. Puedo hacer un poco de cerámica, o de pintura, o de tejer, o de bordado. Y siempre me gusta hacer cosas para adornar la casa o la casa de las personas que quiero. Esas son las cosas que me gustan hacer.

Vignette 2, “Simplified Example”:
Me gustan todos los deportes, pero especialmente la natación. Normalmente, nado dos o tres veces por semana. Otro deporte que me gusta es el tenis. También me gusta el arte. Me gusta ir a conciertos de rock o a conciertos de música clásica, pero cualquier actividad que yo practique, me gusta hacerla con la gente que quiero.
Appendix 2.12 Transcripts of Video Vignettes Used for Second Year Learners in –PI
The clauses and sentences used for pronunciation practice are underlined.

Dictation A for Second Year Learners
Topic: Tell about where you are from
Vignette 1, Speaker from Bogotá, Colombia:
Bueno, yo nací en Bogotá en 1985. Viví casi toda mi vida allá, me mudé en el dos mil cuando tenía quince años, me mudé a los Estados Unidos, pero los recuerdos de Bogotá todavía están muy presentes. Es una ciudad muy linda, hace frío, la temperatura normal fluctúa entre los siete-ocho grados centígrados hasta los veinte grados centígrados cuando hace calor. Hace el mismo clima durante todo el verano porque estamos sobre la línea del ecuador entonces el clima es templado. Es una ciudad muy grande, la ciudad más grande de Colombia. Es muy linda, como dije, hay muchos carros, mucha polución, pero hay verdor en las calles y en las avenidas principales, hay muchos árboles. Se ve mucha pobreza, mucha pobreza como en cualquier país latinoamericano, cualquier ciudad metropolitana de Latinoamérica, hay mucha pobreza. Pero hay formas de aislarse un poco de esa pobreza y de ir a malles, de ir al cine, de ir al teatro. Hay muchos bares, muchas discotecas, la vida nocturna es muy activa, es una ciudad de estudiantes también. Mucha gente que vive en las ciudades más chiquitas de Colombia van a Bogotá a estudiar específicamente ahí. Es un país, es un país especial, es una ciudad interesante.

Vignette 2, Speaker from Lima, Peru:
La migración es un fenómeno muy actual, que hace que las personas de una misma familia sean de distinta procedencia. Los Estados Unidos es el país modelo del encuentro cultural. Hay personas de diversos orígenes y diferentes razas, así como de culturas diferentes. Esta situación internacional ha hecho que algunas personas gocen de una doble y hasta triple nacionalidad. Algunas hasta tienen varios pasaportes.

Dictation B for Second Year Learners
Topic: Talk about yourself
Vignette 1, Speaker from Lima, Peru:
Mi nombre es Regina Ruiz, soy peruana, y tengo 28 años. Soy profesora en la universidad de Texas, y también soy estudiante de la facultad de literatura. Me gusta mucho viajar y creo que me gusta mucho viajar porque viví en muchos lugares en Latinoamérica. He vivido en Brasil, cuatro años, he vivido en Ecuador, en México. He vivido en El Salvador. Y el lugar que más me ha gustado ha sido México.

Vignette 1, Speaker from San José, Costa Rica:
The clauses and sentences used for pronunciation practice are underlined.

Dictation A for Third Year Learners
Topic: Tell about how long has it been since you took a break
Vignette 1, Speaker from Popayan, Colombia:
Bueno, las últimas vacaciones en las que estuve fue en Enero de este año, del dos mil cuatro. Fui a Colombia, obviamente a pasar la navidad con mis papás y decidimos ir a la zona donde se cultivaba el café, el café de Colombia. Entonces yo, bueno, yo fui la elegida para manejar todo el tiempo el carro, lo cual no me molestó porque me gusta mucho manejar. Y fuimos a esta región que se llama El Quindio y conocimos como la gente, las fincas—que son unas casas muy lindas con todos los plantíos de café, sitios muy lindos—y cómo era el proceso por el cual se producía el café que me pareció muy interesante porque no lo sabía... hasta el final, hasta cuando ya queda empacado para ser exportado. Pues era lleno de naturaleza, la gente muy querida, la comida deliciosa, estuvimos en un parque de un museo con la última tecnología que había—que existe creo yo en producción de café y bueno toda la historia de, no sé, desde principio de siglo veinte, cómo fue evolucionando la producción de café y cómo fue mejorando los procesos de producción y recolección de café. Entonces fue una experiencia muy bonita pues porque yo no sabía cómo se hacían las cosas, ni mis papás. Y pasé con mis papás, entonces también fue muy rico. Y bueno, fue muy muy divertido, muy tranquilo, eso sí.

Vignette 2, Speaker from Madrid, Spain:
Muy bien, hace unos cuatro meses fuimos a México a visitar la familia de mi esposo. Pasamos una semana en la ciudad de México y otra semana en la playa, en Acapulco. Nunca había conocido Acapulco así que fue bastante divertido.

Dictation B for Third Year Learners
Topic: Describe an invention that will make you rich some day
Vignette 1, Speaker from Guatemala, Guatemala:
Me gustaría inventar platos desechables que se pudieran comer, de esta manera no habría que descartarlos y así reduciría un buen porcentaje de la basura que se produce en las escuelas públicas. Aunque realmente no es un invento que creo que generaría demasiado dinero, no es algo que me interesa mucho pero creo que sería positivo en cuanto al uso de los recursos de una manera más apropiada.

Vignette 2, Speaker from Lima, Peru:
Bueno si tuviera que inventar algo para hacerme millonario, yo inventaría, bueno es bien difícil pero inventaría la cura contra el cáncer. ¿Cómo lo haría? Bueno, tendría que empezar a hacer un poco de research y bueno trabajar con el apoyo de una compañía porque necesito muchos recursos en ese sentido y luego pues tomaría un montón de tiempo hacerlo y una vez que se pueda hacer y sea factible estoy seguro que con esa medicina o vacuna se podría hacer muchísimo dinero.
Appendix 2.14 Production Test Items

1. ¿Cómo? ‘What?’
2. ¿Qué tal? ‘How are you?’
3. señorita ‘Miss’
4. avenida ‘avenue’
5. Hasta luego. ‘See you later.’
6. perro ‘dog’
7. agosto ‘August’
8. cubano ‘Cuban’
9. rubio ‘blonde’
10. para ella ‘for her’
11. tú y yo ‘you and I’
12. pintar ‘to paint’
13. aburrido ‘bored’
14. ¿Dónde vives? ‘Where do you live?’
15. llegar ‘to arrive’
16. computadora ‘computer’
17. barrio ‘neighborhood’
18. Adiós. ‘Good bye.’
19. número ‘number’
20. tocar ‘to touch’
21. talle ‘figure’
22. cace ‘hunt’ [3 sing., pres. subjunctive]
23. pace ‘graze’ [3 sing., pres. indicative]
24. amaga ‘threaten’ [3 sing., pres. indicative]
25. calada ‘soaked’ [adj., fem. sing.]
26. ara ‘altar’
27. arras ‘dowry’
28. calaba ‘was soaking’ [3 sing., past imperfect]
Appendix 2.15 Production Test Instructions

Instructions: You are going to read some words aloud in Spanish. Your voice will be recorded using SANAKO. SANAKO (“DUO”) should be open already, but if it’s not, open it from the desktop or list of programs. Try a little test run with your microphone first. When you are ready to record, press the RED (record) button. When you’ve finished recording, press the BLUE (stop) button. Press the GREEN (play) button to hear what you recorded. Notice that you don’t need to speak very loudly. Click “File” – “New” – and “Don’t Save”, to clear the memory of SANAKO for your speaking test. Now you’re ready to begin the test.

*Please read aloud the following Spanish words and phrases. You probably won’t know all of the words. Just try your best to sound them out. Try not to read very fast or very slow, but rather at a normal speech rate. Do not repeat words.*

When you’ve finished recording, press the BLUE (stop) button. Press the GREEN (play) button and listen to the first few words you pronounced. If you think the sound quality is good, and your voice is not too quite to hear well, then you have finished. Please get the attention of a researcher, so that we can make sure your file is saved properly. We will save your file in your folder (W://Student/0Teachers/Kissling/StudentFolders/your code number/session 1) as “YourCodeNumberSpeakingPRE.WAV” (*NOT* as an .mp3!).

**There is one last part to this test.** Please go back and type the English translation of the words and phrases. If you don’t know the meaning of a word, just type “?” . You are not expected to know them all!
Appendix 2.16 Background Questionnaire for Spanish Learners

Please complete the following questionnaire. The information you provide will be kept confidential and used only for the purposes of the study. Please underline your answers to multiple-choice questions.

Age: __ Gender: M / F
Your class level (Freshman / Sophomore / Junior / Senior / Other)

If you answered “Other,” please explain. __

Major 1 _____
Major 2 (if any) _____
Minor 1 (if any) _____
Minor 2 (if any) _____

Language Experience
1. Are you of Hispanic and/or Spanish-speaking background? (Yes / No)

If you answered “Yes,” please explain. ______

2. Approximately how old were you when you first began studying Spanish? _____

3. What Spanish courses did you take before college? For each, please describe how many hours per week you were in class, how old you were, and (to your knowledge) if your instructor a native speaker of Spanish or not?

Course name ___ Class hrs/wk __ Age __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Age __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Age __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Age __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Age __ Native Speaker? (Yes/No/Not Sure)

4. What college level courses have you taken? For each, please describe how many hours per week you were in class, and (to your knowledge) if your instructor was a native speaker of Spanish or not.

Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
Course name ___ Class hrs/wk __ Native Speaker? (Yes/No/Not Sure)
5. Have you had any instruction in Spanish sounds or Spanish pronunciation – for instance, with a language teacher, private tutor, or computer program? If so, please describe the instruction.

6. Do you use Spanish outside the classroom? For instance, do you watch Spanish TV or movies or speak Spanish with family, friends, or coworkers? Please briefly describe what you do, in terms of when you use Spanish, how often, with whom, for what purposes, etc.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Frequency in Hours per Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch Spanish TV/movies</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Listen to Spanish music</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Speak Spanish with family</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Speak Spanish with friends</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Speak Spanish with coworkers</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Other (explain): __</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Other (explain): __</td>
<td>0  1  2  3  4  5  6  7  8  9  10</td>
</tr>
</tbody>
</table>

7. Have you traveled to or lived in a Spanish-speaking country? If so, which country? Why? For how long? How did you use Spanish while you were there?

<table>
<thead>
<tr>
<th>Country</th>
<th>Length of stay</th>
<th>Purpose of travel</th>
<th>Use of Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Length of stay</td>
<td>Purpose of travel</td>
<td>Use of Spanish</td>
</tr>
</tbody>
</table>

8. Have you studied any languages other than Spanish? If so, please describe for each which classes you have taken, how old you were, and how many hours per week you were in class.

<table>
<thead>
<tr>
<th>Language</th>
<th>Course name ___ Class hrs/wk ___ Age ___</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Course name ___ Class hrs/wk ___ Age ___</td>
</tr>
<tr>
<td>Language</td>
<td>Course name ___ Class hrs/wk ___ Age ___</td>
</tr>
</tbody>
</table>
Course name ____ Class hrs/wk ___ Age __
Course name ____ Class hrs/wk ___ Age __
Course name ____ Class hrs/wk ___ Age __

9. Do you use a language other than Spanish or English outside the classroom? For instance, do you watch TV or movies in this language or speak the language with coworkers or friends? Please briefly describe what you do, in terms of when you use the language, how often, with whom, for what purposes, etc.
Language ____
Language ____
Language ____

10. Do you have any other language experience that you haven’t mentioned so far? If so, please explain.

_______

Thank you for completing the questionnaire. The information you provided will be kept confidential and used only for the purposes of the study.
Appendix 2.17 Background Questionnaire for Native Spanish Speakers

Participant Code Number (leave blank; a code number will be assigned to you): ______

Please complete the following questionnaire. The information you provide will be kept confidential and used only for the purposes of the study. Please underline your answers to multiple-choice questions.

Age: ___ Gender: M / F Current occupation: ___

Language Experience

1. Where were you born? Country: ___ City: ___
2. Where have you lived other than your place of birth? Please explain where, when, and for how long. ___ ___ ___
3. Did you hear and use only Spanish between the ages of 0 – 5? Yes / No
   If you answered ‘No,’ please explain. ____

4. What other languages do you know? Please explain how old you were when you started learning or using each language, how you used it (e.g., in school, with friends, etc.), the total amount of time or experience you have with it, and about how proficient you currently are in it.
   a. Language: ____
      Age: ____
      How I learned/used it: ___
      Years experience: ___
      Estimated proficiency level: novice / low intermediate / high intermediate / advanced / very advanced
   b. Language: ____
      Age: ____
      How I learned/used it: ___
      Years experience: ___
Estimated proficiency level: novice / low intermediate / high intermediate / advanced / very advanced
c. Language: ____
   Age: _____
   How I learned/used it: ___
   Years experience: ___
Estimated proficiency level: novice / low intermediate / high intermediate / advanced / very advanced
d. Language: ____
   Age: _____
   How I learned/used it: ___
   Years experience: ___
Estimated proficiency level: novice / low intermediate / high intermediate / advanced / very advanced

5. Do you have any other language experience that you haven’t mentioned so far? If so, please explain.
   ____

6. Have you ever taken a class in, or otherwise received training in, phonetics or phonology? Yes / No
   If yes, please explain. ____

Thank you for completing the questionnaire. The information you provided will be kept confidential and used only for the purposes of the study.
Appendix 2.18 Post-instructional Module Questionnaire

Please briefly assess the lesson you just finished. Your responses will be kept confidential and will not affect your grade in your class or any incentive you might get for participating.

1. How long did you spend on the lesson? (Be as exact as possible.) ____ minutes

2. How difficult was the lesson? (Underline a number. 1=Very Easy, 10=Very Difficult)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can:

3. How interesting was the lesson? (Underline a number. 1=Very Uninteresting, 10=Very Interesting)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can:

4. How helpful was the lesson? (Underline a number. 1=Very Unhelpful, 10=Very Helpful)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can:

5. How likely are you to use what you learned from the lesson inside or outside of class? (Underline a number. 1=Very Unlikely, 10=Very Likely)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can:

6. How likely would you be to recommend this lesson to others if they asked you about it? (Underline a number. 1=Very Unlikely, 10=Very Likely)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can:

7. Did the lesson include information that you have heard or been taught before? If so, what exactly?:

8. How could the lesson be improved?:

9. Is there anything else you’d like to add?:

194
Appendix 2.19 Post-study Questionnaire

Please complete the following questionnaire. Be honest. Your responses will not affect your grade in your class or any incentive you might get for participating in the study. Your answers will remain confidential.
Questions 1 – 10 refer to the four lessons you completed (modules or dictations).

1. Please rate the lessons over all, with ‘0’ being the worst and ‘10’ being the best. (Underline a number)
   1  2  3  4  5  6  7  8  9  10

2. How difficult were the lessons? (Underline a number. 1=Very Easy, 10=Very Difficult.)
   1  2  3  4  5  6  7  8  9  10
   Please explain if you can/have time:

3. How interesting were the lessons? (Underline a number. 1=Very Uninteresting, 10=Very Interesting)
   1  2  3  4  5  6  7  8  9  10
   Please explain if you can/have time:

4. How helpful were the lessons? (Underline a number. 1=Very Unhelpful, 10=Very Helpful)
   1  2  3  4  5  6  7  8  9  10
   Please explain if you can/have time:

5. How likely are you to use what you learned from the lessons inside or outside of class?
   (Underline a number. 1=Very Unlikely, 10=Very Likely)
   1  2  3  4  5  6  7  8  9  10
   Please explain if you can/have time:

6. How likely would you be to recommend these lessons to others if they asked you about it?
   (Underline a number. 1=Very Unlikely, 10=Very Likely)
   1  2  3  4  5  6  7  8  9  10
   Please explain if you can/have time:

7. Did the lessons include information that you have heard or been taught before? If so, what exactly?:

195
8. How could the lessons be improved?

9. Is there anything else you’d like to add, about the lessons?

10. Over the past few weeks, have you received additional instruction, outside the lessons themselves, about the Spanish sounds that were presented in the lessons? If so, please explain, in as much detail as possible.: 

Questions 11 – 17 refer to the tests you have taken after completing the lessons.

Questions 11 - 12 refer specifically to the speaking tests.

11. Please rate the difficulty of the speaking tests. (Underline a number. 1=Very Easy, 10=Very Difficult.)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can/have time:

12. Please estimate your score on the speaking tests, with ‘1’ being the worst and ‘10’ being the best.

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can/have time:

Questions 13 - 14 refer specifically to the same/different listening tests, in which you had to decide whether two sound recordings were the same or different.

13. Please rate the difficulty of the same/different listening tests. (Underline a number. 1=Very Easy, 10=Very Difficult.)

   1  2  3  4  5  6  7  8  9  10

   Please explain if you can/have time:
14. Please estimate your score on the **same/different listening tests**, with ‘1’ being the worst and ‘10’ being the best.

```
1 2 3 4 5 6 7 8 9 10
```

Please explain if you can/have time:

Questions 15 - 16 refer specifically to the **identification listening tests**, in which you had to type the nonsense Spanish words that you heard.

15. Please rate the difficulty of the **identification listening tests**. (Underline a number. 1=Very Easy, 10=Very Difficult.)

```
1 2 3 4 5 6 7 8 9 10
```

Please explain if you can/have time:

16. Please estimate your score on the **identification listening tests**, with ‘1’ being the worst and ‘10’ being the best.

```
1 2 3 4 5 6 7 8 9 10
```

Please explain if you can/have time:

17. Is there anything else you’d like to add about any of the **speaking** or **listening tests**?:

*Thank you very much for your participation!*
Appendix 3.1 Discrimination Test Pairings

**Target Phones in AX Pairings (items included in analysis)**

<table>
<thead>
<tr>
<th>PA</th>
<th>PH</th>
<th>PI</th>
<th>PO</th>
<th>PU</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>PH</td>
<td>PI</td>
<td>PO</td>
<td>PU</td>
</tr>
<tr>
<td>TA</td>
<td>TE</td>
<td>TI</td>
<td>TO</td>
<td>TU</td>
</tr>
<tr>
<td>KA</td>
<td>KE</td>
<td>KI</td>
<td>KO</td>
<td>KU</td>
</tr>
<tr>
<td>AB</td>
<td>ABE</td>
<td>UBO</td>
<td>IBE</td>
<td>OBI</td>
</tr>
<tr>
<td>ADA</td>
<td>ADE</td>
<td>UDO</td>
<td>IDE</td>
<td>ODI</td>
</tr>
<tr>
<td>AGA</td>
<td>AYE</td>
<td>UGO</td>
<td>IGE</td>
<td>OGI</td>
</tr>
<tr>
<td>ORO</td>
<td>ORO</td>
<td>IR</td>
<td>IR</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>ARRI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Target Phones in XX Pairings**

<table>
<thead>
<tr>
<th>PHI</th>
<th>PHE</th>
<th>PI</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>TE</td>
<td>TO</td>
<td>TU</td>
</tr>
<tr>
<td>KA</td>
<td>KI</td>
<td>KO</td>
<td></td>
</tr>
<tr>
<td>UBO</td>
<td>IBE</td>
<td>OBI</td>
<td></td>
</tr>
<tr>
<td>ADA</td>
<td>IDE</td>
<td>EDE</td>
<td></td>
</tr>
<tr>
<td>AGA</td>
<td>UGO</td>
<td>OGI</td>
<td></td>
</tr>
<tr>
<td>ORO</td>
<td>IRA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>ARRI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Non-Target Phones in AX Pairings**

<table>
<thead>
<tr>
<th>PA</th>
<th>PE</th>
<th>FE</th>
<th>FEI</th>
<th>3E</th>
<th>3EI</th>
<th>LO</th>
<th>LOO</th>
<th>ZO</th>
<th>ZOO</th>
</tr>
</thead>
</table>

**Non-Target Phones in XX Pairings**

| BA | BE | AE | AAP | FEI | FEI | 3E | 3E |
Appendix 3.2 Acoustic Analysis of Discrimination Test Stimuli /p, t, k/

<table>
<thead>
<tr>
<th>AX Item</th>
<th>Spanish-like Token (Unaspirated)</th>
<th>English-like Token (Aspirated)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pa] / [pʰa]</td>
<td>9</td>
<td>51</td>
<td>42</td>
</tr>
<tr>
<td>[pe] / [pʰe]</td>
<td>8</td>
<td>45</td>
<td>37</td>
</tr>
<tr>
<td>[pi] / [pʰi]</td>
<td>10</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>[po] / [pʰo]</td>
<td>9</td>
<td>69</td>
<td>60</td>
</tr>
<tr>
<td>[pu] / [pʰu]</td>
<td>14</td>
<td>66</td>
<td>52</td>
</tr>
<tr>
<td>average</td>
<td>10</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>[ta] / [tʰa]</td>
<td>13</td>
<td>103</td>
<td>90</td>
</tr>
<tr>
<td>[te] / [tʰe]</td>
<td>12</td>
<td>43</td>
<td>31</td>
</tr>
<tr>
<td>[ti] / [tʰi]</td>
<td>20</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td>[to] / [tʰo]</td>
<td>19</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>[tu] / [tʰu]</td>
<td>19</td>
<td>37</td>
<td>18</td>
</tr>
<tr>
<td>average</td>
<td>17</td>
<td>62</td>
<td>46</td>
</tr>
<tr>
<td>[ka] / [kʰa]</td>
<td>25</td>
<td>98</td>
<td>73</td>
</tr>
<tr>
<td>[ke] / [kʰe]</td>
<td>29</td>
<td>77</td>
<td>48</td>
</tr>
<tr>
<td>[ki] / [kʰi]</td>
<td>47</td>
<td>99</td>
<td>52</td>
</tr>
<tr>
<td>[ko] / [kʰo]</td>
<td>28</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>[ku] / [kʰu]</td>
<td>49</td>
<td>104</td>
<td>55</td>
</tr>
<tr>
<td>average</td>
<td>36</td>
<td>88</td>
<td>53</td>
</tr>
</tbody>
</table>
Appendix 3.3 Acoustic Analysis of Vowels in Discrimination Test Stimuli

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]</td>
<td>291</td>
<td>2299</td>
</tr>
<tr>
<td>[i]</td>
<td>290</td>
<td>2379</td>
</tr>
<tr>
<td>[a]</td>
<td>964</td>
<td>1276</td>
</tr>
<tr>
<td>[a]</td>
<td>896</td>
<td>1277</td>
</tr>
<tr>
<td>[o]</td>
<td>641</td>
<td>2022</td>
</tr>
<tr>
<td>[o]</td>
<td>627</td>
<td>1964</td>
</tr>
</tbody>
</table>
Appendix 3.4 Identification Test Items

- ba.’ra. ða barada
- bi.si.’ye bisigué
- bo.’dor bodor
- bo.’ni.βe bonibe
- bu.’ro.βi burrovi
- di.’ye digné
- di.’rin.se dirrinse
- do.’yn.so doguínzo
- fa.’ran farrán
- fu.’Bo.ti fuvoti
- fu’.óo.ta fudota
- ge.’ðe.sa guedesa
- ka.’po.tro capotro
- ka.’pu.se capuse
- ke.’βe.ru queberu
- ke.’yem.bo queguembo
- ke.yen.’po queguempó
- ‘kin.dor quindor
- ‘kin.sa quinsa
- ko.’yan.por cogánpor
- ko.’ner coner
- ku.’βor cubor
- ‘ku.yo cugo
- ‘ku.mi cumi
- le.’ðe.mo ledemo
- lo.’dor lodor
- lo.sa.’ða losadá
- o.’βi.ri obiri
- o.βi.’to obitó
- o.’ði.ta odita
- o.’siri osiri
- ‘pa.βa paba
- ‘pa.βi pabi
- pa.’ya.ra pagarra
- pe.’ra.le perrale
- pe.’su.ðo pesudo
- ‘pi.fa piña
- pi.ya piga
- po.’ni.le ponile
- po.’su.ði posudi
- po.’su.ðo posudo
- ‘pu.yer púguer
- pu.’ko.yi pucogui
- pu.ror purror
- pu.sor pusor
- ‘re.βen rèben
- re.’βe regué
- re.’le.su relesu
- ‘re.lu.no réluno
- re.’sa.γe resague
- re.’si.γe resigue
- ri.’γe rigué
- ‘sa.βa.ra sábara
- sa.’sa.βa sasava
- ‘si.βe sibe
- so.ði.’par sodipar
- ‘ta.ya taga
- ‘ta.pi tapi
- ta.’ran tarrán
- ta.’ron tarrón
- ‘te.ti teti
- ‘ti.ðe tide
- ti.’ka.ða ticada
- ‘to.fer tófer
- ‘to.ni.pa tónipa
- ‘tr.γe tregue
- tu.’se.ki tusequi
- tu.’su.γo tusugo
- te.’ka.ða tecada
- ‘we.βer huéver
- we.’mi.ðe huemide
- wi.’rin huirrin
- xe.’βer jever
References


perceptual assimilation model. In Goodman, J. & Nusbaum, H.C. (Eds.), The
Development of Speech Perception: The Transition from Speech Sounds to Spoken Words

Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in

reorganization for normative speech contrasts: Zulu click discrimination by English-
speaking adults and infants. Journal of Experimental Psychology: Human Perception and
Performance, 14, 345-360.

Best, C.T. & Tyler, M. (2007). Nonnative and second-language speech perception:
Commonalities and complementarities. In Bohn, O.S. & Munro, M.J., (Eds.), Language
Experience in Second Language Speech Learning: In honor of James Emil Flege (pp. 13-
34). Amsterdam: John Benjamins.

in the pronunciation of a foreign language. Studies in Second Language Acquisition, 19,
447-465.

Contrast on Speech Perception. In Avery, P., Dresher, E. & Rice, K. (Eds.) Contrast in
Phonology (pp. 143-172). Berlin: Mouton de Gruyter.

identify English /t/ and /l/ IV: Some effects of perceptual learning on speech production.


Correa, M. (2011). La enseñanza de fonética y fonología a través de analogías y
metáforas. *Hispania, 94*(2), 360-365.


Nibert, H. (2006). The Acquisition of the Phrase Accent by Beginning Adult Learners of


Polka, L. (1992). Characterizing the influence of native language experience on adult speech


Shafiro, V. & Kharkhurin, A.V. (2009). The Role of Native-Language Phonology in the


