COMPARING THE SUCCESS OF PRONUNCIATION INSTRUCTION IN
ELEMENTARY AND INTERMEDIATE GERMAN LANGUAGE CLASSROOMS

A Dissertation in

German

by

Ashley Roccamo

© 2014 Ashley Roccamo

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2014
The dissertation of Ashley Roccamo was reviewed and approved* by the following:

B. Richard Page  
Associate Professor of German and Linguistics  
Dissertation Adviser  
Chair of Committee  
Head of the Department of Germanic and Slavic Languages & Literatures

Carrie Jackson  
Associate Professor of German and Linguistics

Michael T. Putnam  
Assistant Professor of German and Linguistics

Paola E. Dussias  
Professor of Spanish, Linguistics and Psychology  
Head of the Department of Spanish, Italian and Portuguese

Mary Grantham O’Brien  
Special Member  
Associate Professor of German  
University of Calgary

*Signatures are on file in the Graduate School.
Abstract

The phonology of a second language (L2) is notoriously difficult to acquire. Oftentimes even advanced L2 speakers cannot acquire accurate pronunciation on their own (Grosser, 1997; Jilka, 1999; Munro & Derwing, 2008; Trofimovich & Baker, 2006), and their abilities tend to stabilize after a certain amount of time. To counteract this, training in L2 pronunciation has been found effective in a number of classroom and laboratory studies (Derwing, Munro & Wiebe, 1998; Elliott, 1997; Flege, 1989; Hardison, 2004; Saito & Lyster, 2011). Many researchers recommend introducing pronunciation training into the curriculum as early in the language learning sequence as possible (e.g., Counselman, 2010; Elliott, 1995, 1997; Eskenazi, 1999; Hardison, 2004; Neufeld & Schneiderman, 1980), yet previous experimental studies have not tested the validity of these recommendations. Most training programs are introduced in more advanced stages of L2 proficiency (e.g., Counselman, 2010; Elliott, 1995, 1997; Lord, 2008), and pronunciation instruction is still largely ignored in elementary-level classrooms. Thus the question of when to begin formal pronunciation instruction has yet to be sufficiently answered.

This dissertation addresses the above issue by directly comparing the effectiveness of pronunciation instruction at elementary and intermediate proficiency levels in order to detect the best time course for introducing pronunciation instruction in the classroom. A two-group experimental design was used, and forty-four students in four intact sections of first- and fourth-semester German courses participated in this study. An identical eight-week pronunciation training unit was implemented as a supplement in the experimental sections of the first- and fourth-semester German language classes. Training was divided into eight modules that each focused on one segmental or prosodic aspect of the German language at a time, including lexical stress, palatal and velar fricatives ([ç] and [x]), fricative and vocalized /r/, and the monophthongization of [e] and [o]. Training in both the perception and production of the
targeted features was provided, and training was conducted in the first ten minutes of each class period. At the conclusion of each module, participants in all groups completed partner recordings in which they practiced producing the target features in citation form and in free speech contexts. Following completion of these recordings, students provided peer feedback on pronunciation for their partner. Data come from identical pre- and posttests that were administered for both groups before and immediately after training. Pre- and posttest speech samples were rated by native German speakers for ease of comprehensibility and strength of accent. Pre- and posttest ratings from the experimental and control groups, as well as from first- and fourth-semester students, are compared in order to gauge the effectiveness of pronunciation training in the earliest semesters of German L2 learning.

Results show that first-semester students who received training were able to significantly improve their phonological perception skills, their comprehensibility, and their accent to a greater extent than both those students who did not receive training and fourth-semester students. In comparison to those same groups, first-semester students who received training were able to significantly improve their pronunciation in more of the problem areas for which they had received training. The dissertation concludes that pronunciation instruction is effective for students in their first semester of learning German, and that it may be more effective to train L2 learners in the first semesters than once they have reached more advanced stages of L2 proficiency. I argue that instruction in pronunciation should begin in the first few semesters of L2 German learning for optimal results. As such, this work provides an answer to a gap in the literature and begins to provide answers to the appropriate time course for intervention to begin for the most effective results.
## Table of Contents

List of Figures .......................................................................................................................... viii

List of Tables ............................................................................................................................. x

Acknowledgements ..................................................................................................................... xi

Chapter 1: Introduction and Background Literature .................................................................... 1

1.1 Introduction ............................................................................................................................. 1

1.2 Why Pronunciation? ............................................................................................................... 2

1.3 Previous Pronunciation Training Studies .............................................................................. 5

1.4 Attention and Motivation in Second Language Acquisition .................................................. 17

1.4.1 Noticing Hypothesis ......................................................................................................... 18

1.4.2 Further Development of the Concept of Attention in SLA .................................................. 20

1.4.3 Attention and Noticing in the Classroom .......................................................................... 22

1.4.4 Motivation in Second Language Learning ......................................................................... 26

1.5 Conclusion ............................................................................................................................. 30

Chapter 2: Focus of the Current Study ....................................................................................... 32

2.1 Research Questions ................................................................................................................. 32

2.2 Hypotheses ............................................................................................................................. 34

2.3 Pronunciation Targets Chosen for Training .......................................................................... 35

2.4 Differences between English and German Phonology .......................................................... 36

2.4.1 Lexical Stress ...................................................................................................................... 37

2.4.2 Ich and Ach -- [ç] and [x] .................................................................................................. 38

2.4.3 Fricative and Vocalic /t/ .................................................................................................... 40

2.4.4 The Tense Mid Vowels [e:] and [o:] ............................................................................... 41

Chapter 3: Methodology and Analysis ....................................................................................... 44

3.1 Participants ............................................................................................................................. 44

3.1.1 First-Semester German ...................................................................................................... 45

3.1.2 Fourth-Semester German .................................................................................................. 46

3.1.3 Instructors ......................................................................................................................... 48

3.2 Stimuli ................................................................................................................................... 49

3.2.1 Perception Test .................................................................................................................. 52

3.2.2 Production Test – Word Reading ...................................................................................... 54
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.3 Production Test - Paragraphs</td>
</tr>
<tr>
<td>3.2.4 Production Test – Free Speech</td>
</tr>
<tr>
<td>3.2.5 Posttest</td>
</tr>
<tr>
<td>3.2.6 Pronunciation Attitude Inventory</td>
</tr>
<tr>
<td>3.2.7 Language Background Questionnaire</td>
</tr>
<tr>
<td>3.3 Experimental Procedure</td>
</tr>
<tr>
<td>3.3.1 Perception test</td>
</tr>
<tr>
<td>3.3.2 Production Test</td>
</tr>
<tr>
<td>3.3.3 Posttest</td>
</tr>
<tr>
<td>3.4 Pronunciation Training</td>
</tr>
<tr>
<td>3.4.1 In-Class Training</td>
</tr>
<tr>
<td>3.4.2 Partner Assignments</td>
</tr>
<tr>
<td>3.5 Analysis of Learner Data</td>
</tr>
<tr>
<td>3.5.1 Native Speaker Raters</td>
</tr>
<tr>
<td>3.5.2 Rating Materials</td>
</tr>
<tr>
<td>3.5.3 Rating Procedure</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
</tr>
<tr>
<td>4.1 Perception Test</td>
</tr>
<tr>
<td>4.2 Production Test</td>
</tr>
<tr>
<td>4.2.1 Comprehensibility Data</td>
</tr>
<tr>
<td>4.2.2 Accentedness Data</td>
</tr>
<tr>
<td>4.3 Generalization</td>
</tr>
<tr>
<td>4.3.1 Generalization of Comprehensibility</td>
</tr>
<tr>
<td>4.3.2 Generalization of Accentedness</td>
</tr>
<tr>
<td>4.4 Pronunciation Attitude Inventory</td>
</tr>
<tr>
<td>4.5 Correlations &amp; Relationships</td>
</tr>
<tr>
<td>4.5.1 Perception Test Correlations</td>
</tr>
<tr>
<td>4.5.2 PAI Correlations</td>
</tr>
<tr>
<td>4.5.3 Accentedness and Comprehensibility Correlations</td>
</tr>
<tr>
<td>4.6 Summary of Results</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
</tr>
<tr>
<td>5.1 Targeted Conditions</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Average Overall Perception Test Scores by Group .......................................................... 78
Figure 2: Change in Perception Scores for First-Semester Experimental Group ........................................ 79
Figure 3: Change in Perception Scores for First-Semester Control Group .................................................. 79
Figure 4: Change in Perception Scores for Fourth-Semester Experimental Group ....................................... 80
Figure 5: Change in Perception Scores for Fourth-Semester Control Group .................................................. 80
Figure 6: Percentage of Improved, Declined and Same Scores in Trained vs. Untrained Groups .......... 81
Figure 7: Amount of Change in Average Comprehensibility Ratings for First-Semester Participants ...... 88
Figure 8: Average Comprehensibility Ratings, First-Semester Experimental .............................................. 89
Figure 9: Average Comprehensibility Ratings, First-Semester Control ......................................................... 90
Figure 10: Amount of Change in Average Comprehensibility Ratings for Fourth-Semester Participants ... 92
Figure 11: Average Comprehensibility Ratings, Fourth-Semester Experimental ........................................ 94
Figure 12: Average Comprehensibility Ratings, Fourth-Semester Control ..................................................... 95
Figure 13: Mean Comprehensibility Ratings per Group ................................................................................. 96
Figure 14: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, First-Semester Experimental Group ............................................................................................................. 99
Figure 15: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, First-Semester Control Group .................................................................................................................... 100
Figure 16: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Experimental Group .............................................................................................................. 101
Figure 17: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Control Group .................................................................................................................. 101
Figure 18: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Experimental Group ....................................................................................................... 102
Figure 19: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Control Group ................................................................................................................ 103
Figure 20: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Experimental Group ....................................................................................................... 104
Figure 21: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Control Group ............................................................................................................ 104
Figure 22: Amount of Change in Average Accentedness Ratings for First-Semester Participants ............ 110
Figure 23: Average Accentedness Ratings, First-Semester Experimental ..................................................... 112
Figure 24: Average Accentedness Ratings, First-Semester Control ................................................................. 113
Figure 25: Amount of Change in Average Accentedness Ratings for Fourth-Semester Participants ....... 115
Figure 26: Average Accentedness Ratings, Fourth-Semester Experimental .................................................. 117
Figure 27: Average Accentedness Ratings, Fourth-Semester Control ............................................................ 118
Figure 28: Mean Accentedness Ratings per Group ........................................................................................... 119
Figure 29: Accentedness Ratings of Individual Conditions on the Word-Reading Task, First-Semester Experimental Group ............................................................................................................... 122
Figure 30: Accentedness Ratings of Individual Conditions on the Word-Reading Task, First-Semester Control Group

Figure 31: Accentedness Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Experimental Group

Figure 32: Accentedness Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Control Group

Figure 33: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Experimental Group

Figure 34: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Control Group

Figure 35: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Experimental Group

Figure 36: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Control Group

Figure 37: Generalization of Pronunciation Skills in Comprehensibility

Figure 38: Generalization of Pronunciation Skills in Accentedness

Figure 39: Correlation between Perception Accuracy and Accentedness Ratings at Pretest

Figure 40: Correlation between Perception Accuracy and Accentedness Ratings at Posttest

Figure 41: Correlation between Perception Accuracy and Comprehensibility Ratings at Posttest

Figure 42: Correlation between Perception Accuracy and Comprehensibility Ratings at Posttest

Figure 43: Correlation between Perception Accuracy at Pretest and Total Score on the PAI

Figure 44: Correlation between Perception Accuracy at Posttest and Total Score on the PAI

Figure 45: Correlation between Accentedness and Comprehensibility Ratings at Pretest

Figure 46: Correlation between Accentedness and Comprehensibility Ratings at Posttest

Figure 47: Correlation between Amount of Change in Accentedness and Comprehensibility
List of Tables

Table 1: Descriptive Statistics for each Participant Group .................................................................47
Table 2: Self-Rated Proficiency Levels in German and English ...........................................................48
Table 3: Descriptive Statistics about Pronunciation Instructors .........................................................49
Table 4: Example Stimuli for each Training Condition ........................................................................50
Table 5: Wilcoxon Signed Rank Test Results- Perception Test ............................................................83
Table 6: Results of Wilcoxon Signed Ranks Tasks for Comprehensibility Ratings per Discourse Level ....99
Table 7: Significant Improvements in Comprehensibility per Condition per Group ..............................105
Table 8: Significant Improvements in Comprehensibility per Condition for First-Semester Experimental Group ................................................................................................................106
Table 9: Significant Improvements in Comprehensibility per Condition for First-Semester Control Group ..........................................................................................................................107
Table 10: Results of Wilcoxon Signed Ranks Tasks for Accentedness Ratings per Discourse Level ....122
Table 11: Significant Improvements in Accentedness per Condition per Group ....................................128
Table 12: Significant Improvements in Accentedness per Condition for First-Semester Experimental Group .....................................................................................................................................129
Table 13: Significant Improvements in Accentedness per Condition for First-Semester Control Group ...130
Table 14: Summary of Study Results and Improvements .....................................................................132
Table 15: Percentage of Fourth-Semester Participants who Changed their Perception and Pronunciation Accuracy as a Factor of Previous Language Experience .................................................................................................................134
Table 16: Percentage of Fourth-Semester Participants who Changed their Perception and Pronunciation Accuracy as a Factor of Motivation Level ........................................................................................................137
Table 17: Statistical Analyses of Generalization of Comprehensibility per Task per Group ..................142
Table 18: Statistical Analyses of Generalization of Accentedness per Task per Group ..........................145
Table 19: Mean Scores on the Pronunciation Attitude Inventory split by Group ....................................146
Table 20: Correlations between Perception Test Accuracy and Accentedness Ratings .......................149
Table 21: Correlations between Perception Test Accuracy and Comprehensibility Ratings ................151
Table 22: Correlations between Perception Test Accuracy at Pretest and Amount of Change in Accentedness and Comprehensibility Ratings ........................................................................153
Table 23: Correlations between Amount of Change in Perception Test Accuracy and Amount of Change in Accentedness and Comprehensibility Ratings ........................................................................................................153
Table 24: Correlations between Perception Test Accuracy and Motivation Level ..............................155
Table 25: Correlations between Motivation Level and Accentedness and Comprehensibility Ratings at Pre- and Posttest ........................................................................................................157
Table 26: Correlations between Motivation Level and Amount of Change in Perception Accuracy, Accentedness and Comprehensibility Ratings ........................................................................158
Table 27: Correlations between Accentedness and Comprehensibility Ratings at Pre- and Posttest ......160
Table 28: Correlations between Amount of Change in Accentedness and Comprehensibility Ratings ....162
Acknowledgements

So many people in my life have been involved in getting me here, and I would like to take the opportunity to thank them and recognize their efforts.

A big thank you goes to my advisor, Richard Page, for all of the support, time and effort put in so that I would succeed at Penn State and elsewhere. Your investment in me as a teacher and scholar encouraged me to find and follow the path that is right for me. The confidence you have in me and my abilities means a lot to me, and I’m lucky to have had an advisor that I feel comfortable with and who I work well with.

I would also like to thank the members of my dissertation committee, Carrie Jackson, Mike Putnam, Giuli Dussias and Mary O’Brien. Each one of you has generously contributed your time and expertise to me and this project. Thank you to Carrie for always being eager to help me when I wandered into your office; your feedback and advice has always been valuable to me. Mike, thank you for believing in me and the other Penn Staters, and always being ready to give encouragement. Giuli, I appreciate your agreeing to join my committee in the final hour and providing me with questions to guide the analysis of my results. Thank you to Mary for being a guiding force and mentor for me throughout my graduate career. I appreciate all the time, support and feedback you have given me throughout the years, whether it involved reading my articles, helping me figure out normalization, or sponsoring my trip to Calgary. All five of you have really had my best interests at heart, something I know that not everyone can say. I will never forget all of your support as I have grown up at Penn State, both personally and as a teacher and scholar. Lastly, I would also like to thank Chip Gerfen for all of his help and guidance on the design of this dissertation project during my candidacy and comprehensive exams.
I would like to thank the committees of the Language Learning Dissertation support grant and the RGSO grant from the College of the Liberal Arts, without which I would not have been able to complete my dissertation research as efficiently as I have. This support was invaluable.

In addition, I am forever grateful to Niki McInteer and Courtney Fowler. Without your hard work and commitment to my pronunciation unit, I wouldn’t have any data to work with, and therefore, no dissertation. Thank you!

My friends have been instrumental in making my time here at Penn State one of the best of my life. Thank you all for being by my side throughout it all—without you all I wouldn’t have made it! Alison, I’m so glad that you’ve been by my side for this whole thing. Your support then and now can’t be measured, and a thank you isn’t enough! Niki, Atia & Janice: thank you for keeping me sane through the worst times. You three were my rock when I needed it most and I will never forget it. Thank you, Nick, for listening to me whine about anything and everything technology related for the last couple of years. You were instrumental in helping me solve my million-and-one E-Prime crises and stats questions! Matt, you have been such a great help when looking at the bigger picture that I sometimes miss, and you are exceptionally good at talking me down off the ledge. Thanks for listening to me talk (and talk and talk and talk) about the little things, and the big things too. You played a big part in getting this thing done.

And finally, to my parents. Thank you for teaching me to take pride in always doing my best. You have always supported my education and my choices throughout all of their various and long long loooooong steps. Thanks for proof-reading my papers, being guinea pigs for conference presentations, and helping me figure out what my next step should be. Oh, and thanks for still letting me take my laundry home to wash for free. I owe you. 😊
Chapter 1: Introduction and Background Literature

1.1 Introduction

It is quite obvious to anyone who has tried to learn or teach a foreign language that mastering a new phonological system is a very difficult task. Even highly proficient speakers with native-like ability in syntax, fluency or lexicon often continue to speak with a foreign accent. In some cases, this foreign accent can be so strong as to interfere with communication, leading to misunderstandings, confusion, or even a full-fledged communication breakdown. Yet despite its importance for communication, pronunciation is often overlooked in foreign language classrooms. The reasons for this vary from the belief that it is impossible to teach an adult second language (L2) learner to improve his or her accent at all, to the opinion that pronunciation skills are not an essential part of communicative competence. Especially when confronted with time constraints in the basic language classroom, many instructors adhere to the belief that pronunciation errors will eventually “work themselves out” at higher proficiency levels or after some time abroad. For these reasons, pronunciation instruction is typically undertaken, if at all, in more advanced courses such as conversation or phonetics courses (e.g., Counselman, 2010; Derwing, Munro & Wiebe, 1998; Elliott, 1995, 1997; Lord, 2008).

However, research conducted with L2 learners in the areas of pronunciation training and Second Language Acquisition (SLA) suggests that most adult L2 learners are not able to solve pronunciation problems on their own. They may not even notice or be able to articulate which areas of their speech need more work, therefore it seems necessary for instructors to assist them in coming to this realization by providing instruction about pronunciation in their L2. The following chapter will provide a motivation for teaching pronunciation in the classroom, give a brief overview of previous pronunciation training studies that showed successful improvement,
and provide an outline of various theories in SLA that lend support to the idea of pronunciation instruction for L2 learners.

1.2 Why Pronunciation?

As mentioned previously, unaccented L2 pronunciation is a nearly impossible skill for late L2 learners to acquire, and even advanced L2 speakers may speak with an accent strong enough that it interferes with their comprehensibility (i.e., their ability to be understood by an interlocutor). Yet it is important that all L2 speakers learn adequate pronunciation; there seems to be a threshold for pronunciation skills that is vital for successful communication of meaning by L2 speakers (Hinofotis and Bailey, 1980). Without reaching this basic level, late L2 learners may face difficulties or possibly even an inability to express themselves to other speakers of their target language (Hinofotis and Bailey, 1980; James, 1976).

Despite the important relationship between meaning and pronunciation, many L2 learners may not be aware how their pronunciation can and does affect their ability to communicate in their target language. In a survey by Derwing and Rossiter (2002), 39% of 100 adult immigrants in English as a Second Language (ESL) classes could not identify any specific areas of their speech that caused problems in communication. Of those students who could identify problems, 84% focused only on segmental errors, often those that are salient but carry a low functional load, generally defined as a measure of how frequently a contrast between two phonemes is responsible for distinguishing minimal pairs in a language (see: Brown, 1991). An example of a segment in English with low functional load in English is [ə], which was the segment most commonly identified as needing improvement by the students in Derwing and Rossiter (2002). In a similar vein, Dlaska and Krekeler (2008) reveal that L2 learners were able to recognize less
than half of their inaccurate productions when asked to listen back to their own recordings of familiar German words and assess their pronunciation. Learners’ self-assessments of those inaccurate sounds did not correlate with evaluations of accuracy by native speaker judges. The L2 learners in this study had so much difficulty with self-assessment, in fact, that they even rated some of their accurate productions to be inaccurate, while at the same time failing to recognize the genuine pronunciation errors made. Taken together, these studies confirm that self-monitoring of pronunciation seems to be a very difficult skill for L2 learners to acquire and use, and therefore self-assessment is unreliable. It seems that learners’ awareness of their own pronunciation errors or problem areas is very low to non-existent. Without any awareness of their own mistakes, asking students to improve their pronunciation on their own seems an incredibly challenging request.

This difficulty is reflected in studies investigating the pronunciation of learners both within and outside the classroom. Research has found that pronunciation ability tends to remain stable for classroom learners in the absence of formal instruction (Counselman, 2010; Ducate & Lomicka, 2009; Elliott, 1995, 1997). Similarly, naturalistic and immersed L2 learners often fail to develop new L2 segments or prosodic features beyond initial levels, even after many years of exposure (Grosser, 1997; Jilka, 1999; Munro & Derwing, 2008; Trofimovich & Baker, 2006). Munro and Derwing (2008) appraised the pronunciation of English vowels by low-proficiency learners immersed in an L2 environment. Their findings show that over the course of one year, speakers did in fact show improvement, but largely only within the first six months of immersion. After the great gains made within these first six months, very little improvement was seen. This rapid initial improvement seems to be typical in the early stages of an L2 immersion experience, and is most obvious in the area of segmental accuracy (Flege, 1988; Flege, Munro &
Skelton, 1992). Thus, the Munro and Derwing (2008) study concludes that L2 experience and an immersed setting are not always enough to lead to accurate vowel productions on their own.

It may be the case, however, that additional improvement required more time for exposure than a mere year. In another study investigating pronunciation by immersed learners, Trofimovich & Baker (2006) explore the effects of short- and long-term L2 experience on adult learners' acquisition of various L2 suprasegmental features. Three groups of Korean learners of English with similar daily patterns of L2 use were formed based on length of residence in the United States, ranging from 3 months to about 10 years. Based on native speaker ratings of their low-pass filtered utterances, the inexperienced learners received significantly lower ratings than the two more experienced groups, yet all were rated significantly lower than native English speakers. These results suggest that even with up to ten years of L2 immersion, there remain some features of suprasegmental pronunciation that elude L2 learners and are not acquired with target-like accuracy. In fact, out of the five suprasegmental features chosen for analysis, only stress timing was found to relate to their amount of L2 experience. Thus, the authors conclude that suprasegmental features of pronunciation pose a difficult learning challenge for L2 learners, even with very large amounts of L2 experience. Taken together, these studies imply that even in long-term immersion environments, pronunciation errors will not always simply "work themselves out", as so many instructors assume.

Even among related languages that may share some prosodic features or intonation patterns, such as German and English, L2 learners may have difficulty realizing just how their pronunciation differs from that of native speakers. Grosser (1997) completed a longitudinal study of Austrian children learning English and found that the learners did not develop native-like prosodic patterns in their English. Curiously, the prosodic patterns they used in their L2 did
not occur natively in either German or English. This was the case even when the same or very similar intonational patterns would be used in comparable contexts in both languages and transfer from the first language may not have been terribly detrimental. For example, the L2 learners used more sentential peak accents when speaking in their L2 English than is normal in both native English and their own L1. They also often applied broad focus to their utterances where narrow focus would be appropriate, and vice versa. Thus, the authors conclude that L2 learners have difficulties recognizing differences between their own and target-like pronunciations, even when appropriate patterns are the same or very similar between the two languages.

In conclusion, pronunciation is very important for the communication of meaning in any language. Yet the aforementioned studies suggest that students are generally unaware of linguistic features in their speech that need improvement, and self-assessment of their own pronunciation is usually unreliable at best, and completely inaccurate at worst. Even with many years of L2 experience and immersion, learners may have difficulties acquiring L2 segmental and prosodic patterns. As L2 learners do not typically acquire good pronunciation automatically, one can hypothesize that they need some assistance to improve. It is possible that without help, they may never be able to fix pronunciation problems on their own.

1.3 Previous Pronunciation Training Studies

It is with these difficulties in mind that many researchers have turned their attention to the topic of L2 pronunciation training. The general idea behind much of this research is that if L2 learners cannot acquire target-like pronunciation even after many years of L2 experience, then it may be necessary to train them in this skill. There exists a large and growing body of
research addressing a variety questions surrounding the concept of pronunciation training. The general question at the heart of every pronunciation training study is: can L2 pronunciation be improved with training? Moving a step beyond this, a significant portion of the literature also investigates the details of pronunciation training, answering such questions as: 1) Which methods are most effective in a pronunciation intervention? or 2) Which segmental or suprasegmental features should be trained?

Many studies have compared participants’ L2 pronunciation both before and after explicit pronunciation training and found that a significant improvement of pronunciation in the L2 is a possible and attainable goal for L2 learners. Improvement in pronunciation is measured in a variety of ways—via acoustic analysis of the learners’ productions, ratings of phonetic accuracy provided by phoneticians, native speaker perceptions of things like the strength of accent and the ease of understanding (i.e., comprehensibility), or by changes in how many words out of a given utterance a native speaker can actually understand. More recently, emphasis in pronunciation training studies has shifted away from reducing accentedness and improving phonetic accuracy towards what Levis (2005) names the “intelligibility principle”. This principle states that the goal of instructors and researchers when providing instruction in pronunciation should be to make sure that L2 learners can be understood. The goal of more recently completed training programs is to make learners’ L2 speech more understandable, rather than more native-like (Levis, 2005; p. 370). Improvement in such studies is typically assessed through ratings of comprehensibility, defined as "a listener's perception of how difficult it is to understand an utterance" (Derwing & Munro, 2005, p. 385), and transcription tests that measure intelligibility, or “the extent to which a listener actually understands an utterance” (Derwing & Munro, 2005, p. 385). Becoming more understandable is slowly being understood as a more reasonable
expectation for L2 learners than holding them to a native-like standard which they may never achieve.

The language laboratory has been found to be a successful location in which to train L2 pronunciation, showing positive results and significant improvement of phonetic accuracy and comprehensibility under very controlled settings (Akahane-Yamada et al., 1996; Flege, 1989a, 1989b; Guion and Pederson, 2007; Hardison, 2004; Iverson et al., 2005; Thomson, 2011). It is not absolutely necessary to train L2 learners in an organized laboratory setting, however—many other studies applied intervention methods successfully within the less controlled setting of the L2 classroom (Botero, 2011; Counselman, 2010; Derwing, Munro & Wiebe, 1998; Elliott, 1995, 1997; James, 1976; Lord, 2005, 2008; Neufeld, 1977; Saito & Lyster, 2011; Tanner & Landon, 2009). Such a large body of previous research in the area of pronunciation training suggests that interventions can help improve a variety of pronunciation skills and abilities in multiple different settings.

The types of interventions used in pronunciation training studies vary widely, and there is as yet no agreement on which methods are most effective in obtaining a significant improvement in pronunciation. Typically, studies focus on one of two areas—on the development of the skills that are required for the acquisition of pronunciation, or on the methods used for the intervention. When addressing the skills essential for pronunciation improvement, previous studies often emphasize either perception-based or production-based methods. Studies focusing on the best mode of intervention often concentrate on methods using computer software or the internet, audio and visual aids, or a combination of some or all of the above.

Studies focused on training only an L2 learner’s phonological perception skills are based on the hypothesis from the segmental literature that a learner’s phonological perception skills are
linked to their production abilities in the L2 (Flege, 1995; Ohala, 1981). For example, the Speech Learning Model, a model of phonological perception, posits that adult L2 learners must be able to correctly perceive L2 segments and prosodic patterns before accurate production of these features can occur (Flege, 1995). The ability to discriminate target sounds, does not guarantee the ability to produce them accurately, but the SLM predicts target-like production only for those learners who have established precise mental representations for segments in their L2, termed L2 phonological categories. A key component of the SLM is that learners can develop their L2 phonological categories through language experience and training. It is therefore projected that by training perception, L2 learners can extend this learning to their production of the corresponding sounds or prosodic patterns as well. A study by Bradlow et al. (1997) demonstrates that perceptual training can lead to improvement in production of the same phonemes, even when no production training has been conducted. Japanese speakers learning English received three to four weeks of perceptual training of the /l/-/r/ contrast in English, and participants who perceived the phonemes more accurately after training also produced them in a more target-like fashion. This study and others (Botero, 2011; Counselman, 2010; Flege, 1989a), confirm that perceptual training alone can in fact be successful for the improvement of L2 production.

Production-based intervention methods, on the other hand, rely on plenty of oral practice to help L2 learners improve their production of various segmental and suprasegmental features. This production practice is often combined with explicit presentations about how to produce the target L2 sounds and speech patterns, which may include visual, kinesthetic, or aural presentations (Elliott, 1995, 1997; Flege, 1989b; Hardison, 2004).
One increasingly popular way to conduct a production-based pronunciation intervention is through the use of computer programs or the internet (de Bot, 1983; Ducate & Lomicka, 2009; Eskenazi, 1999; Hardison, 2004; Lord, 2008; Tanner & Landon, 2009; also see O’Brien, 2011 for a review). Within this training method, computer programs use a variety of methods, including automatic speech recognition or visualization software to provide audio, visual, or audio-visual training to learners. For example, visualization software programs typically take on the role of feedback through visual pitch contour displays and allow learners to compare their own utterances to a native speaker target, sometimes even in real time. These seem to be particularly useful when training L2 learners in prosodic features like intonation. Hardison (2004) used visual pitch contour displays in contrasting colors over three weeks of training as feedback for students practicing French intonation. Although training was only received for prosody, results showed a significant improvement for the learners in both prosody and segmental accuracy using this computer program. Improvement also generalized to prosodic and segmental accuracy of novel sentences, further showing the benefits of the use of computer software for pronunciation training.

In addition to researching the in-class or during-training methods that are effective for pronunciation training, many studies examine the use of self- and peer assessment and feedback as a method to improve pronunciation skills. As mentioned above, many L2 learners may not be aware of deficits or errors in their pronunciation and without this knowledge may be unable to identify or correct them (Derwing & Rossiter, 2002; Dlaska & Krekeler, 2008). Therefore, self-assessment of pronunciation seems to be somewhat fruitless without any other intervention. Partner assessment, however, has been shown to be an effective method of feedback for L2 learners in the areas of oral proficiency and spoken fluency (Miller & Ng, 1996), as well as non-
auditory skills like writing (Haughton & Dickinson, 1988; Jacobs, 1988). This method has been extended to pronunciation training studies as well, with positive results.

A unique perception-based method of pronunciation instruction is employed in Counselman (2010), in which students received a two-part intervention. Students received in-class pronunciation instruction, produced recordings of their own speech, and for homework listened to the audio recordings of another class’s speech and gave appropriate corrective feedback. Students were instructed to attend to a specific set of target forms in their peers’ outputs each week. They then gave feedback on these errors and any others that they found to be prominent. Students in this training program were thus able to significantly improve their own pronunciation through the process of noticing and giving feedback on someone else’s. Students in the control group, who did not provide peer feedback for another class but received general pronunciation feedback from the teacher, did not show improvement from pretest to posttest. In another study, Lord (2008) combines the use of technology in pronunciation training with student feedback by incorporating pronunciation-themed podcast assignments into an undergraduate Spanish phonetics class. After recording each production assignment, students analyzed both their own pronunciation and that of their other group members. As a whole, students significantly improved both their pronunciation and their overall attitude toward the importance of correct pronunciation from pretest to posttest.

It is important to note, however, that student feedback alone does not seem to be enough to provide students the opportunity for significant improvements over the course of a semester. Results from both Ramírez Verdugo (2006) and Ducate and Lomicka (2009) demonstrate that without explicit training in class, students are not always able to identify errors in their own speech or in that of others. Ducate and Lomicka (2009) also make use of podcasting and student
feedback to practice pronunciation skills outside of class. Students created eight of their own podcasts with both scripted and extemporaneous speech, then listened to their classmates’ and provided feedback on the content of the podcasts. After completion of the eight podcasts, students failed to show significant improvement in either comprehensibility or accentedness between pre- and posttest when rated by native speakers of the target language. It is possible that this intervention using student feedback failed to be effective because of the lack of explicit training and teacher feedback. Students did not receive any in-class pronunciation instruction, which deprived them of the skills and knowledge both to provide effective peer feedback and to make changes in their own speech. In addition, student attention was focused on the content of the podcasts, and not their peers' pronunciation. It remains to be seen how participants would have fared with more explicit instruction from their teachers, or by focusing their attention on the form of their partners' work, rather than on the content.

Another research question of great importance in the L2 pronunciation training literature is whether it is more beneficial to train learners in segments or in prosody. Traditionally, most pronunciation training studies have successfully focused on improving the production of individual phonemes, especially shibboleths that are particularly salient at marking nonnative speech. Yet research on L2 accent and its perception by native speakers has demonstrated the importance of prosody for communication. Prosodic information is so salient to native speakers that they are able to identify foreign-accented speech on the basis of suprasegmental information alone, as evidenced by native speaker ratings of low-pass filtered speech (Munro, 1995). A growing body of research suggests that errors in prosody may even influence native listener perceptions of an accent to a greater extent than the pronunciation of individual phonemes. Errors in prosody, both lexically (Caspers, 2010; Field, 2005) and sententially or more globally
(Anderson-Hsieh et al., 1992; Hirschfeld, 1994; O'Brien, 2004; Trofimovich & Baker, 2006; Wilkerson, 2007), have been found to be more disruptive to native speaker listeners than errors in the articulation of individual phonemes. The prosodic patterns of L2 speakers can also influence how comprehensible and successful their communication with native speakers is. For example, results from Hirschfeld’s (1994) study of L2 German show that native speaker comprehension of sentences produced by L2 speakers with inaccurate lexical or sentential stress fell to 48.8%, whereas it stood at 95% when the correct stress patterns were used. An example is the German sentence *Es regnet*, wherein the stress is placed on the syllable [nɛt] instead of the appropriate syllable [reːg]. Such errors were also labeled as more “irritating” to native speaker listeners than segmental errors (p. 119). Her findings show that even if words in L2 speech were reproduced with the same segmental errors but the correct stress information, the comprehensibility greatly improved. She concludes that the placement of lexical stress must have a special meaning for interpretation of the speech signal, demonstrating once again the importance of accurate prosodic pronunciation for communication.

For all of the above reasons, good intonation and accurate placement of lexical and sentential stress can be interpreted as just as, if not more, important for intelligibility and comprehensibility of the L2 in communication as segmental accuracy (Derwing & Munro, 2005). Therefore, prosodic training is a recommended and necessary part of any pronunciation training unit, no matter what the chosen methods.

A number of experimental studies confirm the importance of suprasegmental training for L2 learners and suggest that prosodic training is so effective that it may even be beneficial when standing alone. A study by Derwing, Munro and Wiebe (1998) is one of the most cited pronunciation training studies and compares two types of classroom pronunciation training units
for effectiveness. The first experimental group received only segmental training, while the other received a more global approach focusing on suprasegmental aspects of language. Students in the global pronunciation group showed significant improvements on comprehensibility and fluency for both sentences and narrative readings, while those in the segmental group only improved significantly on sentence comprehensibility. Derwing et al. conclude that suprasegmental training is more effective because of its generalizability to other tasks, including more difficult ones such as spontaneous speech. Training in segments alone seems to be less generalizable, and therefore less effective as a stand-alone training program. Hardison (2004) and Ramírez Verdugo (2006) support these findings, as their results reveal that training in L2 prosody can lead to more global gains as well. In both of these studies, participants were also able to improve their pronunciation of individual segments after prosodic training, without receiving any focused training in segmental pronunciation. By raising their awareness of the prosodic patterns of their respective L2s, participants were also paying more attention to, and focusing on improving, the individual sounds as well. Therefore, it is likely that in addition to being more generalizable to new speech situations, prosodic training can also lead to the improvement of pronunciation of individual L2 segments.

While many pronunciation training studies have shown success in improving L2 learners’ productions, others fail to reveal significant improvement from pronunciation training (Ducate & Lomicka, 2009; Gilbert, 1980; Purcell & Suter, 1980). A study by Gilbert (1980) fails to find any effects of training on learners’ ability to identify languages by prosody alone and to mimic native speaker prosody, but these results should be interpreted with caution. This study provided only very short periods of training, which makes it unclear how participants would have performed with longer training periods. In addition to studies which fail to find effects of
training, there is another group of research in the body of literature arguing that explicit pronunciation training and training in production is not even necessary. Studies in this vein assert that large amounts of auditory instruction without production practice may suffice for students to improve their pronunciation (McCandless & Winitz, 1986; Neufeld, 1977). Neufeld (1977) indicates that half of participants were rated by native speaker judges as native or near-native after extensive visual and auditory training without any production on the part of the students. Yet a major criticism of this study is that students received no grammar or vocabulary instruction along with the input. They could not actively use or understand the sentences they were producing, which puts into question what exactly they were acquiring, and if these results would stand once the language was being actively used and understood. Thomson (2011) updates these auditory input methods and supports the use of high variability phonetic training, in which learners are exposed to high levels of L2 input from a large set of different voices producing target sounds. This method has been shown to help students improve their perception of target sounds and contrasts, which then in turn helps them improve their own intelligibility and increase their ability to self-monitor, even without explicit pronunciation instruction (Thomson, 2011). Taken together, these papers demonstrate that explicit production instruction may not always be necessary for improvement, yet they also point to the fact that more research is needed to find the optimal methods that will encourage success.

Despite the wealth of pronunciation instruction studies comparing methods of instruction, the role of feedback, and useful target features, very little research has tackled the question of when the best time to train might be. Waiting until students have more advanced language skills to even begin to address pronunciation is the most prevalent method in the field (e.g., Botero, 2011; Counselman, 2010; Ducate & Lomicka, 2009; Elliott, 1995, 1997; Hardison, 2004; Lord,
2008), yet SLA research suggests that this may not be the best choice. It may in fact be more difficult to reverse pronunciation problems once a certain level of proficiency has been reached (Elliott, 1995; Hammerly, 1982; Maldonado, 1994; Storch, 1994). If pronunciation errors are not addressed as they arise, students may reach a certain level of performance and then not progress beyond it, a phenomenon known as fossilization (Selinker, 1972).

Second language learners who exhibit fossilization continue to use non-target-like structures in what is called their *interlanguage*. This interlanguage shows a number of marked differences when compared with the target language productions of native speakers. Learner interlanguage does not necessarily include any features of the first language or the target language. Incorrect structures that have become fossilized remain available as possible productions for L2 learners’ speech and are used even if such structures would be considered an error in the target language. Pronunciation is an L2 skill that is just as prone to fossilization as syntactic structures if students’ pronunciation errors are not corrected early on (Tarone, 1980). Even with correction, students often continue to produce the incorrect forms from their interlanguage and will have frequent exposure to these forms as a source of language input. Such exposure and frequent input may even *reinforce* their pronunciation errors, leading to phonological fossilization (Elliott, 1995).

In order to mitigate potential fossilization, more and more scholars are suggesting that it may be desirable to begin pronunciation training as early as possible (Counselman, 2010; Derwing & Munro, 2013; Elliott, 1995; Eskenazi, 1999; Hardison, 2004; Maldonado, 1994; Munro, 2013). Counselman’s (2010) research on pronunciation instruction suggests that it may be easier to adjust pronunciation of newly acquired words than that of words learned early on in L2 instruction. This view is seconded by Elliott (1995) and Maldonado (1994), who propose that
it may be more beneficial to the students and less time-consuming for the instructors if pronunciation is stressed from the beginning, as students at upper intermediate levels of proficiency already show some evidence of phonological fossilization that will need intense effort to be reversed.

Beginning pronunciation training as early in the language learning sequence as possible may also have other, indirect results on a learner’s skills in the L2. Results from Hardison (2004) suggest that prosodic instruction beginning in the first year of language study may provide students with the ability to improve their segmental accuracy. Therefore, an indirect result of training in the first semesters may be the improvement of pronunciation training features that were never trained. Eskenazi (1999) also proposes that the benefits of training in the first semester could extend to untrained areas, positing that training in the first semesters may help L2 learners acquire syntactic structures (Eskenazi, 1999). She argues that speakers preparing an utterance must simultaneously organize segments, prosody and syntax and correct any deviations that do not fit within trained prosodic forms, thereby reinforcing to students which word order or segments are correct.

Despite the multitude of recommendations to begin pronunciation training as early in the language program as possible, and the potential advantages to doing so, no previous studies (to this author’s knowledge) have experimentally investigated the effects of pronunciation instruction for beginners.

To summarize, previous pronunciation training research provides many insights into the correct methods and targets for pronunciation instruction, which can act as guidelines for the design of an intervention study. Firstly, Derwing, Munro and Wiebe (1998), Hardison (2004), and Ramírez Verdugo (2006) all demonstrate the importance of prosodic training in improving
L2 comprehensibility and fluency. This type of suprasegmental training is so important because it may be more generalizable to complex tasks, such as spontaneous speech, than segmental training. Results from Hardison (2004) show that training in L2 prosody may even help learners improve their pronunciation of individual segments, and Eskenazi (1999) suggests that prosodic instruction may assist beginning L2 learners in their acquisition of new syntax. These studies highlight the importance of including suprasegmental training in pronunciation instruction units. Training studies by Counselman (2010) and Lord (2008) focus on the success of noticing and student feedback. Participants in these studies attended to their peer’s pronunciation and gave focused feedback, which helped them notice and adjust similar problem areas in their own speech. Previous research has compared a variety of training methods and phonetic targets, but one area that has yet to be addressed is that of the proficiency level at which an intervention is most effective. Pronunciation training beginning in L2 learners’ first few semesters has many potential benefits, from preventing fossilization of incorrect segmental and prosodic forms, to potentially helping students improve their acquisition of syntax (Eskenazi, 1999).

1.4 Attention and Motivation in Second Language Acquisition

The idea mentioned in section 1.2, namely that L2 learners often need assistance in acquiring target phonetic forms, is also a concept in Second Language Acquisition (SLA) more generally. For this reason, theories involving awareness and attention in language learning have a long and disputed history within the field of SLA. It seems that attention is beneficial for learning a language, although it is still unclear to what extent attention and awareness are absolutely necessary for learning. This section will discuss the theoretical background of attention and awareness in second language learning, in order to provide an explanation as to
why pronunciation instruction is effective and why L2 learners might need it to improve their skills.

1.4.1 Noticing Hypothesis

Schmidt’s (1990) Noticing Hypothesis addresses this problematic relationship between L2 input and actual L2 acquisition, and is derived from his observation that L2 learners may not be able to learn important linguistic structures in their second language unless they "notice" them. Noticing is defined within this hypothesis as a focal awareness that divides what is detected or perceived in the input from what is noticed and what the learner has a conscious awareness of. Noticing is necessary for L2 learners to develop this conscious understanding of the target language and its linguistic features in order to acquire the features and produce them accurately in their own speech.

In order to notice, students must first become aware of a stimulus in the input. L2 learners are exposed to large amounts of input that can be used for learning, but is not possible to process every signal in this input. Noticing allows learners to control what parts of the input they let in for further processing. Whatever passes through the learners’ filters, or in Schmidt’s words, “that part of the input that the learner notices” (p. 139), is called intake. A feature in the input must be noticed before it becomes intake, continues on for further processing, and can be used for the acquisition process. This is where the divide between what was taught or what the learner was exposed to and what was actually learned occurs.

With the Noticing Hypothesis, Schmidt argues that noticing and attention are prerequisites for all aspects of L2 learning in adults; anything that is not noticed cannot be learned. Schmidt perceives the first step of L2 learning as the realization that something is
different between the two languages and has to be learned; therefore noticing is the catalyst for every other stage of learning. Global attention, or learning as a subconscious process, may not be sufficient for adult second language learning (Schmidt, 1990, 1995); forms must be attended to in order to be truly learned. Schmidt (2001) acknowledges that subliminal perception is possible, as items or actions can be perceived and alter someone's behavior or thinking without consciousness. However, according to the Noticing Hypothesis, a second language cannot be learned subliminally by adult learners, because it requires the creation of new structures in the brain. Unattended stimuli may be stored in short-term memory, but will not be committed to long-term memory without attention and noticing (Schmidt, 1995). A conscious awareness of the features and their characteristics must be present for further steps towards acquisition to occur.

This noticing and conscious awareness can occur naturally during the language learning process, but Schmidt (2001) posits there are some areas of language learning in which it may be necessary to externally and intentionally focus learner attention on the necessary forms. This is especially true in situations where the target language has coded information that is not attended to at all in the L1, such as lexical tone in Mandarin, and when a certain feature is automatically processed in the L1 without allowing learners the time necessary to focus on their own. Schmidt (2001) argues that new L2 routines can only be established once L2 learners' attention is focused intentionally on the differences and the features of the L2 segments.

In sum, the strong version of the Noticing Hypothesis (Schmidt, 2001) advocates that there can be no acquisition of target L2 forms when the input is processed for meaning alone, without attention and noticing of linguistic forms. Adult L2 learners will not acquire target language forms from the input unless they consciously notice them and realize their importance
in the target language. This noticing can be brought about naturally in the language learning process, but in some instances, must be intentionally made obvious to the learner.

1.4.2 Further Development of the Concept of Attention in SLA

Since its inception, Schmidt's Noticing Hypothesis has been heavily criticized and fallen out of favor. Common oppositions include the fact that it is not possible to measure what an L2 learner has or has not noticed (Gass, 1997), that conscious noticing leads to greater metalinguistic awareness but not necessarily improved competence in the L2 (Truscott, 1998), and that global awareness of the input without conscious noticing can suffice (Truscott, 1998).

In a critical review of the Noticing Hypothesis, Truscott (1998) argues that conscious noticing of L2 structures in the input is not a prerequisite for all language learning to occur. Attention is certainly beneficial for language learning, but global awareness of the input is sufficient in order for L2 learning to occur. He also argues that learners can become aware of gaps in their knowledge, but that this only improves their metalinguistic awareness and that actual language performance does not necessarily benefit. Another key critique of the Noticing Hypothesis, proposed by Gass (1997), is its stance that all noticing is conscious and that language features automatically become intake once they are noticed by the learner. Gass (1997) argues that grammatical forms can be apperceived in the input and processed for meaning even without a learner’s conscious registration of them. Yet the Noticing Hypothesis does not consider these forms noticed until the learner can actually articulate what he or she noticed. Therefore, the major points of contention within the Noticing Hypothesis are whether a theory of SLA should rely so heavily on conscious awareness and whether such awareness is even necessary for learning to occur.
It is true that many critique Schmidt's (1990) Noticing Hypothesis, but this does not mean that attention and awareness are not still considered important aspects in SLA. Indeed, the concepts of attention, noticing, and awareness in language learning remain popular in theories of SLA. A more commonly accepted view of attention in SLA holds that attention and noticing are helpful to SLA but are not absolutely necessary. Details of each individual form do not have to be consciously noticed in order for them to be learned. In addition, other factors, such as input, output, and interaction also play large roles in learning a second language.

A more commonly accepted model of SLA is the Interaction Hypothesis, proposed by Long (1996), which combines the important components of input, interaction and output. This hypothesis is composed of five steps through which L2 input is transformed into output: apperceived input, comprehended input, intake, integration, and output. Apperception is the first stage of language learning; it is the stage where learners realize that there is something to be learned. This is similar to Schmidt and Frota's (1986) concept of “Noticing the Gap”, wherein learners notice the differences between their version of the L2 and that produced by native speakers. In the Interaction Hypothesis, a learner must first apperceive a feature in the input as being different than what they know or what they produce. The apperceived input then progresses through the comprehended input stage, where pieces of the input are understood either for meaning or at a deeper level of syntactic analysis. Afterwards, this input becomes intake, which is processed in the brain and matched against prior knowledge. The integration stage takes what has been processed by the brain and either puts it into storage or uses it to develop the interlanguage grammar. Output is the end product of the acquisition process, when learners begin to produce what they have acquired. Output can be used to test hypotheses or to incorporate feedback into their own grammatical systems. Using the acquired forms in output...
also focuses learners’ attention on both meaning and syntactic processing of the language as utterances are constructed. Interaction and negotiation are also key features of this hypothesis; actually using the language in interactions also assists learners in focusing their attention on parts of the target language that they were not aware of, such as mismatches between their interlanguage and the target language, as they attempt to communicate meaning (Gass & Selinker, 2008). This awareness leads to reassessment of the interlanguage grammar and triggers the five-step process once again.

In this model, attention and awareness are key still components, yet conscious awareness and articulation of what has been noticed is no longer perceived as a requirement for learning. Learners can be aware of the contents of a message that they receive as input, but they do not need to have an awareness of exactly what piece of the input has been processed in order to acquire a language feature (Gass, 1997). Tomlin & Villa (1994) support this viewpoint by arguing that detection, or the registration of a stimulus that is important for further learning and acquisition, can take place without awareness. Therefore, SLA research has generally agreed that no matter which form it takes, the concept of attention remains important for language acquisition.

1.4.3 Attention and Noticing in the Classroom

There are many ways in which the concepts of attention and noticing from SLA research can be utilized in the language classroom. Interaction between learners and explicit instruction are two ways that teachers can make sure that maximal attention is paid to key forms in the input and increase the likelihood of noticing (Gass & Selinker, 2008; Long, 1996; Schmidt, 1995). Additionally, Form-Focused Instruction can manipulate learner attention and make key language
features more salient so that students can notice and process them (Gass & Selinker, 2008). In this type of instruction, communicative, meaning-based assignments are supplemented with instruction on grammatical form. Form-Focused Instruction is effective for bringing mismatches to a learner’s attention by making L2 features more salient and easier for learners to attend to and learn.

There are a number of factors that must be considered when focusing student attention within the setting of the L2 classroom, including decisions about what to make salient, how to incorporate negative evidence, and at which proficiency level this is most effective. According to Harley (1993), Form-Focused Instruction should center predominantly on features that “(a) differ in nonobvious ways for the learner, (b) are irregular, infrequent, or lacking perceptual salience, and (c) do not carry a heavy communicative load” (Harley, 1993, p. 256). It is these features that are most often ignored or subject to misanalysis by learners, therefore it is these features that most benefit from analytic instruction.

In addition to increasing the salience of positive evidence, or what is a possible use of the language, negative evidence can be an important teaching and learning tool. Negative evidence is defined as evidence of what is not allowed in the target language, for instance, which word orders are ungrammatical or which phonological segments are not allowed. Although positive evidence may be sufficient for children to learn a language, adults may need negative evidence to supplement their learning, reorganize their interlanguage structures, and solidify target forms. Results from Trahey and White (1993) reveal that after receiving large amounts of naturalistic input with only positive evidence, French learners of English still were not able to reorganize their interlanguage to forbid forms that were unacceptable in the target language. It seems that
positive evidence, though beneficial, is not sufficient for adult L2 learners to learn what is unacceptable in the target language.

Lastly, research points to the fact that L2 proficiency level has an influence on attention and noticing of L2 forms. Williams (1999) reveals that learners at lower levels of L2 proficiency often do not attend to important language forms independently. This is due in part to the high cognitive demands necessary, especially at beginner levels, for learners to communicate even simple utterances in the L2. He thus concludes that “learner-generated attention to form may not always come naturally and, clearly, may require some pedagogical training” (Williams, 1999, p. 381). Gass, Svetics and Lemelin (2003) discovered in their investigation of focused attention in language learning that focused attention has a greater effect in the early stages of language learning. As L2 learners gain proficiency over time, there is a diminished effect for focused attention on language forms. From these two studies one can infer that L2 learners in beginner levels of proficiency are the least likely to attend to language forms on their own, and also the most likely to benefit from some sort of focused attention training in the classroom.

It therefore seems that the best type of classroom tasks are those whose design draws attention to the form through frequency or perceptual saliency and creates a strong link between the task demands to communicate and accuracy of form. For maximum success, communicative and language-focused instruction should draw learner's attention to target forms in the input that are irregular and unlikely to be noticed by learners. Consciousness-raising activities should also be supplemented with explicit negative evidence. This combination can provide enough evidence for students to notice the differences between the way they produce the L2 and the way that they should be producing the L2. In addition, this combination may be most effective when introduced into classrooms at the lower proficiency levels.
1.4.3.1 Noticing and Attention in Pronunciation Studies

Although the majority of studies conducted on noticing and attention and their extension to the classroom setting have been completed with grammatical features in mind (e.g., Alanen, 1995; Jourdenais et al., 1995), there exists a small set of studies that investigates the use of noticing and attention to improve L2 pronunciation. These studies find that focusing students’ attention on L2 sounds and orienting them to phonetic forms are beneficial for improvement of L2 segmental and prosodic pronunciation (Counselman, 2010; Guion and Pederson, 2007; Pennington & Ellis, 2000; Ramírez Verdugo, 2006; Saito & Lyster, 2011).

Pennington and Ellis (2000) investigate the role of attention in prosodic learning by testing Cantonese speakers learning English on their ability to be trained in English intonational contours such as focus, tags, and boundary contours. After a study session in which their attention was directed to the intonation and meaning contrasts associated with each sentence type, participants had to judge whether follow-up sentences had the same prosodic structure as the original ones they had heard. Results show that awareness of English intonational structures could be improved by focusing learners’ attention during training. However, this attentional focusing was only effective in helping learners improve on the simplest type of intonational contrast tested, the ones used to mark focus. The authors argue that this type of prosodic structure is of a universal and transparent type, and more research is needed to investigate the effects of attention on pronunciation training.

A study by Ramírez Verdugo (2006) also investigates L2 intonation patterns and maintains that L2 learners often begin by using their L1 intonation patterns in the L2. This can be useful because it allows them to communicate and convey meaning, but may actually hinder intelligibility in the L2 when overgeneralized. After a multi-sensory training unit designed to
raise the participants’ awareness of L2 intonation, participants in the experimental group used a wider variety of tones in more appropriate ways and produced important words with more prominent sentential stress. Their intelligibility ratings also increased significantly. The author thus concludes that raising L2 learners’ awareness of their intonation in the L2 is attainable, and can lead to an improvement in intelligibility. She stresses the importance of prosodic instruction, stating that L2 learners must be made aware of how L2 prosody differs from that used in their L1 (p. 142), because they do not seem to be able to recognize the prosodic patterns of their L2 speech without such training.

From these studies one may assert that without explicit training and focusing of attention, even advanced L2 learners will not attend to all of the important phonological information in either their own or others’ speech. Their incorrect L2 pronunciation might be due to a lack of attention to details of L2 phones or prosodic patterns, but concepts of attention and noticing can be successfully applied to L2 pronunciation by raising L2 learners’ awareness of phonological forms.

1.4.4 Motivation in Second Language Learning

In addition to the type of instruction, learners’ levels of motivation have been found to influence language learning by driving a student to begin and continue learning a foreign language. Each individual student differs in his effort and motivation to reach the same goal, but a wealth of research demonstrates higher levels of motivation to learn a language can positively influence language learning and language proficiency (e.g., Dörnyei, 2005; Gardner, 1985; Ramage, 1990; Smit, 2002; Tremblay & Gardner, 1995).
One of the first prominent models of motivation was Gardner’s (1985) Socio-Educational Model. This model assesses the relationships between attitudes toward learning a foreign language and attainment in that language. Two orientations define this model and set the direction for L2 learners’ goals: integrative and instrumental. Integrative orientation is defined as a drive to learn a second language “in order to learn about, interact with, or become closer to, the second language community” (Gardner, 1985; p. 54). Instrumental orientations are those which prime students to learn a language in order to receive some reward or practical benefit. Overall, Gardner’s research (1985) suggests that an integrative orientation is more helpful in acquiring an L2 than an instrumental orientation. Although originally designed to measure motivation among Canadian students learning French in school, this model has been generalized to other second language contexts. The Socio-Educational Model has suitably measured the motivation levels of Jewish students in Israel learning Arabic as a foreign language (Kraemer, 1993), Polish students learning Hebrew as an optional L2 (Okuniewska et al., 2010), and sixth grade students participating in French immersion schooling in the United States (Wesely, 2009). However, the notion that the desire for contact and identification with the L2 community has the most influence on L2 proficiency has been heavily critiqued. Many students who are learning a foreign language do not have significant (if any) contact with speakers of their target language or language community, and it is argued that integrative orientation may be far less relevant for students who are not learning in an immersion or contact environment (Dörnyei, 1990).

Critiques of the Socio-Educational Model often stress the importance of including situation-specific and individual motives into a theory of motivation for L2 learning (Crookes & Schmidt, 1991; Dörnyei, 1990; Oxford & Shearin, 1994), although beyond this point, there is little agreement on which factors play the largest role in learner motivation. Noels (2003)
adapted Self-Determination Theory and intrinsic and extrinsic motivation from general psychology (Deci & Ryan, 1985) to create a system that can measure motivation in language learning. Dörnyei (1990) and Schmidt et al. (1996) stress that an instrumental orientation may be more influential for the achievement of students learning in a foreign language situation who have little or no contact with their L2 community. Setting specific and challenging goals has been shown to influence levels of motivation and motivational effort (Tremblay & Gardner, 1995), while Clement et al. (1994) attributes high levels of motivation to high self-confidence. The L2 Self Model, very influential in recent years, argues that envisioned future selves influence motivational behavior and effort (Dörnyei, 2003, 2005; Kormos & Csizer, 2008). Motivation in this model stems from the desire to reduce distance between one’s current self and one’s future L2 self. This self can either be created by strong outside influences (the “Ought-to Self”), or by the individual himself, and his values (the “Ideal Self”). As one can see, as yet there is no one accepted viewpoint about what factors best increase learner motivation and proficiency.

Although studies have also investigated the relationship between motivation and L2 pronunciation more specifically, this area is much less developed than that of motivation models and general language learning outcomes. On one side of the coin, it is claimed that motivation has no significant impact on pronunciation abilities (Purcell and Suter, 1980). Coates (1986) even goes so far as to say that motivation cannot be measured and connected with pronunciation. Yet other scholars have found a strong positive correlation between motivation to improve accent and overall accentedness. Moyer (2007) found that for international students at American universities, a combination of strong personal and professional motivation proved most successful for students to achieve a more native-like L2 accent than any single type of
motivation in and of itself. She also suggests that a desire to perfect the language rather than assimilate into the target language culture may be more important for improving L2 accent specifically, contradicting Gardner’s (1985) original Socio-Educational Model. Smit (2002), while finding less robust correlations between motivation and students’ grades in a pronunciation course, also suggests that a positive attitude and intrinsic motivation aid success in pronunciation. Although native speaker ratings or acoustic analyses were not used in the previous two studies, final achievement in the course was related to the motivation and confidence levels of the individual learners.

Effects of motivation level and student attitudes have also been found in experimental studies that measure pronunciation improvement after a classroom training unit. In a study similar to the current one, Elliott (1995) discovered a significant correlation between L2 learners’ attitudes towards learning pronunciation and their ultimate level of pronunciation accuracy following a classroom training unit. Importantly, Elliott (1995) also devised a useful measurement of student attitudes towards pronunciation specifically, as opposed to language learning in general. This tool, called the Pronunciation Attitude Inventory, measures student attitudes in a self-reported questionnaire. It asks questions about students’ desire to sound native-like, their opinions about the importance of pronunciation in communication and its role in the classroom, and their beliefs about their own ability to improve their pronunciation. This valuable tool is still used by many scholars who research the effects of pronunciation instruction in the classroom and has unearthed the effects of attitude on pronunciation in many studies since its debut (e.g., Botero, 2011; Counselman, 2010; Ducate & Lomicka, 2009; Lord, 2008).

In sum, motivation levels and student attitudes have been found not only to affect language learning and proficiency in general, but also L2 learners’ performance in the area of
pronunciation. The connection between motivation and pronunciation has been found in connection with pronunciation training studies (Elliott, 1995), or in the classroom more generally (Moyer, 1999; Smit, 2002).

1.5 Conclusion

Acquiring target-like pronunciation in an L2 is a notoriously difficult task that even advanced speakers may never achieve. Good pronunciation is fundamental to the communication of meaning, but adult L2 learners often cannot improve their pronunciation on their own, even with many years of L2 experience and immersion. Therefore, some sort of training in pronunciation may be necessary for learners to improve their skills. A number of pronunciation training programs have been found to be effective at helping L2 learners improve their pronunciation skills. Important findings in the research show that both perception- and production-based methodology can be effective, that peer feedback plays an important role in improving one’s own pronunciation, and that prosodic training is a vital component in a successful pronunciation training unit.

Attention and noticing are key components in the acquisition of language in general, and of a new phonological system in particular. Two hypotheses that deal with the idea of noticing and learner attention as being important for acquiring and accurately using linguistic forms were outlined in this chapter, the Noticing Hypothesis and the Interaction Hypothesis. These hypotheses differ in their estimations of how significant a role noticing plays, but one thing that ties them together is the idea that attention and noticing are very important for learning language forms. They also agree that attention and noticing can be manipulated in the classroom for
beneficial effects. A handful of studies have directly investigated the importance of focusing attention on pronunciation, with successful results.

On the basis of previous research in pronunciation training and noticing in SLA, one can hypothesize that the most successful type of pronunciation training unit would be one that uses a combination of perception- and production- based methods to train a variety of suprasegmental and segmental features. This training unit would use the concepts of noticing and attention from SLA research to supplement communicative activities with language-focused instruction on phonological forms, and would provide students with both negative and positive evidence of what is or is not allowed in the L2. Such a training unit would help students to pay attention to the details of L2 pronunciation that they may otherwise ignore, and by raising their phonological awareness, improvement in pronunciation can be anticipated. It can be predicted that those who are more motivated to improve their pronunciation will demonstrate more improvement over time than those who are less motivated.
Chapter 2: Focus of the Current Study

The current study directly compares the effectiveness of pronunciation training in first- and fourth-semester German language classrooms in order to address a gap in the pronunciation training literature. A communicative pronunciation training unit was designed as a supplement to a traditional four-skills German classroom. This training unit incorporates elements from many of the studies mentioned in section 1.3, including two previously successful methods, peer feedback and consciousness-raising (Counselman, 2010; Lord, 2008; Pennington & Ellis, 2000). The intervention consists of a global method of instruction including training of the perception and production of segmental and suprasegmental features. Accepted methods in pronunciation training research were used for data collection and analysis. Ratings of comprehensibility and accentedness were elicited from native speaker listeners to compare pre- and post-training pronunciation. Finally, the effects of training at both the elementary and intermediate levels is compared in order to find the most effective time course for intervention.

2.1 Research Questions

The current study addresses three research questions. As was mentioned in the previous chapter, many different methods and optimal target features have been investigated in the pronunciation training literature, but one major area that has yet to be researched is the timing of the intervention. There exists a gap in the literature, wherein the most effective time-course for introducing a pronunciation intervention is yet to be determined.

Thus, the main research questions of this study are the following:

1. Can a pronunciation training unit for elementary and intermediate learners in their first and fourth semesters be effective in significantly improving pronunciation when compared to a control group?
2. Is intervention more effective at the elementary or intermediate level?
The current study differs from previous research in significant ways and therefore adds to current discussions in the literature. Participants in previous studies have been predominantly learners at the intermediate level of language study, while the pronunciation training unit in this study is aimed at learners both at the beginner and intermediate levels of proficiency. This study also differs in terms of design, as it will combine training in both segmental and suprasegmental features of the language, rather than providing training either in one area or the other (Botero, 2011; Counselman, 2010; Elliott, 1995; Hardison, 2004; Lord, 2008; Saito & Lyster, 2011), or directly comparing the two types of training (Derwing, Munro & Wiebe, 1998). The planned intervention will also be incorporated into a regular four-skills language course, not a specialized pronunciation or phonetics course. The design of this study combines findings and suggestions from the pronunciation training literature and various hypotheses from the area of SLA. It takes into account SLA research on attention and noticing, previous successful classroom training designs, and standard research methods for pronunciation training in order to address a new question.

The main goal of the training unit is to aid communication in two ways: 1) improve comprehensibility by increasing the accuracy of lexical stress placement and 2) reduce accentedness by increasing pronunciation accuracy of the target segments. The current training unit is designed to investigate whether it is in fact more beneficial to focus on pronunciation as early in the curriculum as possible, as some scholars have suggested, or if the effects are greater at more advanced levels. In addition, it is designed to be an easy-to-implement type of intervention that teachers of all kinds could use in their German classrooms, even those with minimal technical training in phonetics or phonology.
As numerous studies have shown that student attitudes can influence the acquisition of a second language (Dörnyei, 2005; Elliott, 1995; Gardner, 1985; Moyer, 1999), the current dissertation also deals with an additional, secondary research question:

3. Do students’ attitudes towards pronunciation affect their progress in pronunciation improvement throughout the course of the semester?

It is important to measure students’ attitudes about pronunciation and learning a second language due to the possible influence of these factors on their progress. Students’ attitudes will be correlated with the rating data in order to discover any effects.

2.2 Hypotheses

The predictions and hypotheses for these research questions are as follows:

1) It is expected that the pronunciation training unit will help focus students’ attention on L2 sounds and will therefore lead to greater improvement over the course of the semester. Students who receive training will show greater improvement in pronunciation than those who do not, independent of which semester they are currently enrolled in. This hypothesis is supported by the numerous pronunciation training studies, described in section 1.3, in which L2 learners have demonstrated improvement in pronunciation skills as a result of training (Botero, 2011; Counselman, 2010; de Bot, 1983; Derwing, Munro & Wiebe, 1998; Elliott, 1995, 1997; Flege, 1989a, 1989b; Guion and Pederson, 2007; Hardison, 2004; Iverson et al., 2005; Lord, 2005, 2008; Ramírez Verdugo, 2006; Saito & Lyster, 2011; Tanner & Landon, 2009).

2) It is anticipated that students in the first semester will show improvement equivalent to, or perhaps even surpassing that of, students in the fourth semester. This hypothesis is supported by the numerous suggestions in the literature that pronunciation training, as well as focused attentional training, may be more effective early on in a language learning sequence
(Counselman, 2010; Elliott, 1995; Eskenazi, 1999; Gass, Svetics and Lemelin, 2003; Hardison, 2004; Maldonado, 1994). If L2 learners in their first semester of study demonstrate improvement surpassing that of students in their fourth semester, it will lend support to these suggestions, as well as to the hypothesis that pronunciation training may be most beneficial to learners just beginning their study of the language.

3) The third hypothesis predicts that students with more positive attitudes toward pronunciation and learning German will improve more than those with more neutral or negative attitudes. Thus, students who show more motivation to learn pronunciation and place greater importance on learning to speak with good pronunciation will be more receptive to and improve more through training than students who do not hold the same attitudes toward pronunciation.

2.3 Pronunciation Targets Chosen for Training

In-class training consisted of a combination of prosodic and segmental elements. Six German-language segments were chosen for training: [ç], [x], [u], [r], [e:], and [o:]. These sounds were chosen because they encompass a variety of sounds that are typically seen as difficult to pronounce for new speakers of German. In addition, many of these segments serve as strong shibboleths for American English speakers acquiring German. A mixture of vowel and consonant sounds was chosen, because O’Brien (2004) found that a group of native speakers listening to American L2 learners of German determined rhotics, [x] and [ç] to be even more distracting and badly pronounced than vowel sounds. American-like productions of the consonant /r/, such as the retroflex [ɾ], are a prominent marker of an American English accent in German and are also found to interfere with comprehension of nonnative speech by native listeners in languages other than German, such as Spanish (Shairer, 1992). The [ç] and [x] distinction is important because students must not only acquire pronunciation of two new sounds
not occurring in English, they must also learn the alternation between them. The vowels /e/ and /o/ were included in this study because these are sounds that do occur in English but are pronounced slightly differently in German, a difference which many L2 speakers find challenging and do not recognize or produce (see Flege, 1995). It is expected that improvement in the pronunciation of these segments will affect a learners’ perceived accentedness in their L2 productions.

In addition to these six segments, a seventh focus was also tested: lexical stress. Lexical stress is included as a suprasegmental feature, as it is crucial for word recognition and comprehension (Caspers, 2010; Field, 2005). For example, findings from Field (2005) reveal that lexical stress assignment on the wrong syllable had drastic consequences for interlocutor comprehension. As such, improving the accuracy of lexical stress placement is expected to help improve L2 learner comprehensibility in German. English and German lexical stress patterns are similar in many ways, yet stress assignment can cause problems for L2 students. L2 speakers often overgeneralize English stress patterns onto German words, especially among the many cognate pairs that exhibit different stress patterns in German, such as ’mu.sic vs. Muˈsik.

2.4 Differences between English and German Phonology
The aforementioned target sounds chosen for training in the current study differ from English vowels in a variety of ways. These are outlined in detail in each sub-section below. The sub-sections also describe aspects that are particularly problematic for an American English accent in German, as well as potential solutions.
2.4.1 Lexical Stress

Lexical stress is a linguistic means of making certain syllables of a word prominent in the speech stream (Chun, 2002). The location of lexical stress plays a crucial role in helping listeners to recognize the words they hear, and without its accurate assignment, comprehension can be severely impeded (Caspers, 2010; Field, 2005; Hirschfeld, 1994).

English and German have very similar stress systems, although differences can and often do occur. These two languages are similar in that every word has a set pattern of lexical stress and that stress is applied in the same ways—through a combination of pitch, duration and intensity. Of these, vowel duration has repeatedly been found to be the most robust cue to mark stress in English and German (Dogil, 1999; Okobi, 2006; Sluijter & Van Heuven, 1996), and stressed vowels in these two languages are always produced with a longer total duration than unstressed vowels (Jessen et al., 1995).

Due to these similarities in how stress is realized in both languages, English learners of German rarely have trouble with the concept of lexical stress. Instead, their main difficulty lies in deciding which syllable to stress. Stress placement cannot be predicted from the combination of segments in a word alone and varies according to the word’s structure, syllable count, and linguistic origin. It seems that stress may be applied to any syllable, and therefore stress assignment in German can seem random to an L2 learner for a variety of reasons:

1) Even almost identical phonetic shapes may show different stress patterns.

   (1) Al’tar          ‘Alter

2) Placement of stress may shift with the addition of affixes or plural forms.

   (2) Pro’fessor       Profes’soren

   (3) ’Nation         natio’nal
This situation is further complicated by the existence of many disyllabic cognate pairs in which primary stress is placed on the first syllable in English, but on the second syllable in German.

(4) ˈmoment Moˈment
ˈsalad Saˈlat
ˈmusic Muˈsik

In many cases, when L2 learners are unsure of or do not have enough information about the stress patterns or segments in particular words, they will resort to L1 substitutions (Wiese, 1996; Zampini, 2008). This is especially common for English-German learners, who often overgeneralize their L1 English stress patterns onto German words. The similarities between English and German stress realizations may serve as a trap of sorts for English L2 learners of German, especially if they come to believe they may simply apply their L1 stress patterns to the L2. Yet simply transferring all knowledge of their L1 into the L2 will lead to problems in stress placement, and therefore a potentially unintelligible foreign accent. L2 learners must learn which cognates have a different stress pattern in German, and learn to switch stress assignment to a different syllable.

2.4.2 Ich and Ach -- [ç] and [x]
The sounds oftentimes represented as “Ich-Laut” and “Ach-Laut”, or the voiceless palatal and velar fricatives [ç] and [x], are new sounds for English native speakers learning German.¹ These two fricatives appear in complementary distribution, and for the purposes of this dissertation are considered allophones of the same phoneme. The allophone [x] occurs only after central and back vowels and diphthongs, such as [a:], [o:], [u:], or [au], while [ç] occurs

¹ The Ich-Laut ([ç]) does appear in certain environments in English, such as in the words hue or human, but speakers are oftentimes not aware of its use.
everywhere else, including after front vowels, such as [i:] and [e:], after the consonants /n, l, r/, and sometimes in word-initial position (Moulton, 1962). Some examples of the environments for [x] are found in (5) and for [ç] in (6), along with their English glosses:

(5)

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>hoch</td>
<td>high</td>
</tr>
<tr>
<td>Buch</td>
<td>book</td>
</tr>
<tr>
<td>auch</td>
<td>also</td>
</tr>
</tbody>
</table>

(6)

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>dich</td>
<td>you (accusative case)</td>
</tr>
<tr>
<td>Milch</td>
<td>milk</td>
</tr>
<tr>
<td>China</td>
<td>China</td>
</tr>
</tbody>
</table>

English speakers learning German are faced with two difficult tasks when acquiring these new German sounds. They must first learn how to physically articulate the new sounds, one of which they are not at all accustomed to producing, and then they must learn the appropriate distribution for them. When faced with these difficult tasks, L1 English speakers often substitute sounds from their L1 instead. Most American English speakers typically substitute [k] for [x], while the substitutions for [ç] vary (Moulton, 1962). Some speakers will substitute a [ʃ], because it is the closest acoustic equivalent to the German sound, while some will also extend the [k] substitution to this allophone. Another, although rarer, substitution for these two fricatives involves the use of [tʃ], which is most likely an influence of English orthography on their pronunciation of German, as both German fricatives are spelled with a <ch>. Substitutions of L1 sounds in place of difficult L2 sounds lend themselves to a foreign accent and may impede comprehensibility. For example, if L2 learners substitute the segment [k] for [x], native speakers will have greater difficulty in determining what the learner intended to say for many minimal pairs such as *Nacht* (night) or *nackt* (naked).
2.4.3 Fricative and Vocalic /r/

The German /r/ is pronounced with a variety of allophones, none of which occur natively in English and all of which differ considerably from the usual retroflex [ɹ] produced by native American English speakers. Allophones of German /r/, similar to the palatal and velar fricatives mentioned in the previous section, are in complementary distribution and consist of two main types: consonantal and vocalic /r/. Vocalic /r/, realized as [ʁ], is used exclusively after vowels and occurs at the end of words and syllables. This allophone of /r/ is never employed in word- or syllable-initial position; that position is occupied by consonantal /r/. Examples demonstrating the distribution of German /r/ can be seen in (7).

(7) rot [ʁo:t] red
    fahren [fa:ʁn] to drive
    aber [a:br] but
    verstehen [fɛɐʃteŋ] to understand

The allophonic distribution of German /r/ is based on phonological environment alone, not on the root of the word. (8) demonstrates how even words with the same root will be pronounced with different allophones of /r/ based on the position of /r/ in the word:

(8) besser [bɛʁ] --- bessere [bɛʁəʁ] better
    Tier [tɪɛ] --- Tiere [tɪʁə] animal(s)

The consonantal allophone of /r/ is pronounced in different ways by different speakers of German. In Modern Standard German, the voiced uvular fricative [ʁ] is used, although in many dialects an apical trill [r] or a uvular trill [ʁ] may be used. None of the three variants of consonantal /r/ used in German-speaking areas have an equivalent segment in the English language. In contrast, there is only one allophone of /r/ used in English. This is a retroflex [ɹ], and is used in all phonological environments in the English language. The segment is
pronounced by curling the tongue tip back towards the back of the mouth, which produces a fluid sound that is very different than the fricatives and trills used in German.

Thus, American English speakers learning German have to learn both new articulation and distribution patterns, while at the same time suppressing the allophone of /r/ they are used to producing. This is particularly difficult because of the desire to carry over L1 segments to L2 sounds, particularly if the same letter is used in the German orthography. Students of L2 German must learn two new ways to pronounce this consonant, and learn the allophonic distribution of the many German /r/s. The /r/ is potentially the single most robust cue to native speakers that their interlocutor is not a native speaker, and this is particularly so of the American /r/ (Hall, 1992; Schairer, 1992). Only by resisting the impulse to simply substitute an L1 version of /r/ for the /r/ in German can American L2 learner’s accent be improved.

2.4.4 The Tense Mid Vowels [e:] and [o:]

One problematic feature of learning to pronounce L2 segments surfaces when both the L1 and L2 contain similar vowels. According to Flege’s (1995) Speech Learning Model, these cross-language vowels oftentimes appear so similar to learners that they are mapped onto the same phonological categories and therefore pronounced simply as an L1 vowel. The L1 mapping process makes these segments easier for L2 learners to process and use, but oftentimes leads to inaccuracies in production. Examples from German are the tense mid vowels [e:] and [o:]. Although English and German both contain these two mid vowels in their vowel inventories, the articulation of these vowels differs between languages. Mapping the German sound onto an English vowel category and simply substituting the L1 vowel for that in the L2 is not sufficient for target-like German pronunciation.
The biggest difference in articulation between English and German front and back mid vowels lies in their movement in the vowel space as the vowel is articulated. Namely, these vowels are diphthongized in English, but are pronounced as monophthongs in German. This means that in German, the formants of the front and back mid vowels are sustained at the same levels throughout the duration of the vowel. In English, there is a clear movement of the tongue, producing an off-glide towards the end of the articulation of the vowel. Consider:

(9) say [set]
gate [gett]
so [sou]
goat [goot]

This offglide causes the formant frequencies to change. The first formant (F1) corresponds to vowel height, and the second formant (F2) corresponds to tongue advancement. The F1 frequencies for English mid vowels drop in Hertz value as the vowel is produced, whereas those in German typically do not. Many American English learners of German seem unaware of this difference and pronounce German mid vowels with the same diphthongized pronunciation that they use in English (Hall, 1992). This causes an accented pronunciation in German, which will not be remedied until L2 learners stop substituting their English-like mid vowels for German vowels. They must begin pronouncing monophthongized versions of the German mid vowels [e:] and [o:] that are free of off-glides.

In sum, six German segments and one prosodic feature were chosen for training in this study. All of these pronunciation features are hard for American English L2 learners of German to pronounce, and all lend themselves to a strong foreign accent in L2 German. The following
chapter outlines the methodology and analysis of the training unit and the experimental study used to measure results.
Chapter 3: Methodology and Analysis

The purpose of the current study is to compare the effectiveness of pronunciation training on students in both first- and fourth-semester basic German language classes. Two sections of each proficiency level were selected to participate in the study. One section of each proficiency level was designated as the experimental group, and the other as the control. Experimental groups received 10 minutes of in-class pronunciation training exercises per class period over the course of 8 weeks, while the control group received normal instruction without any pronunciation training. All groups completed pronunciation assignments with a partner and gave their partner feedback on their pronunciation. Thus, each group received one of two in-class interventions: a) in-class pronunciation activities and partner assignments or b) the partner assignments only. Data was collected from an experimental study consisting of identical pre- and posttests in which participants recorded themselves producing German in a variety of discourse contexts. These recordings were rated for comprehensibility and accentedness by native German speaker judges.

The following chapter discusses the methodology of the study. It begins with a discussion of the four participant groups. A description of the materials and stimuli used is explained next, and the chapter finishes with an account of the procedures used during the pre- and posttest, the pronunciation intervention, and the rating experiment.

3.1 Participants

All participants were enrolled in German courses at a large university in the Northeastern United States. Four intact classes took part in the study: two sections of a first-semester German course, and two sections of a fourth-semester German course. Within each proficiency level, each section and its students were assigned to either the experimental or the control group. A
total of 60 participants took part in the study, but sixteen were excluded from analysis. Seven participants were excluded for having a native language other than English, two were excluded for accumulating too many absences during the training period, five were excluded for not completing all necessary assignments, and two participants in the fourth-semester groups reported being brought up as English-German bilinguals. By the end of the training and rating procedures, 44 participants remained in the study.

3.1.1 First-Semester German

The experimental section of the first-semester German course contained a total of 13 participants (eight male, five female), with an average age of 19.8. All were native speakers of American English, including one participant who grew up as a Spanish-English bilingual. Eleven of the students had not previously studied German before enrolling in this course, therefore their average age of learning German began at 19.0. Two students in this section had studied German at the high school level, but were not excluded because this previous study was an average of 2 years before the time of this study and their performance did not significantly differ from that of the other students in their group. Only two participants report having spent time abroad in a German-speaking country, and this experience was only for an average of three weeks’ time. Eight participants reported having at least occasional contact with a German native speaker. Participants provided self-ratings of proficiency levels in German for all four skills on a scale from 1-10. Their ratings stood at 4.0 for reading, 3.31 for writing, 2.62 for speaking, and 2.31 for listening, while the same participants’ self-ratings for these skills in English stood at 9.62, 9.54, 9.92 and 10, respectively.

Eleven of the participants (ten male, one female) were in the control section of the first-semester proficiency level. All were native American English speakers, and their mean age was
21.8. The average age of learning was 18.3 years, because three participants had studied German at the high school level for an average of 2.2 years. These participants were not excluded because two of them were in their late twenties and early thirties, therefore this study occurred an average of twelve years ago. Two participants from this section had both spent 6 weeks living in Germany the summer before enrolling in this course. Only four participants reported any contact with a German-speaking friend, relative or acquaintance. The participants self-rated their skills in German reading, writing, speaking, and listening at 3.46, 2.73, 2.36, and 2.91, respectively. Their average self-ratings for English were as follows: reading was rated at a 9.77, writing at 9.18, speaking at 9.55, and listening at 9.46.

3.1.2 Fourth-Semester German
Ten participants (seven male, three female) were members of the experimental section of the fourth-semester German course. All ten were native American English speakers and the average age of participants in this section was 22.4. As opposed to the first-semester students, the majority of students in the fourth-semester courses had begun studying German in middle or high school. Fourth-semester experimental participants had studied German for an average of 3.83 years before enrolling in their current course, with an average age of learning of 18.7 years. Three students had previously studied abroad in Germany and had spent an average of four weeks in the country. Seven of the participants reported contact with German-speaking acquaintances, friends and relatives. Somewhat expectedly, participants self-rated their proficiency levels in reading, writing, speaking and listening in German as much higher than the students in the first-semester had. Their average self-ratings for reading in German were 6.28, writing 6.83, speaking 6.83, and listening 6.17. Their average self-ratings for reading and
listening in English both stood at 10, and they rated themselves at 9.78 for both writing and speaking in English.

The control section of fourth-semester German also contained ten participants (six male, four female). The average age of all participants in this section was 22.2 years, and all were native speakers of American English. The participants in this section had spent an average of 3.6 years studying German before this intermediate course, with an average age of learning German beginning at 15.0 years. Two participants reported having studied abroad in Germany for an average of 5 weeks. Six of the participants reported having contact with German-speaking acquaintances, friends and relatives. Self-ratings of proficiency levels in German stood at 4.90 for reading and writing, 5.40 for speaking, and 6.10 for listening. Self-ratings for the same skills in English stood at 9.8, 9.5, 9.7 and 9.9, respectively.

Table 1 shows a summary of the descriptive details for each section, while Table 2 provides the averaged self-ratings for each skill in German and in English for each group.

Table 1

<table>
<thead>
<tr>
<th>Description</th>
<th>First Semester</th>
<th>First Semester</th>
<th>Fourth Semester</th>
<th>Fourth Semester</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
<td>Experimental</td>
<td>Control</td>
</tr>
<tr>
<td>Number</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Average Age</td>
<td>19.8</td>
<td>21.8</td>
<td>22.4</td>
<td>22.2</td>
</tr>
<tr>
<td>Average Age of Learning</td>
<td>19.0</td>
<td>18.3</td>
<td>18.7</td>
<td>15.0</td>
</tr>
<tr>
<td>Native Languages</td>
<td>English (12)</td>
<td>English (12)</td>
<td>English (10)</td>
<td>English (11)</td>
</tr>
<tr>
<td></td>
<td>English/ Spanish (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># who Studied Abroad</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Average Time Spent Abroad</td>
<td>3 weeks</td>
<td>6 weeks</td>
<td>4 weeks</td>
<td>5 weeks</td>
</tr>
<tr>
<td># who Have Contact with Native German Speakers</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 2

Self-Rated Proficiency Levels in German and English

<table>
<thead>
<tr>
<th></th>
<th>First Semester</th>
<th></th>
<th>Fourth Semester</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
<td>Experimental</td>
<td>Control</td>
</tr>
<tr>
<td>German Reading</td>
<td>4.0</td>
<td>3.46</td>
<td>6.28</td>
<td>4.90</td>
</tr>
<tr>
<td>German Writing</td>
<td>3.31</td>
<td>2.73</td>
<td>6.83</td>
<td>4.90</td>
</tr>
<tr>
<td>German Speaking</td>
<td>2.62</td>
<td>2.36</td>
<td>6.83</td>
<td>5.40</td>
</tr>
<tr>
<td>German Listening</td>
<td>2.31</td>
<td>2.91</td>
<td>6.17</td>
<td>6.10</td>
</tr>
<tr>
<td>English Reading</td>
<td>9.62</td>
<td>9.77</td>
<td>10</td>
<td>9.80</td>
</tr>
<tr>
<td>English Writing</td>
<td>9.54</td>
<td>9.18</td>
<td>9.78</td>
<td>9.50</td>
</tr>
<tr>
<td>English Speaking</td>
<td>9.92</td>
<td>9.55</td>
<td>9.78</td>
<td>9.70</td>
</tr>
<tr>
<td>English Listening</td>
<td>10</td>
<td>9.46</td>
<td>10</td>
<td>9.90</td>
</tr>
</tbody>
</table>

3.1.3 Instructors

In order to control teaching methods and classroom environment as much as possible, both the experimental and control sections of each course level were taught by the same instructor. Two different instructors were used in the experiment: one for both first-semester sections, and another for both fourth-semester sections. Both instructors were advanced graduate student teaching assistants in the German program at the same university. Both had multiple years of teaching experience (mean: 4.5 years), and both had previously taught introductory German courses at the university level. Both instructors also had experience with pronunciation courses. The instructor for the two first-semester German sections had previously taught a pronunciation course in English. The instructor for both sections of fourth-semester German had been a student in a German pronunciation course at the Freie Universität Berlin, and was therefore familiar with different methods of teaching German pronunciation. Although both instructors were native American English speakers, they had been studying German for an average of 15.5 years, and had spent an average of 3.425 years living in German-speaking countries. The table below outlines all of the descriptive details about these two instructors.
### Table 3

**Descriptive Statistics about Pronunciation Instructors**

<table>
<thead>
<tr>
<th></th>
<th>First-Semester German Instructor</th>
<th>Fourth Semester German Instructor</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native Language</td>
<td>Midwestern American English</td>
<td>Midatlantic American English</td>
<td></td>
</tr>
<tr>
<td>Years of Teaching Experience</td>
<td>4</td>
<td>5</td>
<td>4.5</td>
</tr>
<tr>
<td>Types of Teaching Experience</td>
<td>University German – 2 years</td>
<td>University German – 5 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University English – 1 year</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kindergarten English – 1 year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has Taught a Pronunciation Course</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Has Taken a Pronunciation Course</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Beginning Age of Studying German</td>
<td>17.5</td>
<td>11</td>
<td>14.25</td>
</tr>
<tr>
<td>Years Spent Studying German</td>
<td>13.5</td>
<td>17.5</td>
<td>15.5</td>
</tr>
<tr>
<td>Years Spent Living Abroad</td>
<td>4.1 (4 years, 2 months)</td>
<td>2.75 (2 years, 9 months)</td>
<td>3.425 (3 years, 5 months)</td>
</tr>
</tbody>
</table>

Both experimental sections were taught from 11:15 a.m. to 12:05 p.m. This was controlled in order to ensure that the experimental sections would be as similar as possible. The control sections varied in their timing, however. The fourth-semester control section was taught from 10:10-11:00 in the morning, while the control section of the first-semester class was taught from 2:30-3:20 in the afternoon.

#### 3.2 Stimuli

The experiment followed a pre- and posttest design. The pre- and posttests were identical and tested pronunciation at the word, paragraph, and free speech levels. Participants first completed a perception test contrasting German and English sounds, and then completed the production component of the test, which contained three separate tasks. Task one was to read a
list of 75 words aloud. Participants then read six short paragraphs containing the same test words in longer sentences and answered 5 open-ended questions. In addition to recordings of these stimuli, participants filled out two questionnaires: an adapted Pronunciation Attitude Inventory (Elliott, 1995) and a Language Background Questionnaire.

The pronunciation features chosen for training included six German-language segments, \[\varsigma\], \[x\], \[v\], \[r\], \[e:\], and \[o:\], and lexical stress assignment. Within this lexical stress component, all test words fell into one of four sub-conditions. These lexical stress sub-conditions included disyllabic cognates and noncognates with stress either on the first or second syllable of the word. All test stimuli fell under either one of the six target segment conditions or one of the four lexical stress conditions, outlined below in Table 4.

| Table 4

<table>
<thead>
<tr>
<th>Example Stimuli for each Training Condition</th>
<th>Example Stimulus</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Stress: Cognates with 1(^{st}) syllable stress</td>
<td>Lampe</td>
<td>4</td>
</tr>
<tr>
<td>Lexical Stress: Noncognates with 1(^{st}) syllable stress</td>
<td>Abend</td>
<td>4</td>
</tr>
<tr>
<td>Lexical Stress: Cognates with 2(^{nd}) syllable stress</td>
<td>Student</td>
<td>4</td>
</tr>
<tr>
<td>Lexical Stress: Noncognates with 2(^{nd}) syllable stress</td>
<td>bestimmt</td>
<td>4</td>
</tr>
<tr>
<td>[\varsigma]</td>
<td>ich</td>
<td>8</td>
</tr>
<tr>
<td>[x]</td>
<td>machen</td>
<td>8</td>
</tr>
<tr>
<td>[v]</td>
<td>Zimmer</td>
<td>8</td>
</tr>
<tr>
<td>[r]</td>
<td>Reise</td>
<td>8</td>
</tr>
<tr>
<td>[e:]</td>
<td>lesen</td>
<td>8</td>
</tr>
<tr>
<td>[o:]</td>
<td>wohnen</td>
<td>7</td>
</tr>
</tbody>
</table>

Each target segment was tested, to the greatest extent possible, in a variety of phonological environments and different positions in the word. For instance, \[x\] can only be found following back vowels, but the test stimuli feature this segment following as many back vowels as possible, such as \[a\], \[u\], \[ɔ\], and \[aʊ\]. The two vowel conditions, \[e:\] and \[o:\], were tested in a variety of open and closed syllables in similar phonological environments. \[e:\] was
tested word-finally (n=2) and before nasals (n=3), the fricative [z] (n=1), and alveolar and velar stops (n=2). [oː] was tested word finally (n=1), and syllable-finally before nasals (n=3), the fricatives [z] and [h] (n=2), and a bilabial stop (n=1).

Test stimuli consisted of a combination of content and function words with either one, two or three syllables. All words for all tests were real German words chosen from the first two chapters of *Vorsprung*, the first-year German textbook used at the participants’ university. The most important criterion for choosing test stimuli was to use real words that had been covered in the first-semester German classes before the beginning of training. It was deemed most important that students were familiar with the test stimuli before training began, because the purpose of the study is to measure improvement in pronunciation, not acquisition of or familiarity with vocabulary. Presenting completely novel words to first-semester German students would not give an accurate baseline to measure improvement. In that case, any improvement effects could be based on participants’ lack of knowledge of the pronunciation of a particular word before beginning training and not on the effectiveness of the intervention itself. Therefore, familiar words were chosen, and this necessity took precedence over criteria when selecting test stimuli. There were only a few exceptions to this rule. For instance, no tokens for the “Noncognates with 2nd syllable stress” condition were present within the first two chapters of *Vorsprung*. Thus, all words for this condition came from chapters 3 and 4. One token from the [ʁ] condition and one from the [ç] condition also came from Chapter 3, but both of these words (*fahren* and *mich*) have very high frequency and were likely to have been mentioned in the classroom before reaching Chapter 3 in the lessons.
3.2.1 Perception Test

The perception test contained a total of 28 real German words out of the first two chapters of the *Vorsprung* textbook. Four stimuli were chosen for each of the seven test conditions: the six shibboleth segments, and one category of lexical stress. Cognates with second syllable stress were chosen for use in the lexical stress category on the perception test, due to their high contrast with English stress assignment patterns.

The design of the perceptual test was in part inspired by that used by Botero (2011), in which participants listened to recordings of different pronunciations of the /l/ phoneme and were instructed to identify them as ‘Spanish’ or ‘English’. The goal of the perceptual task in the current study was similar to that in Botero (2011) – to raise students’ awareness of the differences between target sounds in German, in this case between those pronounced with a Standard German or an American English-influenced accent. The task was designed in this way to promote the noticing of differences in pronunciation between a native and a nonnative accent.

Audio files used as stimuli in the perception test were prepared by nine different speakers, six female and three male. Two (both female) were native speakers of German, and seven were native speakers of English. All native speakers of English recorded for this task were either graduate students or professors of German linguistics, and therefore were fluent in German and well-versed in German pronunciation. The native English speakers had studied, taught, or otherwise used the German language for an average of 14.83 years, with a range of 6 to 33 years. The average age of the male speakers was 37 years, with a range from 27 to 51. The average age of the female speakers was 29.83 years, with a range from 23 to 37. The two native German speakers had studied English actively for an average of 7.5 years, and both had spent time immersed in an English speaking environment for an average of 7 months.
Each speaker was instructed to first read the German words chosen for stimuli in this perception test in their best Standard German accent. After this, these speakers were instructed to read each German word and pronounce the target segments in a standard American English manner. It was made very clear to these speakers to only alter the target sound; all other sounds were to be pronounced in Standard German. In this way it could be assured that students were only focusing on the target sound and not using any other sounds in the word to make their decision. For example, one of the words to target the consonantal /r/ sound in German was rosa. Speakers were instructed to alter only the /r/ sound (to the American retroflex /r/), and to keep the /o/ sound that followed as much like the monophthongized Standard German variant as possible. This was a difficult task, but speakers repeated the words until they produced a token deemed acceptable to both themselves and the investigator.

The audio files used in the test were selected by the investigator by listening to each speaker’s productions and determining the best productions for each word. All speakers were represented in the final task producing both the Standard German and English-accented pronunciations, in order to prevent students from using the speaker’s voice as a cue to identify the ‘Germanness’ or ‘Americanness’ of the sounds. Each speaker was represented in the final test between five and seven times, demonstrating between two and five instances of either Standard German or English-accented pronunciation. How many times each person was represented producing a particular kind of accent was not influenced by their native language, but rather by the quality of tokens that they produced, as well as which speakers produced the most appropriate tokens for each individual word.
3.2.2 Production Test – Word Reading

The word-reading task consisted of 63 test words and 12 fillers for a total of 75 real German words from the first two chapters of the *Vorsprung* textbook. Most words fell under the semantic categories of “Family” and “Objects in the Room”. All words chosen were fairly frequent in the German language overall, and were used in a variety of classroom and textbook activities.

When choosing test stimuli for the production tasks, a number of precautions were taken. Firstly, no words that had overlapping segments or stress patterns with another condition were included. For instance, the word *rot* could be a contender for either the [ʁ] condition or the [o:] condition, and was therefore discarded. This was done because of the type of analysis being undertaken. Native speaker raters were used to rate comprehensibility and accentedness. If a test stimulus contained more than one test segment or stress pattern, then it would be impossible to tell what the native speakers were basing their judgments on, as well as what was influencing any improvement in the ratings over time. There is only one exception to this rule: the word *Moment*, which contains both an [o] and a special stress pattern. This word was used as a stimulus for the “Cognates with 2\textsuperscript{nd} syllable stress” condition under lexical stress. No other words from the first two chapters were possible, therefore this word was included in order to balance out the number of stimuli for each lexical stress sub-condition.

The second precaution used when selecting test stimuli was avoiding complicated consonant clusters and difficult-to-pronounce segments. For instance, [l] and the four front-rounded vowels in German were avoided as much as possible. These segments are typically very difficult for American English speakers to pronounce when speaking German, particularly at the beginning levels. As these sounds are not the focus of the study at hand, they were avoided to
the greatest extent possible when choosing test stimuli in order to avoid confounding effects on
the native speaker ratings.

Lastly, cognates were avoided in the test stimuli whenever possible (where the condition
did not require cognate status). Research has shown that cognates are processed differently than
non-cognates, which affects both their perception and production (Costa, Caramazza, &
Sebastián-Gallés, 2000; Costa, Santesteban, & Caño, 2005), therefore this was a confounding
factor that was unwanted in the current study. This rule proved very difficult to follow, however,
due to the very high number of cognates between English and German, especially in basic first-
semester vocabulary topics such as family, colors, etc. The cognates that were included due to
the necessity of using words from the first two chapters of Vorsprung were mainly historical
cognates. These cognates have significant phonological differences because of divergent sound
changes, so that their phonological similarity may not be transparent to someone just beginning
to study the German language. An example is Schwester-sister, two words which have diverged
so far from one another that they seem much less similar to English speakers than Vater-father.

A small number of fillers was included for the following reasons. Firstly, there are six
segments and four stress patterns being tested, providing a large variety of sounds and stress
patterns in the pre- and posttest tasks. Secondly, there were 63 test words that had to be tested in
both a word list and paragraph context. As all recordings for the semester were assigned as
homework assignments and incorporated into the grade for the course, it was desirable to avoid
an overly large and cumbersome assignment that may frustrate students and lower morale.
Lastly, task design in many previous research studies did not include fillers or distracter
sentences of any kind (e.g., Botero, 2011; Derwing et al., 1998; Hardison, 2004; Lord, 2005).
The combination of these three motives led to the inclusion of a small number of fillers. Twelve
fillers were included in this experiment. Four words were monosyllabic, four disyllabic, and four trisyllabic. All fillers were noncognates that did not contain any target segments or unusual stress patterns.

3.2.3 Production Test - Paragraphs
Stimuli in the paragraph-reading part of the production test consisted of six different paragraphs, each telling a short story. Each of the test stimuli from the word reading task was included in the paragraphs. Grammatical structures in the paragraphs were appropriately complex for first-semester students. Topics included student life and school studies, family, things in students’ rooms or apartments, and taking a trip. These topics were chosen in order to incorporate all test stimuli in an appropriate context, and aligned well with the topics that occur in the first two chapters of Vorsprung.

The paragraphs varied in number of sentences from 8 to 14 sentences, with an average length of 10.33 sentences per paragraph. A total of 348 words were used in the paragraphs. The shortest sentence was two words long and the longest was twelve words long, with an average length of 5.6 words per sentences.

Each paragraph was comprised of simple and compound sentences. The majority of sentences were either one- or two-clause sentences connected with a coordinating conjunction. Subordinate conjunctions and clauses were not included in the paragraphs because they require a change in word order in German. 74.2% of the sentences in paragraph task contained only one clause (n = 46). 12.9% (n = 8) of the sentences contained two clauses connected with a coordinating conjunction. Four (6.45%) of the sentences consisted of two clauses connected in

---

2 Due to investigator error, one of the target stimuli included in the word-reading task was left out of the paragraph stimuli. The one word that was left out was lesen, which fell under the [e:] condition.
some other way, such as a comma or a question-word, such as wo (‘where’). The last four sentences in the paragraph task were questions (6.45%).

The majority of the sentences in each paragraph, 52 out of a total of 62, contained at least one test stimulus, with a maximum of four per sentence. 54.8% (n = 34) of the sentences contained only one test stimulus, while 25.8% (n = 16) contained two. Only two sentences contained three or four test words, at 1.6% each for 3-test-word and 4-test-word sentences. Ten of the sentences (16.2%) in the paragraphs contained no test words at all, and were just fillers used to further the context.

3.2.4 Production Test – Free Speech

The free speech section of the pretest consisted of five question prompts, designed to elicit the production of each of the test segments. The questions were of an appropriate level for students in an elementary basic language course. Example questions are: Wann hast du Geburtstag? (When is your birthday?) and Was ist dein Hauptfach? (What is your major?).

3.2.5 Posttest

The posttest was identical in content to the pretest. Task order remained the same, but all stimuli were presented in a different order within each separate task block.

3.2.6 Pronunciation Attitude Inventory

A modified version of Elliott’s (1995) Pronunciation Attitude Inventory (PAI) was administered to each participant. The PAI is designed to elicit information about each student’s personal attitudes toward and opinions about the importance of pronunciation and their ability to acquire it.
The original PAI consisted of 12 statements on a 5-point Likert scale, with a rating of 5 being “Always or almost always true of me / Totally Agree” and a rating of 1 being “Never or almost never true of me / Totally Disagree”. Nine of the statements were positively worded, and three were negatively worded. Sample statements include “Acquiring proper pronunciation in a foreign language is important to me.” as a positively worded example, and “I will never be able to speak a foreign language with a good accent.” as a negatively worded example.

Elliott’s (1995) original PAI was altered for this study as follows. Firstly, all of the statements that included the word “Spanish” were changed to “a foreign language”. This was done not only to make the survey applicable to German learners, but to measure participants’ attitudes toward pronunciation in foreign languages in general. This modification had been made in previous studies on pronunciation training that utilized the PAI as well (Ducate & Lomicka, 2009; Lord, 2008). Secondly, a thirteenth question adapted from Botero (2011) was included: “I do not practice a native-like accent in class because of how other students in class would perceive it.” Thirdly, the investigator also included one multiple choice question and one short-answer question at the end of the inventory. These questions were included in order to discover participant’s opinions about the goal of communication in a foreign language and determine their reasons for their interest in learning German.

3.2.7 Language Background Questionnaire

The Language Background Questionnaire was a three-page document distributed to each participant before training began. Questions spanned topics such as: age, language learning history, languages spoken, study abroad experience and amount of time spent practicing German outside of class. In order to provide a measure of language proficiency outside of course level, participants also used this questionnaire to provide self-ratings of their proficiency in reading,
writing, listening and speaking in both English and German. Ratings were measured on a scale from 1 to 10, with 10 being the best possible rating of their skills.

3.3 Experimental Procedure

The pretest and posttest were assigned as homework assignments for all students enrolled in all sections and proficiency levels. Every student was expected to complete all of the pronunciation assignments, and it was written into their syllabus as part of their course grade. Only those students who consented to have their recordings analyzed were included in the final analysis.

The pretest was assigned in the fourth week of the semester, one week before pronunciation training was to begin in the experimental sections. The investigator made individual visits to each classroom to obtain consent from students, assign the pretest, and administer the Pronunciation Attitude Inventory and the Language Background questionnaires. After the completion of these events, the pretest assignment was explained to the students. As this was a homework assignment to be completed outside of class, it was made very clear to the students exactly what the requirements for completion were, as well as the investigator’s expectations for good quality recordings. Students were instructed to produce a good quality recording by following these rules:

1. Make sure that you are recording in a quiet area (no televisions, loud roommates, friends’ conversations, etc.)

2. Please do not type on the keyboard at the same time that you are speaking. If you need to look up a word on an online dictionary, please type while you are not speaking.

3. Try to speak as clearly as possible.
4. Minimize background noise and static from your computer’s microphone, so that your voice is clear and crisp.

5. Please save your file in a .wav or .aiff file format. No other file formats will be accepted for a grade.

The investigator also instructed students on how to download and use Audacity (ed. 1.2.6: http://audacity.sourceforge.net), an easy-to-use recording software available for free online. Lastly, the instructor also shared with students various options for the completion of their recordings if their computer did not have a microphone or the quality of their microphone was bad.

For the pretest assignment, all necessary stimuli and instructions were inserted into two Powerpoint files that were uploaded to the university’s online course management system. One Powerpoint file contained all embedded sound files for the perception test and the other presented all task stimuli for the production test. The Powerpoint also included all of the instructions on how to complete the assignment that were presented in class, for reinforcement and as a reminder. To complete the assignment, students began with the perception test and then moved directly to the production test. Students completed the assignment on their own time and uploaded their files to a dropbox in their online course management system. Students were given one calendar week to complete the assignment and were required to hand in all components before training began.

3.3.1 Perception test

The format of the perception test is as follows. Students were presented with a Powerpoint slide with a German word written at the top and with two audio files beneath it. The audio files were represented with the symbol of a speaker and assigned either the number 1 or
the number 2. Students clicked on the speaker symbols to play each audio file. What they heard on each slide was the designated German word pronounced in two different ways—either with a native-like, Standard German pronunciation, or with a nonnative-like, American English-accented pronunciation. Students provided binary answers to the question “Which one sounds more like a Standard German accent?” by marking either a number 1 or 2 after the corresponding word on a worksheet. Students were permitted to listen to each audio file as many times as they needed to before making a decision.

The audio files were presented to the students in such a way that half (n=14) of the pairs of audio files contained the Standard German pronunciation presented as file number 1, and half (n=14) of the pairs were presented with the American English-accented pronunciation as file number 1. This was done in order to prevent students from only guessing one number consistently, without listening carefully. The stimuli were presented in a pseudo-randomized order, with up to five of the same correct answer appearing in a row. The presentation order was the same for each student, as these were embedded into a Powerpoint file and presented as a homework assignment.

3.3.2 Production Test

For the production part of the assignment, students were directed to the second Powerpoint file. After a few slides reviewing the instructions and rules for submission, the production test began. All three production tasks were included in the same file, with all stimuli and individual instructions appearing before each section. Students were instructed to begin their recordings before each task and save them individually after each task.

The first section that the students encountered was a word reading list, with all word stimuli presented in a pseudo-random order. Each slide contained a single German word, large
and centered within the slide. Students read each word on the slide, and then hit a key on the keyboard to progress to the next slide. Participants repeated this procedure until they had read all 75 words.

The second section for the production test was paragraph reading. Each slide for this part of the assignment had a heading with a number between 1 and 6, and a paragraph for the students to read aloud and record. Students were instructed and encouraged to read the paragraph to themselves silently before beginning to speak. The paragraphs were presented in a completely random order that was the same for each participant.

The third section for the production test was free speech. Each slide had a heading that identified the question (“Question #n), with bullet pointed question prompts underneath. Students were instructed to answer the questions as best as they could. There were no limits on how much or how little they should say; they were told only to answer the question fully and in complete sentences. This allowed for differences in proficiency between the two course levels, and attempted to relieve some of the stress on the first-semester German students who did not have the same speaking ability as the fourth-semester students did.

It was in this manner that students progressed through the production test PowerPoint, moving from the word level to the free speech level, while making recordings of themselves speaking. On the very last slide, they were prompted to upload their three recordings to the dropbox online and were thanked for their participation.

After completion of the pretest, all participants in all groups were offered written feedback on their pronunciation from the investigator. This feedback mimicked the type that is usually given in four-skills German courses, which typically provides statements from the teacher about the sounds of German and how they relate to the orthographic norms of the
language. Such feedback consists mostly of corrections for individual segments that pose problems for the student, or suggestions for improvement of pronunciation of individual lexical items. The feedback given for the pretest offered praise for things done well, pointed out pronunciation errors that need future practice, and gave advice on how to make those sounds.

### 3.3.3 Posttest

The posttest contained all of the same stimuli as the pretest, and participants proceeded through each task in the same order. They began with the listening test and ended with the free speech part of the production test. Within each block, however, all stimuli were presented in a different pseudo-randomized order than they had been in the pretest.

The exact same audio files were used for the perception posttest. Half (14) of all word pairs consisted of the Standard German pronunciation presented first, and half of the stimuli were presented with the English-accented pronunciation first, just as in the pretest. In order to prevent students from picking up on a pattern and simply providing either the exact same or the exact opposite answers from their pretests, the answers to half of the stimuli were switched. Half of the stimuli were presented with the same order and identification of the sound files as before. The other half of the stimuli were switched, so that if the version with the Standard German pronunciation was presented as audio file #1 in the pretest, it was presented as audio file #2 in the posttest. Of those that were switched from the pretest order, half (7) were presented with the Standard German accent first, while the other half (7) were presented with the English-accented pronunciation first. For the posttest, listening test stimuli were again presented in a pseudo-randomized order that was the same for each participant. In this version, up to four word pairs that had the same answer were presented in a row.
For the production test, the order of the stimuli was also pseudo-randomly organized within each block and presented in a different order than in the pretest. This new order was the same for all participants.

3.4 Pronunciation Training

The method of pronunciation training used in this study consisted of two parts: in-class training and partner assignments with a peer feedback component.

In-class training was provided for students in the experimental groups only. The purpose of this training was manifold—1) raise student awareness of the differences in pronunciation between German and American English sounds and stress patterns, 2) train the perceptual system to begin to correctly perceive new and slightly different German segments, 3) demonstrate to students how to produce these segments in their own speech, and 4) provide pronunciation practice in the form of communicative activities, highlighted by occasional drills for skills practice.

The partner assignments were completed by participants in all sections, both experimental and control. One partner assignment was assigned after each module, for a total of four at a rate of one due every other week. The partner assignments were paired with peer feedback, and the goal behind the partner assignments was to raise students’ awareness of their own pronunciation by focusing on that of their peers.

3.4.1 In-Class Training

Students in the experimental group of each proficiency level received in-class training on German pronunciation as a supplement to their normal classroom work. Students received
training on the six target segments and lexical stress, and all training activities were led by the students’ normal instructor.

Instruction was broken up into four separate modules, each of which focused on one or two of the targeted features of pronunciation: “Lexical Stress”, “Ich/Ach Sounds”, “/r/ Sounds”, and “/e/ and /o/”. Each module was taught over a two-week time span during normal class time. Within the module, 10 minutes of class time each day was devoted solely to pronunciation training, for a total of 40 minutes per week, or 80 minutes (1.33 hours) per module. Half of the test stimuli from the pre- and posttests were specifically trained during the intervention; the other half were not trained at all. This was done in order to test the possibility for generalization of the training to words that had not been trained. Each of the words that were chosen for use during in-class training was used in a minimum of two different activities throughout the module. All other words used in the training activities were not among those used as stimuli in the pre- or posttests. The thematic topics of the words and activities chosen for use in the pronunciation lessons closely matched that which was currently being taught in the first-semester class. Topics ranged from family to food, objects in the classroom to personality traits.

The training modules were identical across proficiency levels; all experimental group students received training using identical activities, sentences and vocabulary to avoid confounding factors such as richer vocabulary or syntactic structures for students in the more advanced levels. Training in both perception and production of the feature in focus was provided. The main goal behind training was to heighten attention and potential noticing of the relevant forms. Enhanced input in communicative contexts was used to draw students’ attention to the relevant forms and how they differ between the two languages. A focus was also put on
how these pronunciation patterns and the differences between them affect comprehensibility, intelligibility, and meaning.

Training followed the optimal progression as outlined by Chun (2002): students began work with discriminatory and active listening exercises, participated in a short mimicking stage when necessary, progressed to more active controlled speech exercises, and ended with practicing what they had learned in free speech. Listening activities took place over the first three or four days of training. Listening activities were designed to develop students’ L2 perception and listening skills, and therefore consisted of a mixture of discrimination and active listening exercises. During this phase of training, students worked with speech samples ranging from individual words to longer paragraphs and dialogues. These speech samples were designed to be as authentic as possible, and at least one listening exercise per module featured recordings of native German speakers, so that students received input from a variety of speakers. The students’ first production practice typically began on the fourth or fifth day of training. At least one day consisted of traditional pronunciation training methods, such as drills and mimicking, that provided repetition of the sounds in isolation and in small words. Although drills are often viewed as inappropriate for a communicative classroom, the repetition that these traditional methods provide are vital to phonological acquisition and, when framed in the right way, can be a welcome complement to communicative activities (Isaacs, 2009). It is important that students become familiar with the mechanics necessary to produce unfamiliar segments or prosodic patterns accurately. The last days of training were reserved for controlled and free speech exercises. Controlled speech exercises helped to stabilize the speech patterns that had been practiced during drills. In these exercises, students began with simple utterances of a few words and gradually worked up to practicing their pronunciation in complex sentences and short texts.
Free speech exercises were modeled as closely as possible after natural language situations in which the L2 might be used. Such exercises provide communicative language practice mixed with the opportunity to use the new pronunciation patterns in a “real life” situation and are necessary to solidify the connection between the pronunciation target and the communication of meaning in the students’ minds.

The following example activities demonstrate which types of training were used in the classroom. A common activity involved the use of pairs of German words. With the information that participants perceived during aural presentation of the words, they had to indicate if the two words they heard was the same word pronounced twice and or whether each word they heard was different. This is a standard discrimination task and is effective for developing the perception of L2 sounds. Another activity involved the use of visual and kinesthetic learning via props. Students used rubber bands to symbolize the longer vowels in stressed syllables, and read a list of words while stretching the rubber band on these syllables. The word list for this activity contained words with contrastive lexical stress in English and German, which each student read to his or her partner while stretching the rubber band. The partner then had to guess whether they had heard the English or German style stress pattern, using both the visual and the auditory information they received. A game of “Telephone” was employed to practice producing contrasts between German and English sounds and to communicate the importance of these segments for meaning. A final example involves a combination of grammar review and pronunciation practice. Students conjugated verbs and produced their own sentences using these verbs in order to focus on the alternation between vocalic and fricative /r/ in German. Each of these example activities was used during the modules to promote either perception or production of L2 sounds, or in some cases, both.
All lesson plans were designed by the investigator and shared with the two instructors a week before lessons for the designated module were scheduled to begin. The instructors also held weekly meetings with the investigator. At these meetings, expectations for the units were shared, issues were discussed and solutions found, questions were asked, and lesson plans for the upcoming days and weeks were reviewed. In addition, the investigator made at least one in-class observation for each instructor per module, and gave constructive feedback to the instructor. Through these various methods, it was ensured that the lesson plans were executed according to expectations and that the two instructors taught their lessons as similarly as possible. In-class observations also gave rise to investigator contact with the students, in the event that any issues arose from the students’ perspective.

3.4.2 Partner Assignments

Before beginning the first pronunciation training module, each student chose their own “pronunciation partner”. The pronunciation partner was someone who each participant would work closely with on improving their pronunciation over the course of the semester. Students were allowed to choose their own partners, because a high level of comfort with that person was necessary for proper support, feedback and constructive criticism to occur.

At the conclusion of each module, all participants were assigned a partner recording as part of their coursework. The partner assignments consisted of three different tasks, based on those used in Ramírez Verdugo (2006) and designed to focus participant attention on the pronunciation of the target feature previously studied. For the first task, each student read a short list of five to eight words. This word list included some of the target words that were chosen for in-class training and had been previously included on the pretest, as well as various other words that had been used in the in-class activities. Each partner had a different list of
words, but their lists were always equal in length and contained equal numbers of targeted test stimuli. On the second task, each partner asked a few discussion questions to their partner, who then answered freely and to the best of their ability. The discussion questions were designed to elicit productions of the feature in focus, for example, students answered questions about things they like to eat and cook for the module on [ç] and [x], which would give them the opportunity to pronounce these segments in their answers. The students then switched roles, so that both partners were able to answer the same set of discussion questions.

The third part of the partner assignments was the most crucial of the design. Each partner listened to the recording that they and their partner had just produced and evaluated their partner’s pronunciation. Each student was required to provide at least one positive comment (“He/She did this well”) and two points of constructive criticism (“This still needs work”) for their partner. Students in the experimental groups were encouraged to use what they had learned in class to provide this feedback, while students in the control groups were encouraged to use what they already knew of German pronunciation. The instructions for each partner assignment focused students’ attention on each of the target sounds. They completed this task for each of the four partner assignments.

This method was chosen because while Dlaska and Krekeler (2008) have shown that self-assessment of pronunciation is often inaccurate and ineffective, the success of peer assessment in the development of oral proficiency and pronunciation has been measured in multiple studies (Counselman, 2010; Lord, 2008; Miller & Ng, 1996). Peer feedback can be a vital tool in the development of learners’ pronunciation and was therefore utilized in this study as well. At the same time, Ducate & Lomicka (2009) show that peer feedback may not be effective enough on its own to provide large gains in pronunciation without the help of in-class training. Thus, even
students in the control groups completed this feedback assignment, to keep the instructional
design between the two groups as identical as possible, and the concern that students in the
control group would keep pace with those in the experimental group just by completing this
feedback assignment was not large.

3.5 Analysis of Learner Data

Following standard procedure in pronunciation training studies, analysis of the
participants’ productions was conducted by collecting ratings from native speaker raters. These
raters listened to the audio recordings and provided ratings of accentedness and
comprehensibility on a Likert-scale from 1-7. Results from this analysis measures the effects of
training on L2 learners’ pronunciation, and are reported in the Chapter Four.

3.5.1 Native Speaker Raters

Five native German speakers currently immersed in their L1 environment participated as
raters for the analysis of the participants’ audio files. Four were female, and their average age
was 25.2 years. None had any reported hearing difficulties. All reported German as their
dominant language, and although they originated from different dialect regions in Germany, all
were currently living and studying in the same German university town. All five raters had
begun studying English as a second language at an average age of 9, and they evaluated their
English knowledge at 5.2 out of 10. None had spent any time in an English-speaking country.
Raters also had learned a variety of third and fourth languages. French was the most common,
with 4 out of 5 raters learning French since an average age of 13.25. The raters also had
experience with languages such as Finnish, Arabic, Spanish and Swedish, although none of these
had been learned until college age or studied for an extensive period of time. Three of the raters
were considered naïve listeners, who had never taught German and did not report contact with native English speakers or experience listening to English-accented German.

Two of the raters presented a special case, as they had previously taught German language courses at the B1 and B2 levels. They had also taken a class on German phonology. This could be problematic, as research has demonstrated that teachers are more lenient in their ratings than naïve listeners (Cunningham-Andersson, 1997; Hirschfeld, 1994). Yet they did not differ from the naïve listeners in that they described their knowledge about the sounds and pronunciation of English as “a little”, and rated their level of exposure to English-accented German as “seldom” or “sometimes”. This was in line with the other three raters, therefore these two raters were considered to be no different than the naïve listeners for the purposes of this experiment.

3.5.2 Rating Materials

Each participant’s six audio files from both pre- and posttests were spliced into smaller chunks using Audacity 2.0. (Audacity Team, 2012). All target words and sentences were spliced out of the original audio file and coded appropriately. For the target words, each individual word was saved as its own separate file. For the paragraphs, each sentence containing one of the target words was separated from the original file.

The audio files were cleaned and adjusted, to provide raters with a high quality audio file without distractions. The noise removal filter on Audacity was used to remove any white noise or background noise that may have been present on the original file. All false starts, stutters, and unrelated non-speech sounds (clearing of the throat, coughs) were removed from the final files. As long pauses can alter native speaker raters’ perceptions of accent and negatively influence ratings (Hirschfeld, 1994), pause lengths were adjusted. This was done to ensure that the raters.
were basing their judgments on the actual acoustic features of the speech, and not the speaking rate or pause rate. Pauses longer than 500ms in places where a sentence did not contain a clause break or comma were shortened to 100ms. Pauses longer than 500ms in places where the sentence did contain a clause break or comma were shortened to 300ms. For each file, a short pause was reserved before and after the participant’s speech began, in order to retain all speech characteristics.

Per test, there were 63 individual word files, 57 paragraph sentences with target words, and 5 answers to free speech questions per participant. In sum, the adjusted pre- and posttest data consisted of 250 files per participant, for a total of 11,000. Due to the large amount of files, only one quarter of the cut audio files were used for the rating experiment. A pseudorandomized sample of participants’ pre- and posttest productions was created by randomly choosing from individual participants’ recordings such that all target words, sentences, and free speech prompts were represented as evenly as possible across participants and sections. Per person, a total of one word and one paragraph sentence per stress condition, two words and two paragraph sentences per segmental condition, and two answers to the free speech prompts were used in the final experiment. No target words were repeated across the individual word and paragraph sentence conditions for each participant’s sample. For each participant, each pretest file was matched with its equivalent posttest file to ensure accurate comparison between the two tests. To fill out the sample, files from the participants were mixed with test recordings from two advanced L2 German speakers and four native German speakers. These were added to serve as a check on the native speaker raters, both to see if they used the entire rating scale, as well as whether or not they could identify a native speaker. All audio files were randomly assigned to one of nine
separate test blocks. Each block contained between 374 and 378 audio files to be rated, with a mean of 376.

3.5.3 Rating Procedure

Native speaker raters were tested in a quiet laboratory at the Institut für Germanistische Sprachwissenschaft in Marburg, Germany. Each rater came individually to the institute on three separate days over the course of one week to complete the experiment. Three of the nine total blocks were completed each day, and each block took participants between 30-50 minutes to complete. The order in which the blocks were completed varied for each participant and was counterbalanced so that the same block was not always presented in the same position relative to other blocks (for instance, the third block was not always placed last in the testing order, when participants were most tired). As it is important to battle fatigue when collecting native speaker ratings, raters were always given the option of taking a break between blocks, and were provided with water and various snacks of their choice. The native speaker raters were also paid for their participation.

Before raters began with the experimental portion of testing, they were first required to give consent and were given a short introduction to the experiment. They were advised to base their ratings on their best holistic impression, and to make use of the entire scale. They were told not to attend too much to grammatical errors, but rather to focus on the pronunciation of the words and sentences. All conversation between the investigator and raters and all instructions on the computer were in German.

Raters wore Audio-Technica ATH-M30 closed-back headphones for the duration of the experiment and sat in a comfortable chair in front of a laptop computer. All speech samples were played for the native speaker raters in a random presentation via Eprime software.
The experiment was self-paced and proceeded as follows. The raters read through the instructions on a computer screen, completed one sample rating, and then began with the first test block. Raters were first presented with a fixation point and pressed the space bar when they were ready to continue. This triggered presentation of the sound file over the headphones. Raters heard each audio file only one time, and then provided two ratings directly after hearing it. They could not play the sound again, or go backwards in the program.

After listening to the audio file, the native speakers then provided two different ratings—one for comprehensibility and one for accentedness—on a 7-point Likert scale. A 7-point scale was chosen because it is sensitive enough for raters to define their perceptions accurately while maintaining optimal interrater reliability (Cicchetti et al., 1985; Colman et al., 1997); a 7-point scale was also successfully implemented in a pronunciation training study by Hardison (2004). The rating scale appeared on each of the two rating screens, where 1 was the best possible rating (“perfectly easy to understand” / “definitely a native speaker”), and a rating of 7 was the worst possible option (“impossible to understand” / “definitely not a native speaker”). Questions to aid in the raters’ decisions were provided on the screen, such as Wie verständlich ist diese Person? (How comprehensible is this person?) above the comprehensibility scale and Wie ‘deutsch’ klingt diese Person? (How ‘German’ does this person sound?) above the accentedness rating scale. This question was the only difference between the otherwise identical rating screens. Ratings of comprehensibility were given first, followed by ratings of accentedness. Raters completed this same sequence of events until they reached the end of the block.

Because most raters found the first block completed to be the most strenuous, a short break was required after its completion. During this break, raters filled out a language
background questionnaire. This questionnaire asked them primarily about their language skills, and their exposure to nonnative German in general, and English-accented German in particular. After completion of the questionnaire, raters resumed testing. Each block then continued in the exact same format as previously described, until all blocks for that day were completed.

The first six blocks, conducted on the first two days of testing, have a total of five raters apiece. The last six blocks were conducted on the last day of testing and have a total of four raters apiece.³

³ At the conclusion of the second day (sixth block), one of the participants expressed that she no longer wished to continue with the study, citing lack of time and fatigue with the experiment. She gave consent for the data already collected to be used, but this data had to be excluded from the analysis for other reasons (see Section 4.2).
Chapter 4: Results

Results from the perception test, the production test and the motivational measure are explained in the following chapter. In this chapter, I also describe the various statistical analyses that were conducted. By presenting these results, I shed further light on the research questions that were proposed in Chapter Two. Results from each of the tasks and analyses are presented separately: the perception test is in section 4.1, native speaker ratings from the production test are in section 4.2, and the motivational measure is in section 4.3. This chapter ends with discussions of generalization of training in section 4.4 and the relationships between tasks in 4.5.

4.1 Perception Test

The perception test that was part of the pre- and posttest measured participants’ ability to determine the differences between Standard German and English-accented pronunciation of real German words. Participants listened to two different pronunciations of the same word and had to determine in a binary choice which word was produced with a more typical Standard German pronunciation. There were four tokens for each of the seven categories (six segmental and one lexical stress), and each token was assigned either a score of 1 for a correct answer or 0 for an incorrect answer. Therefore, the highest possible score on the perception test was 28 points.

In a preliminary analysis, it was discovered that an exceptionally high percentages of participants in all four groups could distinguish the differences between the German fricative /ʁ/ [ʁ] and the American English retroflex /ɹ/ [ɾ] at pretest. 80% of all participants (n = 36) scored four points out of a possible four on the pretest—95% of participants in the fourth-semester groups (n = 20) and 67% (n = 16) of the participants in the first-semester groups. Due to such high pretest accuracy on this item, the results from the fricative /ʁ/ category were excluded from further analysis.
After discarding results from the fricative /r/ category, the highest possible score on the adjusted perception test was 24 points. Data from any participant who scored at ceiling at pretest, characterized as a score of 22 or above, were removed and not used for analysis. In addition, data from any participant who scored more than 2.5 standard deviations outside of the mean posttest score were eliminated, resulting in the loss of data from one participant. After this cleaning process, data from 28 participants remained. Eleven were in the first-semester experimental group, and seven in the first-semester control group. Among fourth-semester participants used for analysis on this task, four were in the experimental group and six in the control.

The first analysis undertaken calculated the average total accuracy on the perception test, as well as the amount of change. This was completed for each participant group. Results show that all groups improved their accuracy on the perception test from pretest to posttest and that the first-semester experimental group enjoyed the greatest gains. The first-semester experimental group exhibits the largest change in average total score from Time 1 to Time 2, with an increase of 10%. Participants in this group improved 2.4 points from 17.9 at pretest to 20.3 at posttest. In terms of average scores, this group improved more than twice as much as the group with the next highest amount of improvement. The fourth-semester control group had the second-highest amount of change, improving one point from 20.2 points at pretest to 21.2 at posttest. The fourth-semester experimental group demonstrated a similar pattern. They scored an average score of 18.8 on the pretest, and increased this score to 19.6 on the posttest, for an overall increase of 0.8 points over time.\(^4\) Out of all four groups, the first-semester control group showed

\(^4\) The low accuracy scores for the fourth-semester experimental group are in this case due to the cleaning of the data. After data had been cleaned of any participants who scored higher than 22/24 points, the average of the fourth-semester experimental group dropped to an average score lower than that of the fourth-semester control
the least amount of improvement from pretest to posttest. Their change in average total score was only 0.3 points, from an average of 18.3 points on the pretest to 18.6 points on the posttest. These results are presented graphically in Figure 1 below. Useful to note here is that the first-semester experimental group shows the greatest increase in accuracy, as evidenced by their very steep slope.

![Figure 1: Average Overall Perception Test Scores by Group](image)

A second analysis determined how many participants’ perception test scores showed improvement, a decline, or no change from pretest to posttest. In the first-semester experimental group, 90.9% of participants (n = 10) improved their perception accuracy from pretest to posttest. The other 9.1% of participants (n = 1) achieved the same level of accuracy both times and neither improved nor declined. The first-semester control group shows less systematic results than the experimental group did. In this group, 42.85% of participants (n = 3) showed improvement, but the same number of participants (n = 3, 42.85%) also scored worse on the posttest than they did on the pretest. One participant in this group (14.3%) did not show any change from Time 1 to Time 2. Thus, a vast majority of participants in the first-semester who

---

group. Before the cleaning of the data, the fourth-semester experimental group had accuracy scores of 20.7 at pre-test and 21.2 at posttest, making their scores closer to those of the other fourth-semester group.
received training were able to improve their perception skills over time. The participants who did not receive training, on the other hand, did not show this pattern. Their results were much more varied and do not show a clear pattern of improvement. Figures 2 and 3 summarize these results from the two first-semester groups.

![Figure 2: Change in Perception Scores for First-Semester Experimental Group](image1)

![Figure 3: Change in Perception Scores for First-Semester Control Group](image2)

Similar results can be found in the two fourth-semester groups. After results from those participants at ceiling at pre-test were cleaned from the data, half of the participants (n = 2) in the experimental group showed an increase in their perception scores from pretest to posttest. The other half (n = 2) demonstrated a decline in accuracy at posttest. As with the first-semester groups, the fourth-semester control group shows more chaotic results than the experimental group did. In the fourth-semester control group, even numbers of participants showed increased, decreased, and equal scores from Time 1 to Time 2. One third (33.3%, n=2) of participants improved their perception score, 33.3% (n = 2) received a worse score, and 33.3% (n = 2)
received the same score. So while half of the participants in the fourth-semester experimental group showed improvement in their perception skills, no similar majority was found for those participants in the fourth-semester control group. Results from the two fourth-semester groups are presented in Figures 4 and 5.

![Pie chart showing perception score changes](image1)

*Figure 4: Change in Perception Scores for Fourth-Semester Experimental Group*

![Pie chart showing perception score changes](image2)

*Figure 5: Change in Perception Scores for Fourth-Semester Control Group*

Of the 28 participants in all groups, seven attained a lower score on the perception posttest than on the pretest. Four of the participants who showed a decline in accuracy were in their fourth semester, and three were in their first. Only two participants who declined (28.6%) were in groups that had received training, while five (71.4%) were in control groups and had not received any training. Notably, not one of the participants in the first-semester experimental group declined from pretest to posttest.
Four participants from all groups received a posttest score identical to their pretest score. Half were in their first-semester and half in their fourth. 75% of these four participants (n = 3) were in groups that did not receive training. Only one participant of the four whose scores remained the same received training, and this participant was in the first-semester experimental group.

The vast majority (n = 17, 60.7%) of all participants in this study showed improved perception test scores from pretest to posttest. Of these, 76.5% (n = 13) were in their first semester of learning German, and four (23.5%) were in their fourth-semester. 70.6% of the participants (n = 12) who improved over time were from a group that received training. The majority of the participants who showed improvement were in the first-semester experimental group (n = 10, 58.8%), while the remaining three were in the fourth-semester experimental group. Of the nine participants who improved the most and managed to score 3 or more points higher on the posttest than on the pretest, 66.7% were in trained groups, while 33.3% were in untrained groups. Figure 6 summarizes the results of the previous three paragraphs in visual format.

Figure 6: Percentage of Improved, Declined and Same Scores in Trained vs. Untrained Groups
Results as measured by the perception test suggest that training seems to have had an effect on the improvement of perceptual skills. Among groups in which the participants received training, a majority of participants showed improved scores from pretest to posttest, while the same cannot be said about the two control groups in which participants did not receive any training. A majority of participants who increased their accuracy over time was in groups who had received training. Of particular interest is the first-semester experimental group, in which 90.9% of participants increased their accuracy, and none showed a decline. On the other hand, the majority of participants who decreased in their accuracy from pretest to posttest (71.4%) were members of groups which did not receive training.

In order to conduct statistical analyses, results were compared within and not across proficiency levels, so that first- and fourth-semester groups were only compared to each other. The first analysis completed with these data was a repeated measures ANOVA. This test was conducted using the mean accuracy scores on the perception test, with Time as the within-subjects factor, Group as a between-subjects factor, and the overall score on the perception test as the dependent variable. Results from the ANOVA show that there was a main effect of Time ($F = 10.450, p = .005$) and a Time*Section interaction ($F = 6.428, p = .022$) for the two first-semester groups. Among fourth-semester groups, the ANOVA shows neither a main effect of Time ($F = 1.036, p = .335$) nor a Time*Section interaction ($F = 0.013, p = .912$). Results from this test clarify that for the first-semester groups, there were significant changes from pretest to posttest in the total test score, and that section seems to play an important role in the accuracy scores, while no such changes or influences were present for the fourth-semester groups.

To clarify the situation further, results from the perception test were also submitted to analysis via simple statistical testing, using nonparametric tests due to the low number of
participants per group. This allows for an investigation of how each of the groups did or did not improve over time.

Wilcoxon Signed Rank tests are a nonparametric test used to compare 2 related samples from the same group. Using these tests, it was discovered that only one group, the first-semester experimental group, showed a significant improvement in perception test accuracy over time \((Z = -2.842, p = .004)\). Participants in this group were able to significantly improve their accuracy on the perception test from pretest to posttest, while participants in the first-semester control group who did not receive training were not able to do so. In addition, neither of the fourth-semester groups showed a significant change in accuracy. Their scores on the posttest were higher, but not significantly different from their scores on the pretest. Results from the Wilcoxon tests are summarized in Table 5.

Table 5

<table>
<thead>
<tr>
<th>Wilcoxon Signed Rank Test Results- Perception Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>First-Semester Experimental</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>First-Semester Control</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Fourth-Semester Experimental</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Fourth-Semester Control</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* = significant at the \(p < .05\) level

As participants in the first-semester experimental group improved their overall accuracy on the perception test from Time 1 to Time 2, the amount of change in accuracy over time was also analyzed. Mann-Whitney tests used to compare two independent groups reveal that participants in the first-semester experimental group showed a significantly greater amount of change than those in the first-semester control group \((Z = -1.966, p = .049)\). Yet when
comparing posttest scores between groups, the perception scores for the experimental and control groups did not show significant differences from one another at posttest (First-Semester: \( Z = -0.550, p = .582 \); Fourth Semester: \( Z = -0.101, p = .920 \)). Thus, the first-semester experimental group did show significant improvement over time where the control group matched for proficiency did not, but this improvement was not great enough that the two groups’ scores were significantly different at posttest. Additionally, no significant improvements were found when data was analyzed according to individual condition, although results from the [e:] condition approached significance for the first-semester experimental group only (\( Z = -1.933, p = .053 \)).

In conclusion, training seems to have had an effect on participants’ perceptual accuracy. The first-semester experimental group had the greatest number of participants who improved their accuracy from pretest to posttest, and did not have any participants who declined in accuracy. This group also showed the greatest average improvement over time, with a statistically significant increase in score of 2.4 points, or 10%. Among the nine participants who improved the most, 70.6% had received in-class training, while among the participants who declined, 71.4% had not received any training. However, a higher proficiency level seems to have had a tempering effect on the amount of influence training was able to impart to participants. Participants in the two fourth-semester groups showed a similar amount of improvement over time, and neither group made statistically significant gains in perception test accuracy.

### 4.2 Production Test

Results from the production test consist of the ratings of comprehensibility and accentedness provided by the native speaker raters. The native speaker ratings of accentedness and comprehensibility yielded, on the whole, a high level of agreement. Of the five native
speaker raters chosen for this study, all but one could accurately identify the native and advanced L2 speakers included in the rating sample and used the entire rating scale in their responses. Data from the one rater who could not recognize and accurately rate the native German speakers was excluded.\(^5\) This rater repeatedly evaluated the native German speakers included in the sample at 3 or 4 out of a scale of 7, with 1 being “definitely a native speaker”. In addition, this rater's ratings also failed to correlate significantly with many of the other raters' average ratings, giving more support to the decision to discard them from the final analysis. The remaining four native speaker raters were able to recognize native and advanced speakers very well, as demonstrated by their appropriate ratings. The native and advanced L2 speakers included in the rating sample received an average rating of 1.16 out of 7 for accentedness, and 1.54 out of 7 for comprehensibility, with 1 being the highest rating possible. After the exclusion of one rater’s data, a Cronbach's alpha calculated with the remaining four raters showed a particularly high interrater reliability of 0.974.

4.2.1 Comprehensibility Data

The following section presents the results for the comprehensibility ratings for participants in all groups, both first- and fourth-semester, in addition to the relevant statistical analyses. Mean ratings of comprehensibility were calculated for the overall test as well as for each individual task and each individual condition. Standard practice before analyzing rating data is to exclude any participants who were at ceiling at pretest. Ceiling for this group of participants was set at 2, due to the use of a 7-point rather than a 9-point scale. No participants received average comprehensibility ratings at ceiling, and no data was excluded. There were no

---

\(^5\) It is standard procedure to exclude native speaker raters that do not recognize the native and advanced speakers provided as a check or who do not make use of the entire rating scale (see: Derwing, Munro & Wiebe, 1998).
significant differences between the overall comprehensibility ratings of the experimental and control groups at pretest when compared using a Mann-Whitney U test (First-Semester: 
\[ Z = -.383, p = .702 \] / Fourth-Semester: \[ Z = -.493, p = .622 \]).

4.2.1.1 First-semester Participants

The first results to be presented are the first-semester participants’ mean comprehensibility ratings for the entire pretest and posttest, with all tasks combined. As a reminder, a rating of 1 is the best possible rating and 7 is the worst possible rating, so a lower score on the posttest illustrates improvement. The average comprehensibility rating for the first-semester experimental group was 4.458 at pretest and this changed to 4.145 at posttest. Therefore, participants in the experimental group showed an overall improvement of 0.313 points in their ratings from pre- to posttest. Participants in the first-semester control group also showed improvement from pretest to posttest. Their overall average comprehensibility rating at pretest was 4.599, and improved 0.227 points to an average rating of 4.372 at posttest.

For further analysis, the results of each of the three individual tasks were separated and averaged for each section. Participants in the first-semester experimental group began with a rating of 4.216 out of 7 on the pretest for the word-reading task only. Their rating improved 0.380 points to a posttest rating of 3.836. Participants in the control group also showed improvement on the word-reading task alone. Their pretest score stood at an average of 4.340, and showed an improvement of 0.265 points to 4.075 at posttest.

Results from the paragraph-reading task show a similar pattern. The first-semester experimental group and control group both received better ratings at posttest. Participants in the experimental group began with a rating of 4.859 on the pretest, and improved 0.295 points. These participants received a rating of 4.564 at posttest. Control group participants began with a
pretest rating of 4.818, and ended with a posttest rating of 4.606. This shows an improvement of 0.212 points.

Although average ratings show similar patterns for the overall test and the word- and paragraph-reading tasks, the results from the free speech task reveal something slightly different. While the first-semester experimental group received better ratings at posttest, the first-semester control group did not. In fact, the control group participants received worse ratings at posttest than they had at pretest. They began with a rating of 4.729, but received an average rating of 4.807 at posttest, for an overall decline of -0.078 points in their average rating. This compares to the experimental group participants, who received an average rating of 5.125 on the free speech task at pretest, and improved 0.197 points to an average of 4.928 at posttest.

To summarize, both first-semester groups showed improvement in their overall comprehensibility ratings from pretest to posttest, as well as for their average ratings on the individual tasks as well. On the word- and paragraph-reading tasks, both groups improved from pretest to posttest. Only in the free speech task does the opposite occur: participants in the experimental group received improved ratings at posttest, while participants in the control group declined in comprehensibility. The amount of change in comprehensibility ratings for the first-semester groups over time is summarized in Figure 7 below. The y-axis values have been reversed in order to more easily visualize the improvement made.
Of interest with these data is the fact that while both groups tended to display improvement at posttest, the experimental group always showed more improvement than the control group did. The experimental group participants performed an average of .140 points better at posttest than the participants in the control group did. In addition, the experimental group always received a better/lower average rating at posttest than the control group did. This holds for both the overall test rating, as well as for all individual task ratings, with the exception of the free speech task.

Figures 8 and 9 below show the overall average score in comprehensibility for each person, grouped by section. Note that lower ratings at posttest indicate improvement, therefore the values on the y-axis have been reversed to better indicate this visually. Twelve participants in the experimental group improved from Time 1 to Time 2, while two showed a decline in overall rating. Paired t-tests were used to compare each individual participants’ pre- and posttest ratings. These tests were employed using unaveraged data (i.e., all ratings from the native
speaker raters for all tokens) to determine how many individuals demonstrated statistically significant improvement. These tests determined that of these twelve participants who improved, seven (63.6%) demonstrated an improvement in comprehensibility that was strong enough to be statistically significant. In the control group, all eleven participants improved from Time 1 to Time 2. Paired t-tests conducted with this group’s pre- and posttest ratings reveal that four out of these eleven participants (36.4%) demonstrated significant improvement in comprehensibility over time. Thus, we can see that while roughly even numbers of participants in the experimental and control groups improved their mean comprehensibility ratings, more participants who had received training exhibited statistically significant improvements.

Figure 8: Average Comprehensibility Ratings, First-Semester Experimental
4.2.1.2 Fourth-semester Participants

Participants in the fourth-semester experimental group received an average comprehensibility rating of 3.622 on the entire pretest as a whole. On the posttest, their overall comprehensibility rating declined 0.047 points to a score of 3.669. While those participants in the experimental group showed a slight decline in overall comprehensibility from pretest to posttest, the control group participants showed an opposite pattern. The average comprehensibility rating for the fourth-semester control group on the pretest was 3.766. Their average rating on the posttest was a 3.697, for a total improvement of 0.069 points.
As before, results from each of the three individual tasks were separated for further analysis. On the word-reading task, participants in the fourth-semester experimental group received a comprehensibility rating of 3.361 at pretest and a rating of 3.455 at posttest. Participants in the fourth-semester control group began with an average rating of 3.486, and ended with an average rating of 3.473. This amounts to a slight decline of 0.094 points between the pretest and posttest for the experimental group, and a slight increase of 0.013 points for the control group on the word-reading task.

Results between fourth-semester groups take very different paths on the paragraph-reading task. The fourth-semester experimental group received a comprehensibility rating of 3.801 on the paragraph-reading task at pretest, and 3.814 at posttest. The amount of change shows a small decline of 0.013 points between pre- and posttest for this group. Paragraph-reading task ratings for the control group stood at 3.979 at pretest and 3.871 at posttest. The control group improved their comprehensibility rating by 0.108 over time on the paragraph-reading task. Compared to the amount of change for the tasks previously described, this is a very large improvement for a fourth-semester group.

Results from the free speech task follow the same pattern that seems to be typical for the fourth-semester groups. The experimental group received a comprehensibility rating of 3.930 at pretest, which declined 0.067 points to a 3.997 at posttest. The control group, on the other hand, received a rating of 3.917 at pretest, which improved 0.024 points to 3.893 at posttest. Once again, the experimental group received worse ratings at posttest, while the control group improved over time. This is the opposite pattern that was found for first-semester participants, where the experimental group improved their comprehensibility in free speech and the control group showed a decline.
In sum, the fourth-semester groups present a much messier picture of the change in comprehensibility ratings over time. In contrast to the first-semester experimental group, which improved both overall and on each of the three task types, the fourth-semester experimental group did not demonstrate any improvement on either the overall test or on any task. This experimental group received worse ratings at posttest than they had at pretest. The average amount of change in comprehensibility ratings for the fourth-semester groups are presented visually below. As before, the y-axis values have been reversed so that improvement is graphed above the x-axis.

![Figure 10: Amount of Change in Average Comprehensibility Ratings for Fourth-Semester Participants](image)

An important thing keep in mind when summarizing the data in this way is that although the experimental group were assigned poorer comprehensibility ratings on every task, the amounts of change are fairly small. The amount of decline stands at an average of 0.055 points. In stark contrast to the experimental group, the fourth-semester control group showed improvement in their comprehensibility ratings in every category: both on the tests as a whole, as well as on the
word-reading, paragraph-reading and free speech tasks. On the paragraph-reading task, this improvement is fairly large, at an average of 0.143 points, which is comparable to the average rate of improvement for students in the first-semester groups.

Similar figures as were presented for first-semester results are shown below. These line graphs present the overall average rating in comprehensibility for each person, grouped by section. Both groups are composed of ten participants. In the experimental group, five participants showed improvement from pre- to posttest, while an equivalent number received a worse rating. All ratings received by each participant were subjected to paired t-tests and it was discovered that of the experimental participants, only one showed a significant improvement in comprehensibility, while two showed a significant decline in comprehensibility. In the control group, eight participants improved over time, while only two declined. Paired t-tests conducted for each participant reveal that this improvement was statistically significant for four of the eight control participants who demonstrated improvement (50%). No participants in the control group demonstrated a significant decline in comprehensibility ratings over time.
Figure 11: Average Comprehensibility Ratings, Fourth-Semester Experimental
Figure 12: Average Comprehensibility Ratings, Fourth-Semester Control

Figure 13 presents the changes in mean comprehensibility ratings for all four groups. As with the figures above, the y-axis values have been reversed to connect a positive slope with improvement. One can easily see the large improvement in comprehensibility made by the two first-semester groups. The fourth-semester groups, on the other hand, either show much less improvement than the first-semester groups do, or they do not improve at all.
In sum, both first-semester groups showed improvement in both their overall comprehensibility ratings and on the word- and paragraph-reading tasks from pretest to posttest. The experimental group always showed more improvement than the control group did. Additionally, the first-semester experimental group improved their comprehensibility in free speech, while the first-semester control group showed a slight decline. The fourth-semester group results were not quite as straightforward. The fourth-semester experimental group received poorer comprehensibility ratings on all tasks as well as overall, and did not show any improvement on any task, while the fourth-semester control participants improved their comprehensibility in every measure. Overall, the first-semester participants seem to have shown more improvement in comprehensibility from pretest to posttest than did students in their fourth-semester.

4.2.1.3 Statistical Analyses & Results
In order to measure whether section had an influence on participants’ change of score from pretest to posttest, participants’ mean comprehensibility ratings were submitted to a
repeated measures ANOVA. As with the perception test analyses, Time was the within-subjects factor, Group was a between-subjects factor, and the overall comprehensibility ratings served as the dependent variable. For the first-semester groups, this ANOVA showed a main effect of Time (F = 38.128, p < .001), but no Time*Section interaction (F = 1.414, p = .246). No main effect of Time (F = 0.013, p = .910) or Time*Section interactions (F = 0.907, p = .353) were found via this analysis for the fourth-semester groups. Results from this analysis demonstrate that for first-semester participants, a significant improvement in comprehensibility from pretest to posttest can be found that extends to participants in both groups. The fourth-semester participants, however, did not demonstrate significant improvement in their comprehensibility.

The same type of analysis was completed using mean comprehensibility ratings separately on each individual task. On the word-reading task, ANOVAs show a main effect of Time for the first-semester groups (F = 23.287, p < .001), but no Time*Section interaction (F = 0.724, p = .404). Among fourth-semester groups, neither a main effect of Time (F = 0.015, p = .699), nor a Time*Section interaction (F = 0.280, p = .603) were found when comparing the comprehensibility ratings on this task. ANOVA results from the paragraph-reading task show a similar development: there is a main effect of Time (F = 20.767, p < 0.001), yet no Time*Section interaction (F = 0.573, p = .458) for the first-semester participants, and neither effect for the fourth-semester participants (Time: F = 0.930, p = .347 / Time*Section: F = 1.511, p = .234). No main effect was discovered for the Free Speech task for either proficiency level (First-Semester: F = .203, p = 0.658 / Fourth-Semester: F = .926, p = 0.926). Therefore, ANOVA analyses demonstrate that participants in the first-semester but not the fourth-semester showed significant improvement from pretest to posttest on the word-reading and paragraph-
reading tasks. Neither group showed a statistically significant improvement on the free speech tasks.

At first blush, these results suggest that the section the participants were in, that is, whether or not they received training, did not have an impact on their improvement from pre- to posttest. Yet one must remember that the participant numbers for each group in this study were small, which might not give enough strength to an ANOVA analysis. In order to take a closer look at the data and to decide whether this was truly the case, nonparametric statistical tests were conducted. These tests show exactly where any significant improvement can be found and allows for different comparisons between groups.

Wilcoxon Paired Rank tests compared pre- and posttest results within each section for statistically significant changes. Results reveal that among first-semester participants, those in both the experimental and control group showed significant improvement in their overall comprehensibility ratings (Exp: \( Z = -3.107, p = .002 \) / Control: \( Z = -2.934, p = .003 \)). First-semester students in both groups also showed significant improvement on the word- (Exp: \( Z = -2.760, p = .006 \) / Control: \( Z = -2.667, p = .008 \)) and paragraph-reading tasks (Exp: \( Z = -2.794, p = .005 \) / Control: \( Z = -2.547, p = .011 \)) individually. No significant improvement was found on the free speech task for either first-semester group (Exp: \( Z = -1.290, p = .197 \) / Control: \( Z = -0.211, p = .833 \)).

The fourth-semester participants did not show any significant improvement on either the overall comprehensibility rating, or any individual tasks. Table 6 provides results from the Wilcoxon tests for each of these discourse levels, split by group. Significant differences are marked with an asterisk.
Table 6

Results of Wilcoxon Signed Ranks Tasks for Comprehensibility Ratings per Discourse Level

<table>
<thead>
<tr>
<th></th>
<th>First-Semester Experimental</th>
<th>First-Semester Control</th>
<th>Fourth-Semester Experimental</th>
<th>Fourth-Semester Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>$Z = -3.107^*$</td>
<td>$Z = -2.934^*$</td>
<td>$Z = -0.533$</td>
<td>$Z = -1.172$</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>$p = .002$</td>
<td>$p = .003$</td>
<td>$p = .594$</td>
<td>$p = .241$</td>
</tr>
<tr>
<td>Word-Reading Task</td>
<td>$Z = -2.760^*$</td>
<td>$Z = -2.667^*$</td>
<td>$Z = -0.051$</td>
<td>$Z = -0.652$</td>
</tr>
<tr>
<td>Task</td>
<td>$p = .006$</td>
<td>$p = .008$</td>
<td>$p = .959$</td>
<td>$p = .515$</td>
</tr>
<tr>
<td>Paragraph-Reading</td>
<td>$Z = -2.794^*$</td>
<td>$Z = -2.547^*$</td>
<td>$Z = -0.356$, $Z = -1.599$,</td>
<td>$Z = -0.491$, $p = .110$</td>
</tr>
<tr>
<td>Task</td>
<td>$p = .005$</td>
<td>$p = .011$</td>
<td>$p = .722$</td>
<td>$p = .624$</td>
</tr>
<tr>
<td>Free Speech Task</td>
<td>$Z = -1.290$, $p = .197$</td>
<td>$Z = -0.211$, $p = .833$</td>
<td>$Z = -0.341$, $p = .733$</td>
<td>$Z = -0.491$, $p = .624$</td>
</tr>
</tbody>
</table>

$^*p < .05$

In addition to the overall and task type levels, analyses were also completed with the mean comprehensibility ratings in each individual condition and split by task. Mean comprehensibility ratings on the word-reading task are shown for first-semester experimental group in Figure 14 and the first-semester control group in Figure 15.

![Figure 14: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, First-Semester Experimental Group](image-url)
Figure 15: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, First-Semester Control Group

These figures reveal more about the improvement made by each group. For instance, one can see that large improvements were made by both groups for some conditions, like “Noncognates with 2\textsuperscript{nd} Syllable Stress” or the segment [x]. On the other hand, performance on other conditions was relatively similar for both groups and evidenced very little improvement, like in the [ʁ] and [o:] conditions. On the whole, however, it can be seen that the first-semester experimental group is consistently rated as more comprehensible in the individual conditions than the control group is. Additionally, the first-semester experimental group made larger gains in many individual conditions than the control group did, such as on “Cognates with 2\textsuperscript{nd} syllable stress”, “Noncognates with 2\textsuperscript{nd} syllable stress”, and [e], to name a few.

Figure 16 presents the mean comprehensibility ratings on the word-reading task for the fourth-semester experimental group, and Figure 17 presents these ratings for the fourth-semester control group.
Figure 16: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Experimental Group

Figure 17: Comprehensibility Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Control Group
These figures show that the fourth-semester experimental group often declined in comprehensibility where the control group improved. One notable results shown here is that the fourth-semester experimental group made an improvement of more than one point in their comprehensibility ratings on the word-reading task in the “Cognates with 2nd syllable stress” condition, which was an aspect of pronunciation extensively trained in the classroom.

Figures 18 and 19 show the mean comprehensibility ratings per individual condition on the paragraph-reading task for the first-semester groups. In these figures, one can see that on the whole, the experimental group made greater improvements in more individual conditions than the control group did. They also sometimes made gains where the control group declined in comprehensibility, such as the “Cognates with 2nd syllable stress” condition on the paragraph-reading task.

Figure 18: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Experimental Group
Figures 20 and 21 display the mean comprehensibility ratings per individual condition on the paragraph reading task for the two fourth-semester groups. These figures reveal that the fourth-semester groups made fewer and less drastic improvements than the first-semester groups. The fourth-semester experimental and control groups performed much more similarly on most conditions than the students in the first-semester groups did, and on the whole, the fourth-semester control group seems to have out-performed the first-semester experimental group.
Figure 20: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Experimental Group

Figure 21: Comprehensibility Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Control Group
Wilcoxon Signed Rank tests were conducted with the condition means within groups and reveal that significant improvements were made on the following conditions for the following groups:

Table 7

<table>
<thead>
<tr>
<th>Significant Improvements in Comprehensibility per condition per group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>First-Semester Experimental</td>
</tr>
<tr>
<td>First-Semester Control</td>
</tr>
<tr>
<td>Fourth-Semester Experimental</td>
</tr>
<tr>
<td>Fourth-Semester Control</td>
</tr>
</tbody>
</table>

The first-semester experimental group showed significant improvement in the largest number of individual conditions. They showed significant improvement in seven individual conditions, and approached a significant improvement in one other condition. In the word-reading task, participants in the first-semester experimental group exhibited significantly improved comprehensibility ratings for cognates with stress on the second syllable \((Z = -2.124, p = 0.034)\), noncognates with stress on the second syllable \((Z = 2.670, p = 0.008)\), the “Ach-Laut” \([x]\) \((Z = -2.003, p = 0.045)\), and the vocalic /r/ \([\text{ɾ}]\) \((Z = -2.669, p = 0.008)\) from pretest to posttest. On the paragraph-reading task, students in the experimental group showed significant improvement once more on cognates with stress on the first syllable \((Z = -2.952, p = 0.003)\), noncognates with stress on the second syllable \((Z = 02.156, p = 0.031)\), and the “Ich-Laut” \([ç]\) \((Z = 2.016, p = 0.044)\).
In addition, they approached significance on the “Cognates with 2nd syllable stress” condition ($Z = -1.939, p = 0.052$). The statistical results are summarized in Table 8 below.

<table>
<thead>
<tr>
<th>Module</th>
<th>Condition</th>
<th>Word-Reading Task</th>
<th>Paragraph-Reading Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Stress</td>
<td>Cognates with 1st syllable Stress</td>
<td>-</td>
<td>$Z = -2.952^*$</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 1st syllable Stress</td>
<td>-</td>
<td>$p = 0.003$</td>
</tr>
<tr>
<td></td>
<td>Cognates with 2nd syllable Stress</td>
<td>$Z = -2.124^*$</td>
<td>$Z = -1.939^a$</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 2nd syllable Stress</td>
<td>$p = 0.034$</td>
<td>$p = 0.052$</td>
</tr>
<tr>
<td></td>
<td>Ich and Ach</td>
<td>[ç]</td>
<td>$Z = -2.003^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p = 0.045$</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td></td>
<td>$Z = -2.065^*$</td>
</tr>
<tr>
<td></td>
<td>German /r/</td>
<td>[u]</td>
<td>$Z = -2.669^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p = 0.008$</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monophthongization</td>
<td>[e:]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[o:]</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
a. approaches significance at the $p < .05$ level

Participants in the first-semester control group showed statistically significant improvement in four individual conditions. They showed significantly improved comprehensibility ratings from pretest to posttest for noncognates with stress on the second syllable ($Z = -1.956, p = 0.050$), as well as on the “Ach-Laut” [x] ($Z = -2.192, p = 0.028$) and the monophthongization of long [e:] ($Z = -2.310, p = 0.021$). Participants in the first-semester control group showed significant improvement in only one condition on the paragraph-reading task, which was the monophthongization of long [o:] ($Z = -2.201, p = 0.028$). A summary of these statistical results are presented in Table 9 below.
Table 9

Significant Improvements in Comprehensibility per condition for First-Semester Control Group

<table>
<thead>
<tr>
<th>Module</th>
<th>Condition</th>
<th>Word-Reading Task</th>
<th>Paragraph-Reading Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Stress</td>
<td>Cognates with 1st syllable Stress</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 1st syllable Stress</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cognates with 2nd syllable Stress</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 2nd syllable Stress</td>
<td>Z = 1.956*</td>
<td>p = 0.050</td>
</tr>
<tr>
<td>Ich and Ach</td>
<td>[ç]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td>Z = -2.192*</td>
<td>p = 0.028</td>
</tr>
<tr>
<td>German /r/</td>
<td>[u]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[ʁ]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Monoph-thongization</td>
<td>[e:]</td>
<td>Z = -2.310*</td>
<td>p = 0.021</td>
</tr>
<tr>
<td></td>
<td>[o:]</td>
<td>-</td>
<td>Z = -2.201*</td>
</tr>
</tbody>
</table>

*p < .05

Participants in the fourth-semester groups did not improve significantly in as many individual conditions as the first-semester students did, which is understandable considering the fact that these groups also did not show significant improvement of either overall or task-related comprehensibility. Each of the fourth-semester groups only showed significant improvement in one individual condition. Participants in the experimental group showed significant improvement of comprehensibility ratings on cognates with stress on the first syllable on the paragraph-reading task (Z = -2.319, p = .020). Those in the control group show statistically significantly improvement from pretest to posttest in the cognates with first syllable stress on the word-reading task (Z = -2.207, p = .027).

Comprehensibility results as rated by native German speakers prove most interesting for the first-semester groups following this analysis. Both the first-semester experimental and control groups display significant improvement in their comprehensibility ratings from pretest to posttest on the overall test as well as on two of the three production tasks. Yet the experimental
group that had received training showed significant improvement on more individual conditions than did the control group. The experimental group showed significant improvement in four conditions in the word-reading task and three in the paragraph-reading task, with one more approaching significance. The participants in the control group, on the other hand, did not receive training and only showed significant improvement on two conditions in the word-reading task and one in the paragraph-reading task.

4.2.2 Accentedness Data
Results for the accentedness ratings given by native German speakers are provided in this section. Results are presented for all participants in all groups, as was done with the comprehensibility ratings above. Following standard practice, participants with an average accentedness rating of less than 2 out of 7 were excluded from the analysis. Data from one participant in the first-semester experimental group, five participants in the fourth-semester control group, and six participants in the fourth-semester experimental group were excluded. Data from the first-semester experimental group is presented first, followed by the first-semester control group, the fourth-semester experimental group, and the fourth-semester control group.

4.2.2.1 First-semester Participants
Average accentedness ratings for the first-semester experimental group stood at 2.868 at pretest and 2.494 at posttest. Participants in this group showed an improvement of 0.374 points in their accentedness from Time 1 to Time 2. This is a large amount of improvement and is

---

6 On the whole, average accentedness ratings for all groups were rated lower on the 7-point Likert scale than were comprehensibility ratings. This indicates that the native speaker raters judged participants' speech to be not very comprehensible, but fairly unaccented overall.
similar to the 0.313 point change that occurred in their overall comprehensibility ratings. Improvement was also found in the average overall ratings for participants in the first-semester control group. These participants were given a rating of 2.806 at pretest, which dropped to 2.638 at posttest. This shows an improvement of 0.168 points for first-semester participants in the control group.

As with the comprehensibility ratings, results of each of the three individual tasks were separated and averaged by group. On the word-reading task, participants in the first-semester experimental group were given an average rating of 2.787 out of 7 on the pretest, which improved 0.385 points to a posttest rating of 2.402. First-semester control participants had an average rating of 2.728 on the pretest for the word-reading task. Their average rating was 2.588 at posttest, for an improvement of 0.140 points.

On the paragraph-reading task, the first-semester experimental group began with an average rating of 2.969 at pretest, and changed to a rating of 2.580 at posttest. From Time 1 to Time 2, participants in this group improved 0.389 points. Participants in the control group also improved from pretest to posttest on the paragraph-reading task: they received an average rating of 2.764 on the pretest, and improved 0.190 points for a rating of 2.574 at posttest.

While for the comprehensibility ratings, the experimental group showed an improvement in mean rating for the free speech task and the control group showed a decline, a different pattern emerges for the accentedness ratings. In terms of average accentedness ratings, both groups showed improvement in their free speech from pretest to posttest. Participants in the first-semester experimental group received an average rating of 3.125 on the free speech task at pretest. Their scores improved to an average rating of 2.547 points at posttest, which demonstrates a very large improvement of 0.578 points, the largest improvement for any task or
rating type in these data. First-semester control participants also improved over time, receiving an average rating of 2.756 on the free speech task at pretest and an average rating of 2.604 at posttest. Therefore, although they did not improve their comprehensibility ratings on the free speech task from pretest to posttest, their average accentedness ratings on this task showed an improvement of 0.152 points.

The amount of change in accentedness ratings for the first-semester participants on each task are presented in Figure 22. As with Figures 7 and 10, the values on the y-axis have been reversed so that improvement is plotted positively. Figure 22 reveals that while first-semester groups demonstrated improvement in their accentedness ratings on every task from pretest to posttest, the experimental participants in the first-semester were able to reduce their accent to a much larger extent than the control participants.

![Figure 22: Amount of Change in Average Accentedness Ratings for First-Semester Participants](image)

As was the case with the comprehensibility ratings, the first-semester experimental group always showed more improvement over time than the control group. When looking at the
accentedness ratings, the first-semester control group improved an average of .163 points over time, while participants in the first-semester experimental group improved an average of 0.432 points. These experimental participants performed an average of 0.269 points better at posttest than the participants in the control group. As with the comprehensibility ratings, participants in the first-semester experimental group received better average accentedness ratings on the test as a whole, as well as on every individual task except for one. On the paragraph-reading task, the experimental group received an average rating of 2.580, while the control group received an almost equivalent average rating of 2.574. So although the experimental group did not receive a better rating than the control group on this one task, the numbers are so close as to consider their performance the same.

Figures 23 and 24 present results for mean accentedness ratings per person for both first-semester groups. Once again, the y-axis values have been reversed. The experimental group is shown in Figure 23, while the control participants are shown in 24. In the experimental group, ten of eleven participants improved their accentedness ratings from pre- to posttest, while only one showed a decline in average rating. Paired t-tests conducted with all of each participant’s accentedness ratings reveal that eight of these eleven participants (72.7%) demonstrated a statistically significant improvement in accentedness from Time 1 to Time 2. In the control group, nine participants improved and two declined in accentedness over time. Of these nine participants, three (33.3%) showed enough improvement to be considered statistically significant.
Figure 23: Average Accentedness Ratings, First-Semester Experimental
4.2.2.2 Fourth-semester Participants

Accentedness rating results from both of the fourth-semester groups have been summarized as well. Once data had been cleaned, many of the participants in the fourth-semester groups were excluded because they were already at ceiling. Data from six participants in the experimental and seven participants in the control groups were excluded, leaving data from the remaining five in the experimental and four in the control groups. Results from a Mann Whitney U test reveal that after cleaning of the data, the experimental and control groups were significantly different from one another in terms of accentedness ratings at pretest ($Z = -1.960$, $p$
=.05), but not at posttest (Z = -1.960, p = .05). This is something to keep in mind when interpreting the fourth-semester groups’ results below.

On the test as a whole, the fourth-semester experimental group’s average rating stood at 2.410 at pretest. This average rating changed to 2.360 at posttest, for a small improvement of 0.05 points. Participants in the fourth-semester control group showed a similar pattern, also improving in their average overall rating from pretest to posttest. The control group participants received an average rating of 2.197 at pretest and 2.086 at posttest, for an improvement of 0.111 points.

Results from each of the three individual tasks were once again separated and analyzed. Although for the test as a whole, the fourth-semester experimental group improved over time, this group showed a decline in accentedness ratings on the word-reading task at posttest. Participants in this group received an average accentedness rating of 2.328 at pretest, which declined 0.105 points and stood at 2.433 at posttest. The fourth-semester control group began with an average accentedness rating of 2.281, and ended with an average rating of 2.237. This group demonstrated an improvement of 0.044 points. As was found with the comprehensibility ratings, participants in the control group showed improvement in accentedness on the word-reading task, while the experimental participants did not.

While results from the word-reading task demonstrate an improvement by the fourth-semester control group and a decline by the fourth-semester experimental group, both groups improved on the paragraph-reading task. Average accentedness ratings on the paragraph-reading task stood at 2.306 for the fourth-semester experimental group at pretest, and changed to 2.274 at posttest. This group showed a slight improvement of 0.032 points over time. The control group
received average accentedness ratings of 2.166 at pretest and 2.064 at posttest, which is an improvement of 0.102 points over time.

Data from the free speech task show similar results as on the word-reading task, in that the fourth-semester experimental group showed a decline in average rating, while the control group received better ratings at posttest than pretest. Average accentedness ratings for the fourth-semester experimental group were 2.111 at pretest, and declined slightly to 2.150 points at posttest, for a change of 0.039 points. On the other hand, the control group improved 0.259 points from pretest to posttest, moving from an average pretest rating of 2.459 to 2.200 at posttest. The amount of change in accentedness ratings on the overall test and all separate tasks are presented in Figure 25, with the y-axis values reversed.

![Figure 25: Amount of Change in Average Accentedness Ratings for Fourth-Semester Participants](image)

To summarize, accentedness results for the fourth-semester groups are more unpredictable than those for the first-semester groups, just as they are with the comprehensibility ratings. Participants in the fourth-semester control group showed a reduction of accent on the
overall test and on all tasks, while the fourth-semester experimental participants only improved their accent on the overall test and on the paragraph-reading task.

Another similarity to the comprehensibility ratings of the fourth-semester groups is that for accentedness ratings, the amounts of change are smaller than for the first-semester groups. The amount of change stood at an average of 0.163 for the first-semester control group and 0.432 for the first-semester experimental group, while average amounts of change for the fourth-semester groups were 0.129 for the control and only 0.057 for the experimental. Participants in the fourth-semester groups not only showed less change in their accentedness as perceived by native speaker raters than those in the first-semester, they also received worse ratings in some cases.

The change in average accentedness ratings for each participant are shown in Figure 26 for the fourth-semester experimental group and in Figure 27 for the fourth-semester control group. Keeping in mind that most of the accentedness data from the fourth-semester groups had to be excluded due to ceiling effects, data from four participants in the experimental and five in the control are shown. Out of the four experimental group participants, three improved over time and one showed a decline. Paired t-tests reveal that none of the fourth-semester experimental participants showed a statistically significant improvement, but that one participant showed a statistically significant decline in accentedness ratings. One participant in the fourth-semester control group also declined over time, while one received the same average score at both pretest and posttest, and three participants showed improvement. Only one participant out of the three who improved (33.3%) actually improved enough for the results to be determined statistically significant with a paired t-test.
Figure 26: Average Accentedness Ratings, Fourth-Semester Experimental
Mean accentedness ratings from all four groups are presented in Figure 28. Once again, the y-axis values have been reversed, so that a line demonstrating improvement will move upwards, and a line indicating a decline will move downwards. Figure 28 clearly shows that the first-semester groups showed the most improvement from pretest to posttest. The fourth-semester groups, while receiving better accentedness ratings on the whole, show a less severe change in mean rating.

Figure 27: Average Accentedness Ratings, Fourth-Semester Control
Figure 28: Mean Accentedness Ratings per Group

Results from the accentedness ratings can be summarized as follows. Both first-semester groups showed improvement in their overall accentedness ratings from pretest to posttest, as well as on their average ratings for all three tasks. Both fourth-semester groups improved their average accentedness ratings from pretest to posttest on the test as a whole, but the picture for each individual task is less clear. The fourth-semester experimental group received poorer ratings of accentedness on the word-reading and free speech tasks, but did show improvement on the paragraph-reading task. Conversely, participants in the fourth-semester control group improved their ratings of accentedness on all three individual tasks. As with the comprehensibility ratings, first-semester participants improved more from pretest to posttest than did those in the fourth-semester.

4.2.2.3 Statistical Analyses & Results

Mean accentedness ratings were also submitted to a repeated measures ANOVA in order to discover whether group had an influence on the amount of change over time. Time was used as the within-subjects factor, Group the between-subjects factor, and mean accentedness ratings
as the dependent variable. The ANOVA for accentedness ratings on the test as a whole showed a
main effect for time (F = 22.182, p < .001), but no Time*Section interaction (F = 3.214, p =
.086) among first-semester groups. As with the comprehensibility ratings, no main effect (F =
1.871, p = .214) or interaction (F = 0.260, p = .626) was found for the fourth-semester groups. A
significant change in accentedness over time can be found for all first-semester participants, but
there were no significant differences based on the group that a participant was in. No significant
changes were present among fourth-semester groups.

This type of ANOVA was also conducted for each individual task. A main effect of
Time (F = 16.430, p = .001) and a Time*Section interaction (F = 4.382, p = .048) were both
found on the word-reading task among first-semester groups. However, no such results were
present for the two fourth-semester groups on the word-reading task (Time: F = 0.216, p = .653 /
Time*Section: F = 1.291, p = .285). On the paragraph-reading task, a main effect of Time (F =
15.955, p = .001), but no Time*Section interaction (F = 2.261, p = .148) was discovered among
first-semester participants. Fourth-semester groups did not demonstrate a main effect of Time (F
= 0.819, p = .387) or a Time*Section interaction (F = 0.227, p = .644). Similar results were
found when analyzing the free speech task: a main effect of Time (F = 10.197, p = .005) without
a Time*Section interaction (F = 3.474, p = 0.78) was present for first-semester students, but once
again, no main effect of Time (F = 0.361, p = .565) or Time*Section interaction (F = 0.661, p =
.440) were found for the fourth-semester participants. Therefore, it can be seen that first-
semester participants showed significant improvement from pretest to posttest in all three tasks,
but according to the results of the ANOVA, section only had a significant influence on the
change in score over time for the word-reading task. Fourth-semester students did not show a
significant improvement on any individual task, irrespective of group.
Preliminary results from the ANOVAs suggest that the group that participants were in did not have a significant impact on their change in ratings from pretest to posttest. Yet one must recall that the participant numbers are small in this study, particularly for the accentedness ratings of the fourth-semester groups, which had over half of the same discarded due to ceiling effects. Simple nonparametric statistical tests were chosen to further analyze the data and to see whether training and proficiency level had an effect on accentedness ratings. Wilcoxon Paired Rank tests were completed to compare pre- and posttest ratings within each group.

These Wilcoxon tests reveal that the first-semester experimental group showed significant improvement in their accentedness on the overall test \( (Z = -2.794, p = .005) \), as did the control group \( (Z = -2.401, p = .016) \). The control group did not demonstrate any significant improvement beyond this, however, while the first-semester experimental group also significantly improved their accentedness ratings on each of the individual tasks: word-reading \( (Z = -2.746, p = .006) \), paragraph-reading \( (Z = -2.900, p = .004) \) and free speech \( (Z = -2.472, p = .013) \).

As with the comprehensibility ratings, participants in both fourth-semester groups did not show any significant improvement in their accentedness ratings from pretest to posttest. This holds both for the test as a whole, as well as for each individual task individually. Results from these Wilcoxon tests are split by group and presented in Table 10. Significant differences are marked with an asterisk.
Table 10

Results of Wilcoxon Signed Ranks Tasks for Accentedness Ratings per Discourse Level

<table>
<thead>
<tr>
<th></th>
<th>First-Semester Experimental</th>
<th>First-Semester Control</th>
<th>Fourth-Semester Experimental</th>
<th>Fourth-Semester Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accentedness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-Reading Task</td>
<td>Z = -2.746, p = .006*</td>
<td>Z = -1.779, p = .075</td>
<td>Z = -0.730, p = .465</td>
<td>Z = -0.944, p = .345</td>
</tr>
<tr>
<td>Paragraph-Reading</td>
<td>Z = -2.900, p = .004*</td>
<td>Z = -1.718, p = .086</td>
<td>Z = -0.734, p = .463</td>
<td>Z = -0.943, p = .345</td>
</tr>
</tbody>
</table>

*p < .05

Mean ratings from each of the individual conditions were also calculated by task and group. Ratings per condition on the word-reading task are presented for the first-semester experimental group in Figure 29 below. Mean ratings per condition are shown for the first-semester control group in Figure 30.
Figure 30: Accentedness Ratings of Individual Conditions on the Word-Reading Task, First-Semester Control Group

The mean accentedness ratings per individual condition show us a similar pattern as the comprehensibility ratings did. On the word-reading task, the first-semester experimental group showed more improvement than the control group did on many categories, including: “Noncognates with 1\textsuperscript{st} syllable stress”, “Noncognates with 2\textsuperscript{nd} syllable stress”, and [ʁ]. On four of the conditions, “Cognates with 1\textsuperscript{st} syllable stress”, “Cognates with 2\textsuperscript{nd} syllable stress”, [ʁ], and [oː], the first-semester experimental group showed large amounts of improvement in accentedness, while the control group received worse ratings at Time 2 than they had at Time 1.

Figures 31 and 32 summarize the mean accentedness ratings in each condition for the fourth-semester groups. These figures show that the fourth-semester experimental group also showed more improvement in some individual conditions on the word-reading task than the control group did, such as “Cognates with 2\textsuperscript{nd} syllable stress” and [oː], although as was the case with the comprehensibility ratings, the experimental group actually declined in their accentedness ratings while the control group improved in many conditions.
Figure 31: Accentedness Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Experimental Group

Figure 32: Accentedness Ratings of Individual Conditions on the Word-Reading Task, Fourth-Semester Control Group
Figures 33 and 34 show the mean accentedness ratings on the paragraph-reading tasks, split by individual condition. Figure 33 presents results for first-semester experimental group, and Figure 34 for the first-semester control group.

Figure 33: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Experimental Group
Figure 34: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, First-Semester Control Group

On the paragraph-reading task, the first-semester experimental group showed greater improvement from Time 1 to Time 2 than the control group had in the “Cognates with 2\textsuperscript{nd} syllable stress”, [ç], [ʁ], and [e:], among others. Yet for this task, there were no conditions where they improved while the control group declined.

Figures 35 and 36 present the mean accentedness ratings on the paragraph-reading tasks for the fourth-semester experimental and control groups, respectively. Participants in the fourth-semester experimental group improved more than the fourth-semester control group in four conditions: “Cognates with 1\textsuperscript{st} syllable stress”, “Cognates with 2\textsuperscript{nd} syllable stress”, “Noncognates with 2\textsuperscript{nd} syllable stress”, and [ʁ].
Figure 35: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Experimental Group

Figure 36: Accentedness Ratings of Individual Conditions on the Paragraph-Reading Task, Fourth-Semester Control Group
The means accentedness ratings from each individual condition were also analyzed via Wilcoxon Signed Rank tests for further analysis. Wilcoxon tests reveal that significant improvements were made on the following conditions for the following groups, divided by task:

Table 11

<table>
<thead>
<tr>
<th>Group</th>
<th>Word-Reading Task</th>
<th>Paragraph-Reading Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognates with 2nd syllable stress</td>
<td>Noncognates with 2nd syllable stress</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 2nd syllable stress</td>
<td>[x]</td>
</tr>
<tr>
<td></td>
<td>[e:]</td>
<td>[o:]</td>
</tr>
<tr>
<td></td>
<td>[ʁ]</td>
<td></td>
</tr>
</tbody>
</table>

Approaches Significance:
Cognates with 1st syllable Stress

<table>
<thead>
<tr>
<th>Group</th>
<th>Word-Reading Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Noncognates with 2nd syllable stress</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
</tr>
<tr>
<td></td>
<td>[e:]</td>
</tr>
</tbody>
</table>

Once again, the first-semester experimental group showed significant improvement in the largest number of individual conditions. Participants in this group showed significant improvement in eight individual conditions, five of which were found on the word-reading task and three on the paragraph-reading task. The first-semester experimental group also approached a significant improvement in one other condition on the word-reading task. On the word-reading task, the first-semester experimental group exhibited significantly improved accentedness ratings for cognates with stress on the second syllable ($Z = -2.205, p = 0.027$), noncognates with stress on the second syllable ($Z = -2.675, p = 0.007$), the “Ach-Laut” [x] ($Z = -2.040, p = 0.041$), the vocalic /r/ [ʁ] ($Z = -2.395, p = 0.017$), and the monophthongization of [e:] ($Z = -2.492, p = 0.013$). Of these five individual conditions, all but the last on the list also showed significant
improvement in comprehensibility ratings in the previous section. Participants in the experimental group also showed an improvement in the “Noncognates with stress on the first syllable” condition that approached significance on the word-reading task ($Z = -1.897, p = 0.058$). On the paragraph-reading task, students in the first-semester experimental group exhibited significant improvement on accentedness of noncognates with stress on the second syllable ($Z = -2.310, p = 0.021$), the “Ich-Laut” [ç] ($Z = -2.987, p = .003$) and the monophthongization of [ø:] ($Z = -2.677, p = .007$). Of these, the first two also represented a significant improvement on the paragraph-reading task for comprehensibility ratings. Table 12 summarizes these statistical results.

Table 12

<table>
<thead>
<tr>
<th>Module</th>
<th>Condition</th>
<th>Word-Reading</th>
<th>Paragraph-Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Stress</td>
<td>Cognates with 1st syllable Stress</td>
<td>$Z = -1.897^a$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 1st syllable Stress</td>
<td>$Z = -1.897^a$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cognates with 2nd syllable Stress</td>
<td>$Z = -2.205^*$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Noncognates with 2nd syllable Stress</td>
<td>$Z = -2.675^*$</td>
<td>$Z = -2.310^*$</td>
</tr>
<tr>
<td>Ich and Ach</td>
<td>[ç]</td>
<td>$Z = -2.040^*$</td>
<td>$Z = -2.987^*$</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td>$Z = -2.040^*$</td>
<td>-</td>
</tr>
<tr>
<td>German /ʁ/</td>
<td>[ɐ]</td>
<td>$Z = -2.395^*$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[ʁ]</td>
<td>$Z = -2.492^*$</td>
<td>-</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>[e:]</td>
<td>$Z = -2.492^*$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>[o:]</td>
<td>$Z = -2.677^*$</td>
<td>$Z = -2.677^*$</td>
</tr>
</tbody>
</table>

*a. approaches significance at the $p < .05$ level

*$p < .05$*

In contrast, participants in the first-semester control group showed significantly improved accentedness ratings in a total of four different individual conditions. On the word-reading task,
they showed significant improvement on noncognates with stress on the second syllable \((Z = -1.992, p = 0.046)\), on the “Ach-Laut” [x] \((Z = -2.449, p = 0.014)\), and the monophthongization of [e:] \((Z = -2.524, p = 0.012)\). On the paragraph-reading task, the first-semester control group demonstrated significantly improved accentedness ratings in one condition: the monophthongization of [o:] \((Z = -2.384, p = 0.017)\). These were the same four condition and task combinations that showed significant improvement in comprehensibility ratings as well, described in the previous section. The results of these analyses are summarized in Table 13.

Table 13

<table>
<thead>
<tr>
<th>Module</th>
<th>Condition</th>
<th>Word-Reading Task</th>
<th>Paragraph-Reading Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Stress</td>
<td>Cognates with 1st syllable Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noncognates with 1st syllable Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognates with 2nd syllable Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noncognates with 2nd syllable Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Z = -1.992^*)</td>
<td></td>
<td>(Z = -2.449^*)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.046)</td>
<td></td>
<td>(p = 0.014)</td>
</tr>
<tr>
<td>Ich and Ach</td>
<td>[ç]</td>
<td></td>
<td>(Z = -2.524^*)</td>
</tr>
<tr>
<td></td>
<td>[x]</td>
<td></td>
<td>(p = 0.012)</td>
</tr>
<tr>
<td>German /r/</td>
<td>[ɐ]</td>
<td></td>
<td>(Z = -2.384^*)</td>
</tr>
<tr>
<td></td>
<td>[ʁ]</td>
<td></td>
<td>(p = 0.017)</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>[e:]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[o:]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^*p < .05\)

Parallel to the results for comprehensibility ratings, the fourth-semester groups did not demonstrate significant improvement on accentedness in as many individual conditions. Neither of the fourth-semester groups demonstrated significantly improved accentedness ratings on the word-reading task. On the paragraph-reading task, the fourth-semester experimental group only demonstrated significant improvement in one condition, while the fourth-semester control group significantly improved in two. The experimental participants improved their accentedness
ratings significantly on the fricative /ɾ/ [ʁ] (Z = -1.997, p = .046). The control participants significantly improved their accentedness ratings for cognates with stress on the first syllable (Z = -2.060, p = .039), and the “Ich-Laut” [ç] (Z = -1.992, p = .046).

Accentedness results can be summarized as follows. While both first-semester groups showed significant overall improvement from pretest to posttest, the experimental group displayed significant improvement on each individual task as well. In addition, the experimental group improved in eight individual conditions, as compared to the control group’s four. Participants in the control group did not receive training on these different aspects of German pronunciation, and did not show as much improvement. Participants in the fourth-semester groups showed the least improvement of all. They did not significantly improve on the test as a whole or on any of the three task types, and only improved on a few individual conditions. Another interesting result presented here is that for the most part, the individual conditions that exhibited significant improvement in accentedness were the same as those which did in comprehensibility, a fact that will be discussed further in section 4.5.

Table 14 clarifies each of the improvements made by the first and fourth-semester groups and shows a clear picture as to the effects of training. As one can see in this table, the first-semester experimental group made the greatest gains in both comprehensibility and accentedness. They significantly improved in comprehensibility and accentedness overall, and also on all test tasks except for the comprehensibility of free speech. In addition, they improved in more individual conditions in both comprehensibility and accentedness than any other group. The first-semester control group also made gains, but these were fewer and more sporadic than those made by the first-semester experimental group, resulting in a less clear picture.
<table>
<thead>
<tr>
<th>Perception</th>
<th>First-Semester Experimental</th>
<th>First-Semester Control</th>
<th>Fourth-Semester Experimental</th>
<th>Fourth-Semester Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Comprehensibility</td>
<td>improved</td>
<td>improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility – Word-Reading</td>
<td>improved</td>
<td>improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility – Paragraph-Reading</td>
<td>improved</td>
<td>improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility – Free Speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility – # of Individual Conditions with significant improvement (out of 10)</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Overall Accentedness</td>
<td>improved</td>
<td>improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness – Word-Reading</td>
<td>improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness – Paragraph-Reading</td>
<td>improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness – Free Speech</td>
<td>improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness – # of Individual Conditions with significant improvement (out of 10)</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
What is striking in the above table is how little significant improvement was made by the fourth-semester participants. Training seemed to have had beneficial effects on first-semester students, yet the same training program did not seem to affect the fourth-semester students in similar ways. Further analysis of the fourth-semester groups is necessary in order to get a clearer picture of what, if any, effects training had on these participants. As the two fourth-semester groups are very small, and much of their data was cleaned from the analysis because it was already at ceiling, the following breakdown will contain descriptive factors only.

The fourth-semester groups were analyzed according to several different factors that may have affected their performance before and after training. By the very nature of the course, students in the fourth semester have had much more experience learning German than students in the first-semester. Participants in the fourth-semester groups had been studying German for more than three years on average before the time of the study. Therefore, previous language experience is a key factor in analyzing the performance of the fourth-semester students. Two further factors taken into account in this secondary analysis are study abroad experience and motivation levels.\(^7\)

Previous study of the German language varied greatly among fourth-semester participants. Fourth-semester experimental participants had studied German between 1.5 and 6 years before taking the fourth-semester course, while experience in the fourth-semester control group varied from 1.5 to 6.5 years. Averages for each group were 3.8 years for the experimental and 3.6 years for the control. Within each group, a median split was used to create two sub-groups of five participants: a high experience and a low experience group. Each of these groups had five participants in it. The effects of previous German experience on perception and

\(^7\) Motivation levels will be discussed briefly in this section and further in section 4.4.
pronunciation accuracy were measured. A description of improvement by experience group and training section is presented in Table 15.

Table 15

_Percents of Fourth-Semester Participants who Changed their Perception and Pronunciation Accuracy as a Factor of Previous Language Experience_  

<table>
<thead>
<tr>
<th>Training Group</th>
<th>Experience Group</th>
<th>Perception Accuracy</th>
<th>Comprehensibility Ratings</th>
<th>Accentedness Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth-Semester Experimental</td>
<td>High</td>
<td>80% at ceiling</td>
<td>40% improved</td>
<td>60% at ceiling</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>40% at ceiling</td>
<td>60% improved</td>
<td>60% at ceiling</td>
</tr>
<tr>
<td>Fourth-Semester Control</td>
<td>High</td>
<td>80% at ceiling</td>
<td>80% improved</td>
<td>60% at ceiling</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>20% improved</td>
<td>20% declined</td>
<td>20% improved</td>
</tr>
</tbody>
</table>

Participants who were assigned to the high experience groups received better ratings on comprehensibility and accentedness at pretest than the low experience groups. The majority of participants in the high experience groups were also already performing at ceiling in perception and accentedness accuracy at pretest, suggesting that more experience can lend itself to better initial pronunciation skills.

All of the high-experience participants not at ceiling improved their accuracy in perception and accentedness when they received classroom training. Participants with high amounts of previous experience who did not receive training did not show as much consistency; some of the participants in the control group declined in accuracy in perception and accentedness. Results from the comprehensibility ratings are contradictory, however; a majority of the high-experience students in the experimental group received a worse score at posttest.
Participants in the low experience groups were rated as less comprehensible and more accented at pretest than the high experience group. They also showed less consistent improvement over time and many participants received poorer scores at posttest on comprehensibility, accentedness, and perception accuracy. However, low-experience participants who did improve demonstrated greater gains than those in the high experience group. Participants in the experimental section who were in the high experience group improved an average of 0.101 points, while those in the low experience group improved an average of 0.163 points.

These results indicate that previous experience with the German language does seem to have an impact on the ability to improve pronunciation accuracy over time. More language experience affects how likely a participant is to already perform at ceiling at pretest and to receive better pretest ratings. Previous language experience also seems advantageous in improving perception and accentedness skills, but this seems to only be the case when this language experience is combined with the added benefit of training. From these data, one cannot say that the combination of language experience and training will result in better pronunciation accuracy by intermediate students in all measured areas. Yet overall, this analysis indicates that the combination of training and previous experience with the L2 is beneficial for intermediate students, and may also help a few students with less experience and lower pronunciation abilities improve to a greater extent.

Study abroad experience also seems to affect pronunciation performance by intermediate-level participants. Three participants in the experimental group had spent an average of one month studying abroad before this study. Of these three participants, all were already at ceiling for the perception test and accentedness ratings. Moreover, these three students also received the
three best mean ratings for comprehensibility at pretest, and two of them (67%) were able to improve their comprehensibility even more. Just one of the participants with previous study abroad experience (33%) was rated as less comprehensible at posttest. The students in the experimental group who had not studied abroad before enrolling in their course show a less consistent pattern. Except for accentedness ratings, in which 43% were at ceiling and 43% improved, students typically displayed equal numbers of students who improved and declined in ratings over time.

Two participants in the fourth-semester control group had studied abroad for an average of 5 weeks. These students also seemed to perform in a more consistent way than the students who had not studied abroad. Both participants improved their comprehensibility ratings over time, although they did not begin with the best scores at pretest as those in the experimental group had. One participant improved their accentedness ratings while the other was already at ceiling. On the perception test, one student was already at ceiling and the other demonstrated the same score at pretest as at posttest. Of the students who did not have previous study abroad experience, 75% were able to improve their comprehensibility, but only 25% were able to improve their perception test accuracy and accentedness ratings.

It seems that previous study abroad experience has an effect on pronunciation improvement. Of the five participants who had previous study abroad experience, the vast majority were already at ceiling or were able to improve their pronunciation skills. This ability to improve was not so clear among students who had not studied abroad, and many students without study abroad experience also declined in their accuracy. Important to note is that study abroad experience does not seem to positively affect how much a student will improve over time. Many of the students who had studied abroad, while improving from Time 1 to Time 2, made
less drastic gains than students who had not studied abroad. Yet these same students tended to begin with better ratings in comprehensibility and accentedness than the others. This may be similar to the high experience groups described above; it is possible that students who had not studied abroad had more room to improve than those who had.

Motivation level was the last factor analyzed with the intermediate-level data. A more in-depth discussion of the results collected from the Pronunciation Attitude Inventory is provided in Section 4.4, but a preliminary analysis is presented here. As with experience level, participants were divided into high and low motivation groups via a median split based on their answers on the Pronunciation Attitude Inventory, resulting in four groups of five participants each. Results as analyzed by motivation level and training group are summarized in Table 16 and described below.

Table 16

<table>
<thead>
<tr>
<th>Training Group</th>
<th>Motivation Group</th>
<th>Perception Accuracy</th>
<th>Comprehensibility Ratings</th>
<th>Accentedness Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Semester Experimental</td>
<td>High</td>
<td>60% at ceiling</td>
<td>80% improved</td>
<td>80% at ceiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% improved</td>
<td>20% declined</td>
<td>20% improved</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>60% at ceiling</td>
<td>20% improved</td>
<td>40% at ceiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% improved</td>
<td>80% declined</td>
<td>40% improved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% declined</td>
<td></td>
<td>20% declined</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Semester Control</td>
<td>High</td>
<td>20% at ceiling</td>
<td>100% improved</td>
<td>60% at ceiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% improved</td>
<td>60% improved</td>
<td>40% improved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% declined</td>
<td>40% declined</td>
<td>20% declined</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>60% at ceiling</td>
<td>60% improved</td>
<td>40% at ceiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% improved</td>
<td>40% declined</td>
<td>20% improved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20% declined</td>
<td>20% declined</td>
<td>20% no change</td>
</tr>
</tbody>
</table>
Most of the participants with a high motivation level who had received training were already at ceiling in perception and accentedness skills at pretest. Only 20% of participants in this group received worse scores in comprehensibility and perception accuracy at posttest; the rest of the participants not already at ceiling improved in their skills.

Participants with higher motivation levels were also able to benefit more from in-class training. 80% (n=4) of students who had received training but were classified as low motivation received worse comprehensibility ratings at posttest than at pretest. This stands in stark contrast to those students with higher motivation levels, where the majority of participants improved their comprehensibility ratings over time. Additionally, participants in the experimental section with high motivation who improved were able to increase their comprehensibility ratings by 0.155 points on average, while those who had also received training but demonstrated lower levels of motivation increased their comprehensibility ratings by an average of 0.072 points.

Motivation level also seemed to have effects on students who had not received training. All students in the high motivation group received better comprehensibility ratings at posttest, and all those not at ceiling at pretest received better accentedness ratings as well. Among students who did not receive training, participants with high motivation did not demonstrate larger improvement than students with lower motivation. In fact, the opposite is true: high motivation students improved an average of 0.128 points on comprehensibility, while those with low motivation improved 0.244 points.

These results suggest that motivation seems to influence an intermediate-level students’ improvement in perception, comprehensibility and accentedness more than training does. Students with higher motivation levels were able to demonstrate more and larger amounts of improvement than those with lower motivation levels, independent of whether they had received
in-class pronunciation training. Even students who had received training, but who had lower levels of motivation, did not improve to the same extent as highly-motivated students without training. This suggests that motivation may override training in terms of pronunciation accuracy.

A combination of high motivation levels and large amounts of previous experience with German seemed to be particularly beneficial for participants in this study; participants exhibiting this combination fared best and were able to improve their skills in all areas. For example, the one participant with this combination in the experimental group was at ceiling for perception and accentedness and improved their comprehensibility scores. The combination of low motivation and low experience seemed to be particularly harmful to students, independent of group; these participants received worse scores on every skill at posttest than they had at pretest.

Overall, this breakdown of results demonstrates that previous language experience, study abroad experience, and motivation level all have effects on an intermediate-level participants’ ability to improve their pronunciation. Attitude seems to have the largest effect on the fourth-semester groups’ performance on the pronunciation tests administered in this study. Among high motivation groups in both the experimental and control sections, the majority of participants were either already at ceiling or improved. There were more individual participants in the low motivation groups who declined in accuracy over time, and 80% (n=4) of the participants in the experimental low motivation group received worse ratings in comprehensibility at posttest. Yet the combination of training and high motivation levels seems to make it more likely that participants will improve; these participants improved their comprehensibility ratings to a greater extent than participants who had low levels of motivation. This pattern is not replicated among participants who did not receive training.
4.3 Generalization

A key goal in pronunciation training studies is the hope that participants will be able to take what they have learned during training and generalize that knowledge to other words and sentences. An ideal training paradigm would be one that aids L2 learners in improving their pronunciation not only of the words and sentences that specifically trained, but also those that they have never encountered before. This gives their knowledge a practical use in L2 communication and can improve their comprehensibility and accent as a whole within the L2.

While this experiment did not have true novel words or paragraphs, because the pre- and posttest contained the same stimuli, a different method of testing generalization is available. During the eight-week training period, half of the individual word stimuli were targeted for training. These words were used repeatedly in partner activities, games, and exercises. Participants practiced their pronunciation of these words many times and received feedback from the instructor and their in-class partners. The other half of the stimuli were not trained and were purposely excluded from all activities during training. Thus, in this analysis, there are two groups of words—trained words and untrained words.

Pre- and posttest mean ratings in comprehensibility and accentedness were calculated for both trained and untrained words on the word- and paragraph-reading tasks, for a total of sixteen rating categories per group (8 on the pretest and 8 on the posttest). Analyses of trained and untrained words were also conducted for the two control groups, despite the fact that neither of these groups received any in-class pronunciation training. Their results are analyzed for purposes of comparison only, and to shed light on whether any generalization that may be found among experimental groups can be explained by training.
4.3.1 Generalization of Comprehensibility

The amount of change from Time 1 to Time 2 is presented in Figure 37 for each word group, test task and training group. Mean comprehensibility ratings of trained words improved for both first-semester groups on both tasks. Participants in both fourth-semester groups showed poorer comprehensibility ratings for trained words on the word-reading task at posttest than they had at pretest, but both fourth-semester groups showed improved ratings for trained words on the paragraph-reading task at posttest. Among untrained words, both first-semester groups showed improvement in comprehensibility on the word-reading and paragraph-reading tasks over time. The fourth-semester experimental group also showed improvement on untrained words on the word-reading task, while the fourth-semester control group showed a slight decline in mean rating over time. Both fourth-semester groups improved on untrained words on the paragraph-reading task from pretest to posttest.

*Figure 37: Generalization of Pronunciation Skills in Comprehensibility*
What is surprising here is the fact that the first-semester control groups improved more than the first-semester experimental groups did when the data is split by trained and untrained words. This does not happen among fourth-semester groups, although it is also useful to note that all changes, both positive and negative, were very small for the fourth-semester groups, typically less than 0.1 of a point.

The mean ratings were divided by student group and submitted to nonparametric alternates to paired t-tests, the Wilcoxon Signed Ranks test. Wilcoxon Signed Ranks tests reveal that mean ratings for trained words on both the word- and paragraph-reading tasks showed significant improvement in comprehensibility for the first-semester experimental group, while the first-semester control group showed improvement among trained words on the paragraph-reading task only. Results from the Wilcoxon tests conducted with comprehensibility ratings are presented in Table 17 below.

Table 17

<table>
<thead>
<tr>
<th>Group</th>
<th>First-Semester</th>
<th></th>
<th>Fourth-Semester</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
<td>Experimental</td>
<td>Control</td>
</tr>
<tr>
<td>Trained Words in</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-reading Task</td>
<td>Z = -3.140*</td>
<td>p = .002</td>
<td>Z = -1.478</td>
<td>p = .139</td>
</tr>
<tr>
<td>Untrained Words in</td>
<td>Z = -2.746*</td>
<td>p = .006</td>
<td>Z = -2.578*</td>
<td>p = .010</td>
</tr>
<tr>
<td>Word-Reading Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained Words in</td>
<td>Z = -2.229*</td>
<td>p = .026</td>
<td>Z = -2.191*</td>
<td>p = .028</td>
</tr>
<tr>
<td>Paragraph-Reading Task</td>
<td></td>
<td></td>
<td>Z = -1.112</td>
<td>p = .266</td>
</tr>
<tr>
<td>Untrained Words in</td>
<td>Z = -2.118*</td>
<td>p = .034</td>
<td>Z = -2.045*</td>
<td>p = .041</td>
</tr>
<tr>
<td>Paragraph-Reading Task</td>
<td></td>
<td></td>
<td>Z = -1.156</td>
<td>p = .248</td>
</tr>
</tbody>
</table>

* = significant at p < .05 level

These results demonstrate that the first-semester experimental group was able to improve significantly on both words that were trained in class, as well as those that were avoided during
training. This seems at first-glance to point to generalization effects. Yet the participants in the first-semester control group did this as well, demonstrating significant improvement in comprehensibility among untrained words on both tasks. These results are particularly striking considering that these participants did not receive training. The fact that participants in the control group still showed significant improvement in comprehensibility on three of the four tasks seems to contradict the idea that the first-semester experimental group improved because of generalization effects due to training alone. The elementary-level improvement in comprehensibility on both trained and untrained words may derive from some factor other than training.

4.3.2 Generalization of Accentedness

Figure 38 presents the mean accentedness ratings for each training group by word group and task. Mean ratings of accentedness for both trained and untrained words improved on both tasks for both first-semester groups. Conversely, mean accentedness ratings for trained and untrained words on the word-reading task declined for both fourth-semester groups. On the paragraph-reading task, participants in the fourth-semester experimental group showed a decline in ratings at posttest on trained words, while those in the control group showed improved ratings at posttest for trained words. For untrained words both the fourth-semester experimental and control groups showed improved accentedness ratings on the paragraph-reading task.

In comparison to the comprehensibility rating data, the data for accentedness ratings demonstrate a marked improvement for the first-semester experimental group, while the participants in the first-semester control group did not show as large of an improvement. However, compared with change in mean ratings for the first-semester groups, those for the fourth-semester groups are much smaller.
Wilcoxon Signed Ranks tests were also conducted with accentedness ratings. The tests show a very different picture than that which was found for comprehensibility ratings. The first-semester experimental group was the only group that was able to demonstrate a statistically significant improvement in accentedness ratings over time. This group showed significant improvement for all word types on all tasks (trained and untrained; word- and paragraph-reading). Improvement was not found to be significant for the first-semester control group on any task or for any group of words, and the changes exhibited by the two fourth-semester groups were also not found to be significantly different. Table 18 presents the results of the Wilcoxon tests calculated for accentedness ratings.
Table 18

**Statistical Analyses of Generalization of Accentedness per task per group**

<table>
<thead>
<tr>
<th>Group</th>
<th>First-Semester</th>
<th></th>
<th>Fourth-Semester</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
<td>Experimental</td>
<td>Control</td>
</tr>
<tr>
<td>Trained Words in Word-reading Task</td>
<td>Z = -2.341*</td>
<td>Z = -1.511</td>
<td>Z = -0.978</td>
<td>Z = 0.420</td>
</tr>
<tr>
<td></td>
<td>p = .019</td>
<td>p = .131</td>
<td>p = .328</td>
<td>p = .674</td>
</tr>
<tr>
<td>Untrained Words in Word-Reading Task</td>
<td>Z = -2.103*</td>
<td>Z = -1.511</td>
<td>Z = -1.468</td>
<td>Z = -0.652</td>
</tr>
<tr>
<td></td>
<td>p = .035</td>
<td>p = .131</td>
<td>p = .142</td>
<td>p = .515</td>
</tr>
<tr>
<td>Trained Words in Paragraph-Reading Task</td>
<td>Z = -2.856</td>
<td>Z = -1.541</td>
<td>Z = -0.459</td>
<td>Z = -0.459</td>
</tr>
<tr>
<td></td>
<td>p = .004*</td>
<td>p = .123</td>
<td>p = .646</td>
<td>p = .646</td>
</tr>
<tr>
<td>Untrained Words in Paragraph-Reading Task</td>
<td>Z = -2.480</td>
<td>Z = -1.362</td>
<td>Z = -1.007</td>
<td>Z = -0.889</td>
</tr>
</tbody>
</table>

* = significant at p < .05 level

These results demonstrate that the first-semester experimental group was able to generalize their improvement in accentedness to words that had not been specifically trained. This improvement could be found on both the word-reading and paragraph-reading tasks. The control group did not show any significant improvement in accentedness for any type of words on either task. Additionally, the fourth-semester experimental group did not show any significant differences at posttest on any task for any group of words. Therefore, preliminary analysis suggests that the pronunciation instruction program potentially aided elementary-level students generalize their improved accent to untrained words.

When comparing groups to each other based on performance on the trained vs. untrained words on each task at posttest only, a Mann-Whitney test revealed no significant differences between groups at posttest. Differences between the experimental and control groups were not significant for any group of words (trained or untrained) or any task (word- or paragraph-reading). Thus, although each group showed different amount of improvement on the tasks and
within word groups, training was not sufficient to make the experimental groups’ posttest performance significantly better than that of the control groups.

From the results in this section, a preliminary inference can be made that a pronunciation instruction program can help elementary-level students generalize their improvement in both comprehensibility and accent to words that were not trained.

4.4 Pronunciation Attitude Inventory

The revised version of Elliott’s (1995) Pronunciation Attitude Inventory (PAI) provides self-reported measurements of attitudes and motivation level to acquire pronunciation in an L2. Possible scores ranged from 13, representing a negative attitude towards pronunciation, to a maximum score of 65, representing a positive attitude towards pronunciation. Reponses to the three negatively worded statements on the questionnaire were reversed, so that a response of 5 always characterized a positive attitude and a 1 a negative one. Mean total scores were calculated for each group and are reported in Table 19 below.

Table 19

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean PAI Score (out of 65)</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Semester</td>
<td>49.00</td>
<td>7.71</td>
<td>38-60</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Semester Control</td>
<td>45.64</td>
<td>7.01</td>
<td>34-56</td>
</tr>
<tr>
<td>Fourth-Semester</td>
<td>52.10</td>
<td>7.61</td>
<td>44-61</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourth-Semester Control</td>
<td>51.80</td>
<td>5.17</td>
<td>36-60</td>
</tr>
</tbody>
</table>

The group with the highest score on the PAI was the fourth-semester experimental group, whose answers on the questionnaire average to 52.10 points out of 65. The participant with the lowest score on the PAI in this group still scored 44 out of 65, while the lowest-scoring
participant in every other group scored in the mid-thirties. Thus, the fourth-semester experimental group can be said to have the most positive attitudes about the value of pronunciation, as well as the highest motivation to improve their pronunciation in German. As measured by the PAI, students in the first-semester control group have the most negative attitudes towards pronunciation and a lesser desire to improve. Participants in this group scored an average of 45.64 points out of 65 through their answers. Results of a Mann-Whitney statistical test show that the differences between group mean scores on the PAI are not significant.

Results from the PAI are most valuable due to the insights they might give into why a certain participant group improved or did not improve. Investigating scores on the PAI can give clues as to the relationships between attitude about pronunciation and motivation to improve and actual performance and improvement over time. Thus PAI scores will be tested for significant correlations with accuracy on the perception test and ratings on the production test in section 1.5 below.

**4.5 Correlations & Relationships**

As the pre- and posttests involved so many different components, this section is devoted to reporting on the relationships between each component. Responses and scores from each of the four different components—the Pronunciation Attitude Inventory, the perception test, and the production test including both accentedness and comprehensibility ratings—were analyzed using Pearson correlations.
4.5.1 Perception Test Correlations

Correlations for the perception test scores were calculated as follows. Accuracy scores on the perception test at pretest and posttest were compared with native speaker judgments of accentedness and comprehensibility at both pre- and posttest via a Pearson’s Correlation. The amount of change (either improvement or a decline) on the perception test was also compared with the amount of change in accentedness and comprehensibility ratings.

Scores on the perception pretest were significantly correlated with native speaker ratings of accentedness on both the pretest ($r = -0.386$, $p = .027$) and posttest ($r = -0.369$, $p = .034$). The higher the participants scored on the perception test, the better the accentedness ratings they received. Scores on the perception test at posttest did not correlate significantly with accentedness ratings at either pre- or posttest, although the correlation did approach significance with the ratings at posttest ($r = -0.336$, $p = .056$), and most likely a greater number of participants would lead to significance. In this case, there does seem to be a relationship between perception skills and accentedness, and that this relationship was at play in for the participants in this study. Table 20 shows the correlations between the perception test accuracy and the accentedness ratings, and Figures 39 and 40 present these data in scatter plots.

---

8 The $r$ scores for these correlations are calculated as negative correlations, because lower ratings demonstrate greater improvement. The higher the accuracy on the perception test, the lower the rating, thus the negative $r$ number.
### Table 20

*Correlations between Perception Test Accuracy and Accentedness Ratings*

<table>
<thead>
<tr>
<th></th>
<th>Perception Pre</th>
<th>Perception Post</th>
<th>Accent Ratings Pre</th>
<th>Accent Ratings Post</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td>1</td>
<td>.635**</td>
<td>-.386*</td>
<td>-.369*</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.000</td>
<td>.027</td>
<td>.034</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>45</td>
<td>45</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td>1</td>
<td>-.213</td>
<td>-.336</td>
<td></td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.234</td>
<td>.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>33</td>
<td>33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

*Figure 39: Correlation between Perception Accuracy and Accentedness Ratings at Pretest*
Accuracy scores from the perception test were also tested for correlation with the comprehensibility ratings from the native speaker raters. Accuracy on the perception test was correlated with comprehensibility ratings at both pretest and posttest. Pre-training perception test scores correlated significantly with comprehensibility ratings at both pretest ($r = -.444, p = .002$) and posttest ($r = -.451, p = .002$). The same goes for post-training perception test accuracy; it correlates significantly with the comprehensibility ratings at pretest ($r = -.430, p = .003$) and posttest ($r = -.504, p < .001$). Therefore, the higher the accuracy on the perception test, the better comprehensibility ratings a participant received, whether at pretest or posttest. Table 21 outlines these results, and the scatter plots are shown in Figures 41 and 42 below.
Table 21

*Correlations between Perception Test Accuracy and Comprehensibility Ratings*

<table>
<thead>
<tr>
<th></th>
<th>Perception Pre</th>
<th>Perception Post</th>
<th>Comp. Rating Pre</th>
<th>Comp. Rating Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>.635**</td>
<td>-.444**</td>
<td>-.451**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>-.430**</td>
<td>-.504**</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.003</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>45</td>
<td>45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 41: Correlation between Perception Accuracy and Comprehensibility Ratings at Posttest
Both overall perception skills at pretest and the amount of change on the perception test were also measured against the amount of change in accentedness and comprehensibility ratings. It was found that the amount of change in perception test accuracy, whether positive or negative, was not significantly correlated with the amount of change in either accentedness or comprehensibility ratings. Tables 22 and 23 show the results of these two correlation analyses.

Figure 42: Correlation between Perception Accuracy and Comprehensibility Ratings at Posttest
Table 22

Correlations between Perception Test Accuracy at Pretest and Amount of Change in Accentedness and Comprehensibility Ratings

<table>
<thead>
<tr>
<th></th>
<th>Perception Pre</th>
<th>Amount of Change - Accentedness</th>
<th>Amount of Change - Comprehensibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>1</td>
<td>-.216</td>
</tr>
<tr>
<td>Perception Pre</td>
<td>Sig. (2-tailed)</td>
<td>.228</td>
<td>.260</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td>Amount of Change -</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.727**</td>
</tr>
<tr>
<td>Accentedness</td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Amount of Change – Comp.</td>
<td>Pearson Correlation</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 23

Correlations between Amount of Change in Perception Test Accuracy and Amount of Change in Accentedness and Comprehensibility Ratings

<table>
<thead>
<tr>
<th></th>
<th>Amount of Change - Perception</th>
<th>Amount of Change - Accentedness</th>
<th>Amount of Change - Comprehensibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.326</td>
</tr>
<tr>
<td>Amount of Change -</td>
<td>Sig. (2-tailed)</td>
<td>.060</td>
<td>.323</td>
</tr>
<tr>
<td>Perception</td>
<td>N</td>
<td>34</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.727**</td>
</tr>
<tr>
<td>Amount of Change -</td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Accentedness</td>
<td>N</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Amount of Change -</td>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

From these reports, one can see that participants’ accuracy rates on the perception test correlated significantly with accentedness and comprehensibility ratings. As accuracy on the perception test increased, accentedness and comprehensibility ratings became better. A
relationship between perception scores and accentedness ratings was not found significant at posttest, but it is believed that a stronger $n$ would have produced a significant relationship. These results suggest a strong link between perception skills and comprehensibility of speech, and a moderate relationship between perception skills and foreign accent. On the other hand, there does not seem to be a connection between perception skills and how much or little a participant was able to improve; perception skills alone do not seem to be enough to guarantee improvement.

4.5.2 PAI Correlations

Pearson's correlations were also performed with results from the PAI to gauge the influence of attitudes and motivation on participants' pre- and posttest perception and production scores. PAI scores were correlated with accuracy on the perception test, as well as with ratings of accentedness and comprehensibility.

To test the relationship between a participant's attitude toward learning and improving pronunciation and their perceptual skills, scores on the PAI were correlated with accuracy on the listening test. Pearson correlations demonstrated that perception test accuracy was significantly correlated with the results on the PAI questionnaire. This is true for pretest perception accuracy only, however ($r = -.327, p = .032$); the correlation between posttest perception accuracy and PAI results is not significant at posttest ($r = -.275, p = .074$). Table 24 presents the results from the correlations of perception accuracy and accentedness.
### Correlations between Perception Test Accuracy and Motivation Level

<table>
<thead>
<tr>
<th></th>
<th>Perception Pre</th>
<th>Perception Post</th>
<th>PAI Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception Pre</td>
<td>Pearson Correlation 1</td>
<td>.635***</td>
<td>.327*</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed) .000</td>
<td>.032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N 45</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Perception Post</td>
<td>Pearson Correlation 1</td>
<td>.275</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.074</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N 43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAI Total</td>
<td>Pearson Correlation</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The following scatter plots show the relationships between perception skills and PAI answers in graphic form.

**Figure 43: Correlation between Perception Accuracy at Pretest and Total Score on the PAI**
Mean scores on the PAI were also compared with accentedness and comprehensibility ratings on the production tests. Average PAI scores correlated significantly with accentedness at pretest ($r = -.383, p = .031$) and posttest ($r = -.437, p = .012$), as well as with comprehensibility ratings at both pretest ($r = -.541, p < .001$) and posttest ($r = -.536, p < .001$). Results from these correlations are presented in Table 25 below.
Table 25

Correlations between Motivation Level and Accentedness and Comprehensibility Ratings at Pre- and Posttest

<table>
<thead>
<tr>
<th></th>
<th>PAI Total Score</th>
<th>Accent Rating Pre</th>
<th>Accent Rating Post</th>
<th>Comp. Rating Pre</th>
<th>Comp. Rating Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>-.383*</td>
<td>-.437*</td>
<td>-.541**</td>
<td>-.536**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.031</td>
<td>.012</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>32</td>
<td>32</td>
<td>44</td>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>

| Accent Rating Pre | Pearson Correlation | 1 | .880** | .936** | .827** |
| Sig. (2-tailed)   | .000              | .000              | .000          |
| N                 | 34                | 34                | 34            |

| Accent Rating Post | Pearson Correlation | 1 | .863** | .929** |
| Sig. (2-tailed)    | .000              | .000              |
| N                 | 34                | 34                |

| Comp. Rating Pre  | Pearson Correlation | 1 | .909** |
| Sig. (2-tailed)   | .000              |
| N                 | 46                |

| Comp. Rating Post | Pearson Correlation | 1 |
| Sig. (2-tailed)   | .000              |
| N                 |                   |

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

These significant correlations suggest that there is a strong relationship between a participant’s attitudes toward the importance of pronunciation when learning German and their actual pronunciation performance. This seems particularly true for attitude towards pronunciation and comprehensibility of speech; the correlations between these two items were very strong and highly significant in this study.

Scores on the PAI were also correlated with amount of change (either improvement or decline), in order to see if attitude had an effect on the amount of improvement a participant was bound to make over time. Pearson correlations were calculated for PAI responses compared to the amount of change in accentedness and comprehensibility ratings and the perception test.
accuracy. However, no significant correlations were found. The total PAI score did not correlate with change in any skill from pretest to posttest. Results from this analysis can be found in Table 26.

Table 26

*Correlations between Motivation Level and Amount of Change in Perception Accuracy, Accentedness and Comprehensibility Ratings*

<table>
<thead>
<tr>
<th></th>
<th>PAI Total Score</th>
<th>Amount of Change - Perception</th>
<th>Amount of Change - Accentedness</th>
<th>Amount of Change - Comp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>-.049</td>
<td>-.121</td>
<td>-.254</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>44</td>
<td>32</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amount of Change - Perception</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>34</td>
<td>46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amount of Change - Accentedness</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>34</td>
<td>46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amount of Change - Comp.</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>34</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

In order to compare pronunciation attitudes with performance on the task for each group, average scores on the PAI were compared with accentedness and comprehensibility ratings for each individual group and proficiency level. PAI responses were significantly correlated with comprehensibility ratings on the pretest for the fourth-semester control group only ($r = -0.670, p = .034$), although the relationship approached significance for the first-semester experimental group as well ($r = -0.551, p = .051$). No significant correlations were found between comprehensibility ratings on the pretest or posttest and mean PAI scores for students in the first-semester control or fourth-semester experimental group.
This analysis was repeated with participants' accentedness ratings on the pretest and posttest. Only the first-semester experimental group showed a significant correlation between mean PAI scores and accentedness ratings at pretest ($r = -0.558, p = .048$). As with the comprehensibility ratings, there were no significant correlations between PAI scores and accentedness ratings on the pretest or posttest for students in the first-semester control or fourth-semester experimental group, but there were also no significant correlations found for participants in the fourth-semester control group.

From these results, we can assume that attitudes about pronunciation, as well as the desire to improve one's pronunciation ability, has an effect on actual pronunciation ability. Responses on the PAI correlated significantly with the pre- and posttest accentedness and comprehensibility ratings, but did not show a significant correlation with the amount of improvement. It seems that a positive attitude towards pronunciation can influence a learner's performance, but cannot predict whether he or she will show more improvement than someone with a negative attitude. Yet at the same time, the results are not clear-cut enough to suggest that a positive attitude towards learning pronunciation can lead to better comprehensibility or a more target-like accent, nor that a negative attitude can have the opposite effect. For example, it is interesting to note that while the fourth-semester experimental group demonstrated the highest average score on the PAI and therefore can be described as possessing the most positive overall attitudes towards pronunciation, participants in this group did not make any significant improvement on any task from pretest to posttest. Some nonsignificant improvements were made when looking at the data itself, as was done in section 4.2.2.3. Yet even among students who have previous study abroad experience and might be expected to have higher motivation levels, attitude does not seem to predict significant improvement. In fact, as a group these participants showed a decline in
comprehensibility at posttest, both in overall ratings as well as on each individual task within that posttest.

4.5.3 Accentedness and Comprehensibility Correlations

Lastly, correlations between accentedness and comprehensibility ratings were calculated on the pre- and posttests, as well as with the amount of change in ratings from Time 1 to Time 2.

There was a significant correlation between accentedness and comprehensibility at pre-test ($r = .936$, $p < .001$) and posttest ($r = .929$, $p < .001$). This means that on the production test, if a participant was assigned a good rating for accentedness, they most likely also received a good rating for comprehensibility. Such high $r$ values as seen here suggest a very strong relationship between a participant's accent and how easy or hard it is to understand what they are saying, as perceived by native speaker judges. Table 27 summarizes these correlations, while Figures 45 and 46 present them as scatter plots.

Table 27

<table>
<thead>
<tr>
<th></th>
<th>Accent Rating</th>
<th>Accent Rating</th>
<th>Comp. Rating</th>
<th>Comp. Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Accent</td>
<td>Pearson Correlation</td>
<td>.880**</td>
<td>.936**</td>
<td>.827**</td>
</tr>
<tr>
<td>Rating</td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Pre</td>
<td>N</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Accent</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.863**</td>
<td>.929**</td>
</tr>
<tr>
<td>Rating</td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Post</td>
<td>N</td>
<td>34</td>
<td>34</td>
<td>.909**</td>
</tr>
<tr>
<td>Comp.</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Rating</td>
<td>Sig. (2-tailed)</td>
<td>N</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>Pre</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Figure 45: Correlation between Accentedness and Comprehensibility Ratings at Pretest

The amount of change in accentedness and comprehensibility that each participant demonstrated from Time 1 to Time 2 was also compared via Pearson correlations. It was found that the amount of improvement or decline in accentedness and comprehensibility exhibited by participants in this study was significantly correlated ($r = .727, p < .001$).
Table 28

Correlations between Amount of Change in Accentedness and Comprehensibility Ratings

<table>
<thead>
<tr>
<th>Amount of Change - Accentedness Pearson Correlation</th>
<th>Amount of Change - Comprehensibility Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of Change - Accentedness Sig. (2-tailed)</td>
<td>Amount of Change - Comprehensibility Sig. (2-tailed)</td>
</tr>
<tr>
<td>Amount of Change – Comp. Pearson Correlation</td>
<td>Amount of Change – Comp. Sig. (2-tailed)</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table 28 presents the correlations of change between accentedness and comprehensibility, while Figure 47 presents the data in a scatter plot.

![Figure 47: Correlation between Amount of Change in Accentedness and Comprehensibility](image)

The relationship between accentedness and comprehensibility is very close in this particular study. Good accentedness ratings are strongly correlated with good comprehensibility ratings, indicating that a more native-like accent in German makes an L2 speaker easier to understand for the native speakers chosen as raters in this study.
In sum, perception test accuracy was significantly correlated with accentedness and comprehensibility ratings. Correlations also demonstrate relationships between pronunciation attitudes and perception skills, accentedness ratings, and comprehensibility ratings. The strongest correlations that were found were those between accentedness and comprehensibility ratings. These results suggest that the better an L2 learner’s perception skills, the more comprehensible their speech and the less severe their accent will be. In addition, positive attitudes toward pronunciation improvement can influence a learner’s ultimate L2 pronunciation performance. Also, the more native-like a learner’s accent is judged to be, the more comprehensible they will be to native speakers.

When looking at the amount of change that occurred between pretest and posttest for each participant, no significant correlations were found between improvement in accentedness or comprehensibility in either perception skills at pretest, the change in perception test accuracy or the level of motivation gauged via the PAI. It seems that how much a person improves or does not improve cannot be predicted by changes in perception skills or attitude alone. Yet changes in accentedness are strongly correlated with changes in comprehensibility, further demonstrating the strong relationship between these two factors.

4.6 Summary of Results
The previous chapter outlined the results of the present dissertation study. Results from the perception test were presented first, followed by those of the motivational questionnaire, each of the three separate tasks on the production test, and lastly, the connections and relationships between each of these different components were outlined. These results were presented in order to address the three research questions posed in Chapter Two.
Descriptive analyses and nonparametric statistical measures indicated that participants in the first-semester experimental group made the largest gains in both perception and production skills after training. The first-semester experimental group showed significant improvement from pre- to posttest on the perception test, whereas the first-semester control did not. Participants who received training in the first semester showed significant improvement in both comprehensibility and accentedness as rated by native German speakers, both on the production test as a whole, as well as on the three individual tasks (with the exception of comprehensibility ratings on the free speech task). First-semester control participants also showed significant improvement in comprehensibility on the production test as a whole and the word- and paragraph-reading tasks, but this improvement was not found among their accentedness ratings. Additionally, participants in the first-semester experimental group showed significant improvement the largest number of individual conditions that were trained, whereas the control group did not show as much improvement. The first-semester experimental group showed significant improvement in accentedness and comprehensibility on both words that were trained during the instruction unit and those that were not, for both the word- and paragraph-reading tasks.

Participants in both fourth-semester groups, on the other hand, showed the least amount of influence from training. Participants in the fourth-semester groups did not show any significant improvement on either the perception or production tests. Each group also only improved significant in one of the individual conditions targeted for training.

Pearson correlations demonstrate significant correlations between all of the different components of this study, yet significant correlations could not be found between improvement in accentedness or comprehensibility over time and perception skills at pretest, the change in
perception test accuracy, or motivation as recorded by the PAI. The lack of significant correlations where one might expect them may be due to a low participant number. Yet these results indicate that, standing alone, perception skills and motivation are not predictors for the amount of improvement a participant is likely to make in their production over time. It seems that explicit pronunciation instruction is what can lead to an improvement in production, especially among students who are just beginning their language learning program.

The next chapter discusses the implications of the findings described in this chapter.
Chapter 5: Discussion

The current study was undertaken in order to answer three research questions. The first research question inquired as to whether a pronunciation training intervention could be effective at helping elementary-level learners improve their comprehensibility and accentedness. The results of the experimental classroom pronunciation training unit described in this dissertation confirm that training can be effective in helping elementary German learners significantly improve their perception skills, comprehensibility, and accentedness. This was the case both on the experimental posttest as a whole, as well as on separate tasks and in individual conditions.

The second research question that inspired this study determined whether this intervention was more effective at the elementary or intermediate level. The improvements in comprehensibility and accentedness found in this study were more prominent among first-semester participants than among fourth-semester participants, suggesting that training may be more effective at elementary than intermediate levels of proficiency. Lastly, this study also aimed to determine the interactions between a learner’s attitudes about pronunciation and their pronunciation performance. It was found that attitude levels had an effect on pronunciation performance, although they were not able to predict the amount of improvement a certain learner was bound to make. The outcomes of this study and their implications for research and language learning will be discussed at length in the following section.

5.1 Targeted Conditions

First-semester German learners who received pronunciation training as part of their normal classroom activities were able to significantly improve their comprehensibility and accentedness in more individual segmental and suprasegmental training conditions than the learners who had not received any training. The beginning learners who received training were
able to significantly improve their comprehensibility in seven individual conditions and their accentedness in eight individual conditions. Students in their first semester who did not receive training also showed some significant improvements, but in fewer individual conditions than those who had received training. Untrained learners in the first semester improved in the same four individual conditions both in terms of comprehensibility and accentedness.

Training seems to have had a direct effect on participants in the early semesters of learning German. Participants in the experimental group made greater gains in comprehensibility and accentedness in the conditions that were specifically targeted for classroom training, while similar gains were not made by the control group. If the improvement found in both first semester groups were related to the same factors, then we would expect each group to show similar patterns of improvement in comprehensibility and accentedness in the individual conditions. This pattern is not found in the results; therefore, the improvement of the control group must be related to something other than training.

Elementary-level learners who took part in the pronunciation training program showed greater improvement in those conditions which are most essential to improving comprehensibility and communicating meaning. Critical among these is accurate placement of lexical stress, which is central to native speaker comprehension of L2 speech. Research has shown that nonnative deviations in expected lexical stress patterns can have drastic consequences for comprehensibility (Caspers, 2010; Field, 2005; Hirschfeld, 1994). Elementary-level learners who received training in this study, however, were better able to produce accurate lexical stress assignment in order to improve their comprehensibility and reduce their accentedness. Their improvement was demonstrated in five different condition and task-type combinations dealing with lexical stress: cognates and noncognates with stress on the second syllable on the word-
As mentioned in section 2.4.1, perhaps the most crucial lexical stress pattern that English learners of German must learn to produce accurately is that for cognates with second syllable stress. German and English have many such cognates which appear to be the same, yet differ in stress placement. Participants in this study who received training were better able to avoid the transfer of L1 stress patterns onto words such as Musik [muˈzi:k] or Student [ʃuˈdɛnt] that look and sound similar to words in English. Training gave them the tools they needed to place lexical stress on the second syllable of the German word more accurately. Trained participants were more successful in this task than the first-semester students who had not received training, who showed significant improvement of comprehensibility or accentedness in only one condition related to lexical stress assignment. Pronunciation training enabled beginning L2 learners of German with English as an L1 to better improve their comprehensibility and accentedness in an area that may be the most difficult stress pattern to acquire and that is also fundamental to the communication of meaning.

Results from the control group provide further evidence that the experimental participants’ increased ability to correctly apply lexical stress in German is attributable to training. Participants who did not receive training in the classroom only showed significant improvement of comprehensibility and accentedness in one lexical stress condition: noncognates with second syllable stress. As this is generally an uncommon stress pattern in German, all of the test items for this condition began with one of a set of German prefixes that never carry lexical stress, such as be-, emp-, ent-, er-, ge-, ver-, and zer-. These prefixes can never be stressed in German, and all words that begin with these prefixes must be stressed on the second
syllable. Perhaps the two most common of these prefixes are ge- and be-, which were both used in the test items for the noncognates with second syllable stress condition. Both first semester groups showed significant improvement in this condition, independent of training, which suggests that an “unstressed prefix” rule is relatively easy for first-semester German students to grasp and use, even after only one semester of language learning and without focused pronunciation training. Other lexical stress assignment rules are not acquired so easily, and it seems that training may be necessary for beginning learners to develop the ability to properly apply lexical stress in German words that do not use unstressed prefixes. This seems particularly true for more difficult stress patterns like cognates that apply stress to different syllables in English and German.

5.2 Improvement on Test Tasks

In-class pronunciation training also prepared beginning learners to improve their pronunciation abilities on more difficult production tasks such as paragraph reading and spontaneous speech. Students in the first-semester experimental group showed significant improvement in comprehensibility on the word- and paragraph-reading tasks and accentedness on all three test tasks, including free speech. Students in their first semester who did not receive training, however, improved their comprehensibility on the word- and paragraph-reading tasks only, and did not show significant improvement in accentedness on any test tasks. Students in the experimental group also significantly improved their comprehensibility and accentedness in nine individual conditions on the word-reading task and six on the paragraph-reading task. Control group students, on the other hand, significantly improved their comprehensibility and accentedness in six conditions on the word-reading task and in only two conditions on the
paragraph-reading task. Students who received training were better able to improve their pronunciation abilities on more difficult tasks than students who were not trained in-class.

In their seminal paper about classroom pronunciation training, Derwing et al. (1998) put forward the idea that different speech tasks have different levels of difficulty for L2 learners, and that these varying difficulty levels impact pronunciation abilities. They state that pronunciation accuracy typically worsens as task demands increase. For instance, when L2 learners must read words in isolation, fewer language processing resources are in demand than when producing utterances within longer contexts, and pronunciation does not suffer greatly. But when learners must produce longer utterances and spontaneous speech, their mental resources are more consumed with lexical, morphological, and syntactic concerns, and pronunciation tends to suffer as a result. In part, this is why these authors argue that suprasegmental training is so essential – it allows L2 learners to apply pronunciation knowledge developed during training even as language production tasks become increasingly difficult.

An experimental study measuring the impacts of task type on pronunciation improvement and ability by Liu (2011) supports these claims by Derwing et al. (1998). Liu (2011) provided training in pronunciation to Chinese learners of English and, as in the current study, measured improvement at three different discourse levels: word-reading, sentence-reading, and free speech. Results reveal that pronunciation improvement due to training was greatest on the word-reading task, and then gradually waned as participants progressed to sentence-reading and free speech tasks. The author explains that L2 learners can devote more time and mental resources to self-monitoring their pronunciation when they are focusing on words in isolation. Yet as task demands increase and L2 learners must begin producing their own original utterances, their attention must be divided among a number of language processes related to lexical access and
syntactic formulation and organization. In free speech situations, L2 learners have the fewest resources to expend on pronunciation, and they are not always capable of self-monitoring their pronunciation accuracy. Therefore, their overall pronunciation ability appears to decline, and they demonstrate less improvement as a result of training.

From the explanations in these two studies, one might expect that the beginning L2 learners in the current study would demonstrate fewer improvements in pronunciation as they performed the language tasks with high demands. Yet participants in this study who received training were rated both as more comprehensible and less accented on all tasks. Students who received training significantly improved their comprehensibility on the paragraph-reading task and also showed nonsignificant improvement on the most difficult task out of the three they completed – the free speech task. Nine (64.3%) of the participants in the first-semester experimental group were able to improve their comprehensibility by an average of 0.44 points on the free speech task from pretest to posttest. Only four students in this group (28.6%) were rated as less comprehensible at posttest than at pretest. Participants in the first-semester experimental group also significantly improved their accentedness ratings on all tasks, including the most difficult free speech task. They were also able to significantly improve their comprehensibility and accentedness in three times as many individual conditions on the more difficult paragraph-reading task than the control participants did.

The combination of these results suggests that the beginning learners who received training in this study were still able to improve their comprehensibility despite increasingly complicated task demands. Therefore, it can be argued that in-class pronunciation training allowed the participants to employ their knowledge from training and to improve their pronunciation abilities even as more of their mental resources were consumed with other
language processes. This is striking when considering that focusing on a singular language form such as pronunciation while communicating in the L2 has been argued to be exceedingly difficult for beginning learners who still lack sufficient L2 experience (Williams, 1999).

Once again, results from the control group support the suggestion that training was responsible for the experimental group’s gains. Control participants did not show significant improvement of accentedness on the more difficult paragraph-reading and free speech tasks, and their comprehensibility was actually rated as worse at posttest than at pretest. A total of 45% (n=5) of control participants received poorer comprehensibility ratings on the free speech task at posttest. Thus, although students in the control group were able to show improvement in the comprehensibility of their speech without training when reading words and paragraphs, they were unable to do so on the most difficult task. As the task demands increased and their mental resources were engaged by other burdens, the pronunciation abilities of the control participants both individually and as a group failed to show the same level of improvement as the experimental group had.

Lastly, students in the first-semester experimental group were able to extend their improvement in comprehensibility and accentedness to new words and speech situations to a greater extent than the control group could. This is important to note, as the ultimate goal of any pronunciation unit is that L2 learners will be able to extend what they have learned beyond the words used for training. Students in the first-semester experimental group were able to generalize their pronunciation abilities to untrained words and significantly improved their accentedness in free speech situations. These results suggests that they were able to extend what they had learned during training to a variety of topics and speech situations in their L2. Thus, first-semester participants who received training were, in fact, able to apply the lessons from
their in-class pronunciation activities to new words and new situations, whereas those participants who had not received training were less able to do so.

5.3 Improvement in First-Semester Groups

As was previously discussed, the first-semester group that received training made larger gains in comprehensibility and accentedness in more areas vital to communication and on more difficult tasks. Yet both the experimental and control groups show significant improvement in comprehensibility and overall accentedness ratings at posttest. In training studies, it is desirable that the group that does not receive training should not make similar gains as the experimental group does. Therefore, a deeper investigation into the control group is necessary in order to determine whether the improvements made by the experimental group were truly an effect of training. Upon further analysis, several reasons for the improvement of both groups become apparent.

The most likely explanation for the improvement shown by the control group is that of language experience. Twenty-one of the 25 first-semester participants (84%) had no classroom exposure to the German language prior to enrollment in the beginner course, and began training with only four weeks’ knowledge of the language and its sound system. They had most likely not yet formed any solid impressions about the German language or German pronunciation, but were exposed to a great variety of new input and continual four-skills language practice as they progressed through their language lessons over the course of the semester. Research from SLA demonstrates that a combination of input and practice leads to great gains in learners’ ability to produce and understand their L2, particularly for those at the elementary levels (Gass & Selinker, 2008; Krashen, 1987; Trahey and White, 1993). Due to the great influx of input and language practice in the classroom, supplemented by the extra practice provided by the partner recordings
that were part of this project, the elementary-level students in this study were also likely to make large strides in their language development.

This research from SLA can also be extended to pronunciation development. It is likely that the greatest gains in pronunciation also occur at the beginning stages of classroom learning for all learners, regardless of the presence of classroom training. In fact, a similar phenomenon has been found among L2 learners in immersion environments. When L2 learners are newly immersed in an L2 environment, their most noticeable and significant gains in pronunciation develop very rapidly. After this rapid improvement within the first few months of immersion, pronunciation abilities tend to stagnate (Derwing & Munro, 2013; Flege, 1988; Flege, Munro & Skelton, 1992; Munro & Derwing, 2008). For instance, Derwing and Munro (2013) measured the pronunciation skills of two groups of immigrants living in Canada over a seven-year period. Results of this longitudinal study reveal despite seven years of immersion and interaction with native speakers of Canadian English, native speakers of Slavic and Mandarin languages did not demonstrate significant improvement in their accent after the first two years of immersion.

While one must keep in mind that immersion and classroom learning environments exhibit extreme differences and cannot be directly compared, it is possible that learners in a classroom environment will display similar patterns. If this were the case, increased exposure to L2 input, concentrated practice with the language, and the subsequent changes in general proficiency and ability may lead to improvement that is similar to that made by learners in the first few months of naturalistic L2 immersion. Just as the pronunciation abilities of learners in L2 immersion environments tend to stabilize after the early stages, L2 learners in the classroom may begin to plateau after the first semester. It may not be until some point after the conclusion
of these immediate gains in the first semester that the process of phonological fossilization begins (Elliott, 1995; Maldonado, 1994).

Although the significant gains experienced by both first-semester groups is most likely due to the fact that all participants were in the beginning stages of language learning, additional possibilities should be considered. With only eight weeks between the pre- and posttests, it is possible that test effects were present in the form of participants retaining some of the information from the pretest and using it to their advantage in the posttest. In this instance, significant improvement by both groups would be expected on all aspects of the posttest. This is not the case, however, as the control group was not able to improve on all aspects, and actually declined in their comprehensibility accuracy in spontaneous speech from Time 1 to Time 2. We would also expect the participants in the control group to pattern more like the experimental participants, if test effects were an issue. The control group participants also were not able to improve in the same areas that the experimental group did, further closing the case on the possibility of test effects in this study.

Another possibility for the improvement exhibited by the control group is that these participants had more practice with the language, either through a more in-depth language learning history, study abroad experience, or more daily use of German. They may also have had more interaction with native and advanced L2 speakers. The Language Background Questionnaire provides insight into these possibilities. Participants in the first-semester experimental group had begun learning German at an average age of 19.0, while those in the control group had begun at the average age of 18.3. Only two participants in each group had previous language experience with German before their first-semester course began. One difference that does emerge relates to study abroad experience. The same number of participants
(n=2) from each group had spent time studying abroad in a German-speaking country. The students in the control group had spent an average length abroad that was double (n=6 weeks) that of the length for the experimental section (n=3 weeks).

Contact with native German speakers and daily use of the German language was also measured on the LBQ. Six participants in the experimental group and just one in the control group reported regular contact with native German speakers. One participant in each group reported daily contact with a native German speaker, while the remaining participants in the experimental group reported monthly or yearly contact with their native German friends and acquaintances. It is not clear how much of this contact takes place in German and how much in English, although some clues can be gathered from questions estimating daily language use on the LBQ.

Among first-semester experimental participants, between 53.8 and 69.2% (n=7-9) reported that they spend between 1-30 minutes each day reading, writing, listening and speaking German. Between 15.4 and 30.8% (n=2-4) report using the different German skills between 30-60 minutes each day. Up to 23.1% of participants (n=3) reported not doing any skill practice outside of class, and up to 7.1% (n=1) reported using German for over an hour each day. The skills most commonly practiced outside of class time were writing and reading.

The amount of daily language use among control participants painted a slightly different picture. Participants in this group were less consistent in their daily language use. Most participants in this group reported spending 30-60 minutes each day practicing writing (45.5%, n=5) and reading (54.6%, n=6) outside of class time, but reported spending 1-30 minutes each day practicing speaking (45.5%, n=5) and listening (36.4%, n=4) skills. Up to 27.3% of participants (n=3) reported spending no time using their German skills outside of class time, and
up to 9.1% (n=1) reported using more than an hour of German outside of class. The skills most 
commonly practiced outside of class were reading, writing and speaking. Participants in the 
control group did, on the whole, report more daily use of German skills outside of the classroom 
than participants in the experimental group, which may have contributed to their significantly 
 Improved German pronunciation skills.

Interestingly, the participants in the experimental group who reporting having daily 
contact with German friends and acquaintances were the ones who reported either no outside 
practice of their German skills, or between 1-30 minutes per day. The one participant in the 
control group who reported contact with native speakers reported spending between one and two 
hours each day listening to German outside of class and between 1-30 minutes speaking German. 
According to self-report, this person received more exposure to the German language, possibly 
through contact with their native speaker friend. However, it is still not clear that this person 
was actively producing German during conversations with their friend. Thus, it can be assumed 
that due to their lower proficiency level, participants’ contact with native speakers was not 
conducted entirely in German, and further, that the significant improvements made by the control 
group were unrelated to contact with native German speakers.

One last possibility as to why the first-semester control group was also able to 
significantly improve their comprehensibility and overall accentedness over time is more 
anecdotal in nature. As each of these groups consisted of an entire intact class and was not 
randomly assigned from a participant pool, classroom dynamics and personalities may have 
played a role. Claims of this type are not experimentally measurable, but are valid options that 
can be supported by any language instructor. The instructor of the first-semester classes in this 
study mentioned repeatedly that students in the control group seemed outwardly more excited.
about learning German and had a more interactive and friendly classroom dynamic. This may have influenced their abilities to learn pronunciation, as they knew they were part of a pronunciation study in their course.

Of all the possible reasons for the significant improvement of the first-semester control group in some aspects of their pronunciation abilities, the most likely one is that all first-semester students are going through drastic changes in general language and pronunciation ability due to their increased exposure to L2 input and practice. This is supported by the revelation that the control group participants also reported spending more time outside of class practicing their German skills than the experimental group participants did. Yet even though the control group made some pronunciation gains and it is possible that the majority of L2 learners in their first-semester will show similar improvements, it does not mean that training was ineffectual. One must keep in mind that students who did receive pronunciation training exhibited superior results in a greater variety of areas than the control group. Students in the first-semester experimental group showed greater improvement on the targeted pronunciation aspects, especially in areas most important for the communication of meaning, and on more difficult tasks. Therefore, we can see that training did have a positive effect on beginning learners’ pronunciation abilities, and gave them a boost in ability that the control participants lacked.

Results from the perception test drive this point home, as they most clearly demonstrate positive results as a result of pronunciation instruction. Students in their first semester who received training were the only group of participants that was able to significantly improve their accuracy on the perception test from Time 1 to Time 2. Students who did not receive training did not show significant improvement. In addition, the majority of the nine students (66.7%, n = 6) who showed the greatest gains over time had received training.
The goal for the perception test was determining the differences between German words pronounced with a native-like German accent or with a typical American accent. Raising students’ awareness and helping them learn to attend to the relevant forms was one of the goals of this pronunciation unit, as research in SLA proposes that a first step to learning is that students must be aware of the fact that something needs to be learned (Gass & Selinker, 2008; p. 346). This goal was met in this study, and training provided experimental participants with a way to attend to and perceive the differences in German and American English sounds and stress patterns to a greater extent than a semester of learning German alone did.

Testing and training students’ L2 perception skills is an important first step in improving pronunciation skills. Chun (2002) proposes that all pronunciation training units should first begin with a sensitization and active listening stage. This stage allows L2 learners to notice the relevance of the target pronunciation features and how they should sound, and develops a good base for the production training that comes later. Furthermore, attending to phonetic contrasts and new segments in the L2 may help learners acquire the necessary perceptual skills that are often cited as necessary for proper production. One such model, Flege’s (1995) Speech Learning Model (SLM), deals specifically with adult L2 learners’ speech perception and its links to the production of speech. A key tenet of the SLM, and one ultimately driving pronunciation training studies, is that adult L2 learners are physically able to make the new sounds and can learn how to produce them with increased proficiency and experience. The reason for a foreign accent in the L2 lies in perceptual problems rather than in physiological limitations. The model posits that within a nonnative language, unfamiliar sounds are processed through the filter of the native language and mapped onto already existing L1 phonological categories for decoding. This model requires accurate phonological perception as a prerequisite for accurate production,
although the two skills do not necessarily develop in parallel. The ability to discriminate between L1 and L2 sounds does not guarantee the ability to produce them accurately, unless the creation of a new L2 phonological category has occurred. The SLM predicts target-like production only for those learners who can categorize unfamiliar L2 segments as separate from L1 sounds and have established new categories for them (Flege, 2007).

Participants in the first-semester group who did receive training were given a boost in their perceptual abilities. Training helped them to become more accurate at perceiving differences between German and American English sounds, which was an ability that the participants in the control group did not develop over the course of their first semester. While all first-semester participants showed improvement in production, only those in the experimental group were able to improve their perception. This is a crucial fact when working within the SLM. The SLM states that accurate perception skills precede accurate production skills, and in fact, perception accuracy was significantly correlated with comprehensibility and accentedness ratings in this study (see section 4.5.1). Thus, we can predict that participants who received training are primed to improve further than the results of this study can reveal. Their perception skills have increased and they may be on their way to creating new categories that the participants who did not receive training do not seem to be developing. The participants in the control group may have shown some overall improvement in production due to the influx of input and language practice that they were receiving, but their phonological categories may not develop until much later. Bohn and Flege (1992) suggest that L2 phonological categories are only formed when L2 learners reach a certain stage of language experience. The first-semester control participants in this study have demonstrated that one semester of learning German in class is not enough to push them beyond this threshold and improve their perception skills.
Training seems to be the key that provides first-semester students with an added level of language experience necessary to improve their perception skills, which is a crucial first step in the formation of L2 phonological categories. It is possible that even greater separation of skills would appear between the two groups at a later point in time, as the control group passed the point where their initial improvements take place and the experimental group gets more time to fully develop their L2 categories and improve their production skills further.

The perception test results, as well as the various improvements in production demonstrated by the first-semester experimental group, provide excellent support to the argument that pronunciation training positively affected first-semester L2 learners in ways that the students who did not receive training could not keep pace with. Training has provided the first-semester students in this study with an added boost in their perception and pronunciation skills such that they were able to become more comprehensible and less accented in their L2 productions.

5.4 (Lack of) Improvement in Fourth-Semester Groups

Overall, the pronunciation training unit produced a multitude of positive effects for learners who were in their first semesters of learning German, but, somewhat surprisingly, training did not seem to affect participant performance for learners at intermediate levels of proficiency. Fourth-semester participants who received training did not significantly improve their comprehensibility or accentedness from pretest to posttest, either on the test as a whole or on any separate task. They did show a significant improvement in these two pronunciation skills when the results were divided into individual pronunciation conditions, but this was only the case for one condition. A further analysis of the results showed that the amount of previous language experience, motivation and previous study abroad experience all seem to have had a
connection with pronunciation improvement. A combination of training and higher motivation levels led to greater gains in perception, comprehensibility and accentedness. In addition, those with longer periods of German language study, whether abroad in a German-speaking country or in the United States, were more likely to show improved pronunciation skills after training than those who had not. It is important to keep in mind, however, that none of these criteria are a guarantee for improvement in pronunciation, even when combined with in-class pronunciation training. Thus, these statistical and post-hoc analyses do little to clarify the effects of training among fourth-semester participants on their own.

There are a number of other possible explanations for the fact that the intermediate students did not seem to respond to in-class pronunciation training when the first-semester students made such large gains. For instance, the two fourth-semester groups were much smaller than the first-semester groups – each group began with only ten participants because of much smaller enrollments. In addition, they show remarkable differences in pronunciation when compared directly with first-semester students, despite still being far from native-like. Many of these participants were rated at ceiling and their data were removed from further analysis. Thus, by the time any statistical or individual analyses were completed, the fourth-semester groups could be as small as five participants. Such small participant numbers makes it harder to see any prominent trends in performance. It is possible with larger intermediate groups, more consistent trends in pronunciation improvement or larger impacts of previous experience and motivation could have been found.

The second possible explanation for a lack of improvement by the fourth-semester experimental group lies within the training unit itself. The exact same training unit was implemented at the elementary and intermediate levels for direct comparison to be possible, but
the unit was designed primarily around the needs of beginning learners. Unit topics and vocabulary items were simple enough for beginners to complete, and it stands to reason that the training program might not have been challenging enough to keep learners on-task and engaged. If fourth-semester students determined that the pronunciation exercises were too easy for them or even insulting to their level of ability, it would have been difficult for even the best teacher to keep them motivated. In his Input Hypothesis, Krashen (1987) suggests that for optimal effect on language learning, input should be slightly beyond students’ current ability level. The Input Hypothesis also makes clear that input that is too far beyond a learner’s current ability level would not be useful for learning. Perhaps the intermediate students in this study did not benefit from the words and structures used in these pronunciation exercises because they were too simple for the level of ability at which they currently found themselves. Unfortunately, due to the nature of the comparison needed for this study, it was deemed most crucial that the elementary learners would be able to understand and use the vocabulary and sentences that were being practiced during pronunciation exercises, in order to make the unit as meaningful and relevant as possible for all learners.

In line with the fourth-semester participants’ attitudes about the pronunciation unit, another possibility is that these participants simply did not view pronunciation as important, or did not view their own pronunciation as bad enough to need intervention. While no direct measure of this was taken in the Language Background Questionnaire or the Pronunciation Attitude Inventory, students may have viewed their pronunciation skills as already very good. Similar to the common refrain from students learning L2 grammar that if they say something wrong, other speakers will understand them anyway, the fourth-semester participants may have believed that their pronunciation was good enough as is. After all, they had been successfully
communicating with teachers and their peers at their current level of pronunciation skill for many semesters. They may have been of the opinion that someone would have corrected their pronunciation before the fourth semester if there had been any drastic problems with it.

One final possibility to consider is the risk of phonological fossilization. Students in the intermediate levels may have already passed the point where the largest gains in pronunciation ability are possible (Derwing and Munro, 2013; Flege, 1988; Flege, Munro & Skelton, 1992; Munro & Derwing, 2008) and may have begun the process of phonological fossilization. The participants in the fourth-semester group, either as a group or individually, did not improve to the same extent as first-semester students on many of the pronunciation skills. This suggests that while improvement at the intermediate levels is far from impossible, it may be more difficult to achieve. Elliott (1995) suggests that even after receiving training, students may have great difficulties reversing years of ingrained pronunciation habits that were developed through their own errors and reinforced by the errors of their peers. This would make the task of pronunciation improvement for the intermediate students in this study a more daunting task than that faced by the beginning learners, who have yet to begin the process of fossilization.

Closely related to the idea of phonological fossilization is research by Counselman (2010) that proposes that reversing pronunciation errors may be easier for words that were learned recently, rather than for words that were learned early on in an L2 learner’s experience. All of the test items in this study came from the first two chapters of the first-semester textbook. Beginning learners had only just learned these words as pronunciation training began; therefore altering their pronunciation of these words should not have posed as much difficulty for them. But if, as Counselman (2010) asserts, it truly is harder to change the pronunciation of vocabulary items learned in the first few semesters once students have progressed to higher levels of
proficiency, this would mean that the fourth-semester participants would have the greatest
difficulty improving the exact items chosen as target stimuli in this study. In addition to possibly
having already passed the threshold for beginning phonological fossilization, the test items in
this study were most likely learned by the intermediate participants at least a year and a half
before their training occurred. If vocabulary words that were newer for students at the
intermediate levels had been included in the test stimuli and in-class pronunciation activities,
more positive patterns of improvement may have been found among fourth-semester
participants. Yet as they stand, results suggest that participants in the intermediate levels had
more difficulty changing their pronunciation patterns than those in the elementary levels.
Although more research on this topic is necessary, the results from this study lend themselves to
the argument that pronunciation training may be less effective for intermediate learners, and that
pronunciation habits developed over years of study may be harder to break.

5.5 Effects of Motivation on Pronunciation
A secondary goal of this paper was to measure the influence of motivation and student
attitudes on pronunciation improvement. In fact, high scores on the PAI, signaling more positive
attitudes towards learning German pronunciation, were significantly correlated with higher
perception test accuracy, lower accentedness, and higher comprehensibility. Students in this
study who were more motivated to improve their pronunciation did show better performance
than those who showed less motivation. Importantly, however, motivation and attitude levels did
not significantly correlate with either the amount of positive or negative change that a participant
underwent. Students who wanted to improve the most and had the most positive attitude toward
improving did not necessarily truly improve the most. These results mirror those of Elliott
(1995), who found that scores on the PAI were a significant predictor of ultimate performance
but not how much or little a student would improve. The implication of this is that although motivation and attitude have an influence on pronunciation accuracy, some other factor is at work in determining the amount of improvement a particular student will display over time. In line with Elliott (1995), I suggest that this was the training program implemented in the first- and fourth-semester classrooms. It seems that without training, students who have positive attitudes towards L2 pronunciation will be able to perform well, but it is the explicit training program that can predict how much a given student will be able to improve.

5.6 Implications for Classroom Pronunciation Instruction

The results of this study lead to a number of practical implications for German classrooms; these will be addressed in this section.

5.6.1 When should we train German learners?

The in-class pronunciation training unit designed for this study had an effect on first-semester learner’s ability to improve their perceptual accuracy and comprehensibility and reduce their accentedness. The training unit improved experimental participants’ perception skills, overall comprehensibility, and accentedness to a far greater extent than that of untrained participants, such that they were able to improve these abilities on every task and in more conditions. In addition, training helped beginners develop more quickly on more difficult language tasks, in areas that are most important for communication and comprehensibility and with new words and speech situations. It can be concluded, then, that pronunciation training most definitely has a place in elementary-level classrooms. Training provides students with a
boost that helps them improve their pronunciation abilities in ways that typical four-skills language practice does not.

Beginning pronunciation instruction in the first few semesters may have additional benefits for learners, beyond increasing accuracy and performance (Counselman, 2010; Elliott, 1995; Eskenazi, 1999; Hardison, 2004). One commonly-cited benefit is that pronunciation training in the first few semesters may help prevent phonological fossilization. While this cannot be revealed with the data at hand, fossilization may have had an effect on the lack of improvement by fourth-semester learners. Intermediate learners did not show significant improvement from pretest to posttest on any area, independently of having received training or not. This may have been an effect of fossilization, as very little gains were made by fourth-semester learners over time. Learners in their first semester were most likely too early in their language development to have fossilized any pronunciation habits, and were more likely to improve through training. They were forming proper pronunciation habits from the beginning onward, instead of trying to reverse years of incorrect pronunciation habits and reinforcement of those habits through their own and their peers’ utterances. Therefore, pronunciation training should be included in beginner classrooms so that students can develop their pronunciation skills as they are learning the language itself. And although phonological fossilization most likely begins at a later point than the first-semester learners in this study have reached, it is possible that providing training in the first semesters may be a way to counteract or delay this process.

In addition to helping improve pronunciation before phonological fossilization sets in, the training provided in this study may have benefited students because it helped them to focus their attention on the relevant forms. Attention seems particularly important for acquiring accurate pronunciation in an L2 (Chun, 2002; Pennington and Ellis, 2000; Ramírez Verdugo, 2006).
Providing students in the beginning levels with pronunciation training that works to focus their attention can provide a number of potential benefits beyond those found in this study. Teaching students to attend to pronunciation may cause them to focus on the pronunciation of their teachers and native speakers more than students who have not received training may do. In time, this focused attention may lead to improvement of pronunciation aspects that had never been trained, as found by Hardison (2004). In this study, a training unit in L2 intonation patterns led to improvement in individual segments as well, despite the fact that the students never received segmental training. She concludes that attending to one aspect of pronunciation can help students to focus their attention on untrained elements of the L2 as well. Attending to pronunciation may also make it easier for beginning learners to understand the differences between their own speech and that of more advanced or native speakers, something which is necessary for improvement but very difficult for L2 speakers to do (Derwing & Rossiter, 2002; Trahey & White, 1993).

Focusing L2 learners’ attention on relevant forms may have a greater effect in the early stages of language learning. Gass, Svetics and Lemelin (2003) have found that L2 learners in beginning levels of proficiency are the most likely to benefit from some sort of focused attention training in the classroom. Their study was conducted with English L2 learners of Italian, but these results can easily be extended to pronunciation more generally, and this study more specifically. The beginning learners in this study were the ones who made the greatest gains, and the effects of training did not seem to extend to the fourth-semester learners in this study. Thus, it seems likely that pronunciation training is more effective for beginners because focusing learner attention early on has greater benefits than waiting until higher levels of proficiency.
Overall, results suggest that pronunciation training should have a place in beginner German language classrooms. Providing training to beginners allows them to make gains more quickly than their peers, and may even provide them with the tools to develop their pronunciation skills further with even more training or experience. The possibilities for long-term effects of pronunciation instruction provided in the first semester are great. As mentioned before, focusing learners’ attention in the first semesters seems to be most beneficial for L2 learning in general and pronunciation learning in particular. This consciousness raising may help beginning learners to continue to develop as they progress through their language learning sequence, once they get used to listening to other speakers and noticing the differences between their own pronunciation and those of others. In addition, generalization effects were present in this study, and this ability may also help students to use what they have learned in the classroom and extend it to the new words they learn in the future. If students are able to do this, then their pronunciation skills should only keep increasing as they continue to study the German language. The potential for beneficial long term effects is yet another reason, along with possibly avoiding fossilization and the importance of focused attention on form for beginning learners, why pronunciation instruction should be incorporated into elementary-level classrooms.

5.6.2 What should be trained?
After establishing that pronunciation training should be provided to beginning learners as early in their language learning process as possible, a new question arises: What is the best aspect of pronunciation to train for maximal results? This concern can be addressed via a combination of previous research and the results in this study.

The results of this study demonstrate that lexical stress should be a major part of any pronunciation training unit introduced in the basic language classroom. Not only is accurate
lexical stress placement vital for interlocutor comprehension of L2 speech (Caspers, 2010; Field, 2005), it was also the aspect of pronunciation with the most successful results in this study.

Most of the conditions that showed significant improvement from Time 1 to Time 2 were in the realm of lexical stress assignment. Adjusting their pronunciation of lexical stress seems to be the easiest thing for students to accomplish. This is logical, as stress placement is lexically-based and meaning-driven, and because the realizations of lexical stress are very similar in English and German. The fact that lexical stress was the condition that demonstrated the most significant improvements in comprehensibility and accentedness as a result of training supports Firth’s (1992) claims that classroom training in lexical stress produces the greatest gains for beginners. This combines well with Chun’s (2002) recommendation that any pronunciation training unit begin with lexical stress activities; it seems that students learning to adjust their lexical stress assignment in German will be able to notice their own improvement in this area and become motivated to continue.

In terms of segmental aspects of German that should be focused on for training, results from this study can provide us with some insights when analyzed via Flege’s (1995) Speech Learning Model. The SLM predicts that categorization of L2 sounds may occur in one of the following ways: assimilation or dissimilation. Instances of category dissimilation are the most easily remedied by learners, as the solution is the formation of a brand new category for an L2 sound that does not occur in the L1. This occurs most easily for L2 sounds that are perceived as being the most phonetically and acoustically different from the sounds already existing in the L1. This study contained four new L2 sounds that do not exist in the English language—the voiceless fricatives [ç] and [x], and the rhotics [ʁ] and [ר]. As these sounds are dissimilar to any
sounds in English, the SLM predicts that these segments will be easier for L2 learners to perceive and, in turn, produce accurately.

Category assimilation, on the other hand, occurs when learners persist in perceiving L2 sounds as instances of an L1 category, and formation of a new phonological category is thus blocked. This type of categorization operates most often when L2 speech sounds are perceptually close, but not acoustically identical, to those in the L1. Learners initially produce the L2 speech sound as if it were the corresponding L1 sound, but a merged L1-L2 category may be formed over time. This category will be used to process and produce both the L1 and the L2 sounds, in a way fundamentally different than that of native speakers. Two segments in this study fall under Category Assimilation—the two vowels, long [e:] and [o:]. Both English and German have these vowels, yet they are not produced in the same way; therefore L2 learners must work harder to realize the differences and create new L2 categories for them.

When looking at the segmental conditions trained in this study, it seems that participants who received training had an easier time improving their comprehensibility and accentedness of new sounds as compared to similar sounds. First-semester experimental participants were able to significantly improve both their comprehensibility and accentedness of three new sounds on the word- and paragraph-reading tasks: [v̥], [ç], and [x̥]. The fourth-semester experimental group only improved their accentedness ratings over time in one individual condition, and this is for [ʁ̥], which is considered by the SLM to be a new sound. The first-semester experimental group also demonstrated significant improvement in accentedness for the similar sounds [eː] and [oː], but did not receive improved ratings of comprehensibility in these conditions.

Taken together, the above results suggest that training was most effective for the new sounds that students had to learn, both in terms of likelihood for improvement and influence on
native speaker ratings of accentedness and comprehensibility. A successful pronunciation training unit can address these new segments, like voiceless fricatives and the /r/ sounds, and expect improvement in comprehensibility and accentedness for language learners. This is not to say that providing training for similar sounds is unsuccessful, but the gains evidenced in this study are not as strong for these sounds. The participants did not demonstrate much improvement in their production of these sounds, which is especially striking when one notes that the posttest was completed directly after the training in [e:] and [o:] had concluded. It is true that students did not have as much time to practice these similar sounds as they had had with the sounds trained earlier in the semester, but should have been able to improve somewhat during training.

One can conclude from this, other pronunciation training studies, and Firth’s (1992) recommendations, that instructors should train those things which beginning learners can most hope to accomplish. According to the SLM, it would be more difficult to accurately produce similar sounds like [e:] and [o:]. This study demonstrates that a two-week pronunciation training module does not seem to be enough time for beginning and intermediate learners to develop their perception skills and create new L2 categories for these two segments. In addition, monophthongization of these two vowels is not contrastive in German; if one pronounces them in an American-like manner, it does not greatly interfere with comprehensibility by changing the meaning of the word. This proposal is supported by the native speaker ratings in this study; the participants who did show improved production of the segments [e:] and [o:] were perceived as being less accented but not more comprehensible. It is possible that in addition to the L2 learners, the native speaker raters also judged these two phonemes as the same across English and German, which would have less of an impact on ratings of comprehensibility as opposed to
accentedness. Previous research on pronunciation instruction in the classroom by Osburg (1999) and Murphy (1991) emphasizes the importance of including meaning in any pronunciation instruction unit. L2 learners need to understand the connections between the sounds and features they are learning and their uses for communication, otherwise nonnative speakers may either ignore new sounds or classify them as bad exemplars of a sound they already know from their L1 (Flege, 1995; Osburg, 1999). Acquisition of these forms must be prompted by an understanding of their importance for distinguishing meaning within the L2.

Thus, it would seem to be more beneficial to instructors and L2 German students to focus practice on lexical stress assignment and those segments which do not exist in English. Such a combination would help learners be able to communicate most effectively in their L2 German. The new sounds trained in this study, [ç] and [x], [ɐ] and [ʁ], were the segments most responsive to training. Focusing training the combination of lexical stress and these segments will not only help students become more comprehensible and help them to avoid confusion in their communications by reducing their accent, but will also boost their confidence about their ability to improve their pronunciation instead of frustrating them.

5.6.3 Addressing Instructor Concerns

After choosing which segments or prosodic features to train, the next concern for many instructors is how to actually introduce pronunciation activities into the classroom. The current study was designed to address these concerns. The pronunciation unit was split into four two-week modules, each of which focused on pronunciation training for just ten minutes per class period. This format is well-suited to a elementary classroom because it does not take too much time away from other important course activities, such as those practicing the other four skills, vocabulary or grammar. It is also useful for instructors, as it was intended for use as a warm-up
activity. Students can feel comfortable coming in each day and knowing that they will have a warm-up pronunciation activity each morning.

Another important consideration for instructors who wish to introduce pronunciation into their own basic language classroom is the coordination of activities with appropriate vocabulary, grammar and topics. All vocabulary words chosen for training, as well as activity content, must be coordinated with the topics covered in the course textbook and other classroom materials. For instance, the training topics chosen for each of the modules in this unit coincided with the vocabulary and themes of the chapter the first-semester learners were working with at the time. This is done so that pronunciation activities can also serve another purpose: vocabulary and grammar review. Vocabulary and grammatical structures should be practiced alongside pronunciation skills within the units; for instance, verb conjugations or plural endings can be practiced with alternating pronunciations of /t/ or /<ch>. Instructors should also make sure to choose frequent words for their in-class pronunciation activities. German students in their first semester have limited vocabularies, and choosing frequent words serves two purposes. When students understand the words they are using in the pronunciation activities, it serves as valuable and meaningful input for them and reinforces the connection between meaning and pronunciation in their minds. In addition, training frequent words in the classroom ensures that students can then communicate most efficiently with native speakers after training.

Important for many instructors is the desire to feel comfortable teaching the course material to their students. Many instructors avoid pronunciation instruction because they do not know quite how to go about doing it, or feel that they do not understand enough about the topic themselves (Foote et al., 2012). The pronunciation activities in this unit do not rely too much on technical phonetic instruction and make use of familiar warm-up activities such as partner work,
games and listening activities. Thus, this pronunciation unit and its activities remain within the comfort zone for most instructors and makes sure that even those instructors who do not have training in pronunciation or phonetics can feel comfortable conducting these modules and can successfully teach pronunciation to their students.

In sum, the ideal classroom pronunciation instruction unit for beginning learners should focus on lexical stress assignment and new sounds that are important for distinguishing between words in German, such as the previously mentioned example contrasting the [x] in Nacht and a [k] in nackt. The unit should make use of short warm-up activities each day, which will allow for pronunciation practice in the classroom without taking up too much class time. The warm-up activities are also important because they provide instructors with a familiar basis off of which to introduce pronunciation to their students, and these activities should closely match the topics currently being covered in class in order to best review the relevant vocabulary and grammar points.

5.7 Limitations of Current Study

This dissertation study is one of the first that has been conducted in the fields of Second Language Acquisition in general and pronunciation training in particular that both measures the effectiveness of pronunciation training for beginning learners and directly compares the effectiveness of training at different proficiency levels. Although many insights have come from a research project of this nature, there were also several limitations which would need to be addressed in further studies.

The first limitation of the current study is the small participant numbers. Many of the results in this study should be generalized to larger groups of German learners with caution. Classroom pronunciation training research has its challenges, one of them being the level of
attrition. As with any classroom, students do not always do what is asked of them. The first-semester courses in this study began with a count of 24 students, and a good number of them either did not turn in all of the required components, did not turn in work that was considered suitable for analysis, or did not come to class on the training days. In addition, the fourth-semester groups began the semester with a very low enrollment of only 12 students per section. The combined effects of smaller enrollments and attrition led to group numbers of roughly ten participants. However, it was impossible to include more than two sections of each proficiency level because of section offerings and the fact that the same instructor was required to teach both the experimental and control groups. Without the same teachers, the ability to make a direct comparison between groups would have been compromised. Yet a greater number of participants would have allowed for higher statistical power and more possibilities for generalization of the results.

Another limitation of this study is that the training unit was designed primarily for use in a first-semester classroom, but was also implemented as-is for the intermediate students. The fourth-semester students did not show significant improvement as a result of instruction, but the training program itself may have been a reason for the lack of improvement. Having an identical training unit for all students was necessary in order to perform exact comparisons, but the unit may have been too simple for the fourth-semester students. For instance, all vocabulary and grammatical structures were from the first few chapters of the first-year textbook. Without enough challenging material, it may have been difficult for fourth-semester students to learn from this unit. They may have been bored with the more simple vocabulary and structures, or may have been unable to be convinced of its value for their language development. In future
studies, perhaps a compromise between difficulty and suitability could be made in studies that compare across proficiency levels so that it can be challenging and enjoyable for all students.

Due to time constraints and the number of participants as a whole, it was necessary to provide students with the pre- and posttests as homework assignments to be completed at home. This limitation affected the design of the tests. For instance, there were fewer fillers involved in the word- and paragraph-readings tasks in order to keep the overall task length more manageable for a homework assignment. The investigator also had less control over what the students were doing and how the recordings were completed. For the most part, students completed all assignments to the desired level of detail and clarity required for analysis, so this was not a major concern in the study. But for the perception test, there was no controlling how many times students listened to a certain contrast or how long they took to decide, as the task could not be timed nor a choice forced in any way.

In sum, these limitations mean that the results of this study, while very interesting and encouraging for instructors of beginning German learners, should be interpreted and generalized with caution. As this was one of the first studies to compare the effects of pronunciation training on beginner and intermediate learners, much work on the most efficient training design and implementation still needs to be completed. Future studies should take the limitations of this study into account to develop research in this field even further.

5.8 Areas for Future Research

As this was one of the first studies to investigate the effects of a pronunciation intervention for beginners, future research for elementary-level learners of German is wide open. One goal of future research should be to see if the same results can be replicated with other groups of students, at other universities, and with other teachers.
One of the many potential benefits of training beginning learners that has been put forward in the literature is that of the generalization of pronunciation skills to untrained segments (Hardison, 2004). A study of potential interest and importance would investigate the indirect effects of pronunciation training, that is, the experimental groups’ performance on other aspects of pronunciation that were not trained. Example features to investigate include front-rounded vowels, intonation patterns, or sentence accent. If students who receive pronunciation instruction in class are improving in part because of their developing ability to attend to their own pronunciation and speech and that of others, it is likely that this focus would extend to other sounds as well. Participants who are increasing their consciousness and attending to pronunciation would likely pay attention to aspects of pronunciation that were not trained, compare them with their own productions, and make the necessary changes. A study of this nature would be able to inform researchers as to the probability of the suggested potential benefits of training beginning learners, as well as inform literature on attention in SLA by focusing on pronunciation and phonological acquisition.

Another potential benefit outlined by researchers as an argument for beginning pronunciation training in the first few semesters is that of preventing phonological fossilization. Long-term effects of this nature cannot be measured from the current study, but a longitudinal study can investigate if the boost provided to first-semester learners by pronunciation training really can keep this type of fossilization at bay. A program that continues pronunciation training over two or three elementary-level semester could reveal greater differences in student performance that appear with time and explore how training in the early semesters aids students as they continue on in their language study. Providing beginning students with even more practice in pronunciation may further encourage the development of their comprehensibility,
accentedness, and perceptual abilities. It is possible that over the course of their language study, differences between trained and untrained groups would continue to grow. Students who received training for their first three semesters may be able to develop their attentional and pronunciation skills more and further separate themselves from students who studied German for three semesters without training. In addition, such a combination of an early start plus continued training may be what is necessary to produce the greatest benefit to L2 German learners. A longitudinal study would clarify whether beginning training from the start of language learning and continuing this practice throughout the entire elementary curriculum truly can prevent fossilization of incorrect forms.

Although goals of pronunciation studies vary, one common goal is to measure the most effective methods for training pronunciation. This research question was not addressed in the current study, and most of the previous research has investigated the best methods for training intermediate or advanced learners. Methods from previous studies conducted with intermediate and advanced learners were adopted in the current study, but the field still lacks an examination of the appropriateness of these methods for beginners. Previous research in pronunciation training has also attempted to measure which features are most valuable to train. A combination of suprasegmental and segmental features was chosen for the current study, but deeper investigation into this topic in regard to beginning learners is also needed. Research into Functional Load (Brown, 1991) for German could be conducted, to determine which German contrasts carry the most weight and importance for communication, and which are not as necessary to perfect for effective communication. In addition, acoustic analyses of learner production data could be completed and combined with native speaker ratings in order to make connections between what types of errors are being produced and their consequences are for
native speaker perceptions of comprehensibility and accentedness. Studies of this nature would be able to scientifically address the question of which features of pronunciation to train in elementary-level classrooms. It stands to reason that pronunciation training units for beginners should be designed to address both those sounds that carry a high Functional Load in German, as well as those features that most drastically affect native speaker comprehension. Lastly, future pronunciation research studies for beginners could even modify the pronunciation unit designed for this study in order to fit other languages or language pairs and perform pronunciation training research with students of other languages as well. Research in pronunciation instruction is limited in other languages outside of English, but other world languages, such as German, hold an important place in the discussion.

In summary, this dissertation has made inquiries into aspects of pronunciation training research that have yet to be made regarding the most effective time course for introducing classroom interventions. Results demonstrate that students in their first semester who receive training in pronunciation are able to significantly improve their comprehensibility and reduce their accent to a greater extent than students who did not receive training. In addition, beginning learners seemed more receptive to pronunciation training than students in intermediate proficiency levels did. These findings suggest that pronunciation training has a place in elementary-level language classrooms, and that this training may be even more effective at this level than for more advanced learners. Furthermore, this dissertation is one of the first studies to measure the effects of pronunciation training on beginning learners specifically. It has opened doors to further research that will lead to a better understanding of how beginning learners can best acquire the pronunciation skills that will help them to communicate more effectively in their L2.
References


## Appendix A: Stimuli for Pretest and Posttest

### Word-Reading Task

<table>
<thead>
<tr>
<th>Test Items</th>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>warum</td>
<td>ich</td>
<td>Katze</td>
</tr>
<tr>
<td>zurück</td>
<td>sich</td>
<td>Tante</td>
</tr>
<tr>
<td>Uhren</td>
<td>nicht</td>
<td>Neffe</td>
</tr>
<tr>
<td>spazieren</td>
<td>endlich</td>
<td>Student</td>
</tr>
<tr>
<td>verheiratet</td>
<td>gleich</td>
<td>Musik</td>
</tr>
<tr>
<td>Reise</td>
<td>Nichte</td>
<td>August</td>
</tr>
<tr>
<td>rufen</td>
<td>Geschichte</td>
<td>Moment</td>
</tr>
<tr>
<td>fahren</td>
<td>mich</td>
<td>Abend</td>
</tr>
<tr>
<td>Zimmer</td>
<td>machen</td>
<td>Augen</td>
</tr>
<tr>
<td>Fenster</td>
<td>lachen</td>
<td>Junge</td>
</tr>
<tr>
<td>DVD-Spieler</td>
<td>suchen</td>
<td>Leute</td>
</tr>
<tr>
<td>Schweizer</td>
<td>Woche</td>
<td>Gepäck</td>
</tr>
<tr>
<td>ihr</td>
<td>Hauptfach</td>
<td>bestimmt</td>
</tr>
<tr>
<td>Kinder</td>
<td>Mittwoch</td>
<td>Geschäft</td>
</tr>
<tr>
<td>aber</td>
<td>nach</td>
<td></td>
</tr>
<tr>
<td>Schwester</td>
<td>auch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lampe</td>
<td></td>
</tr>
</tbody>
</table>

### Fillers
- Tisch
- Wand
- klein
- Kunst
- Deutschland
- Tafel
- Kuli
- Eltern
- Universität
- Landkarte
- Enkelkind
- Verwandten
Paragraph-Reading Task


Free Speech Task

1. Wie alt bist du?
2. Was ist dein Hauptfach?
3. Was gibt es in deinem Zimmer?
4. Wann hast du Geburtstag? Was machst du an deinem Geburtstag?
5. Beschreiben Sie Ihre Familie.
Appendix B: Pronunciation Attitude Inventory

Pronunciation Attitude Inventory
Adapted from Elliott (1995) and Botero (2011)

Please rate these statements on a scale from 1 to 5. Ratings are described below.

5 = always or almost always true of me (totally agree)
4 = usually true of me
3 = somewhat true of me
2 = usually not true of me
1 = never or almost never true of me (totally disagree)

1. I’d like to sound as native as possible when speaking German.
2. Acquiring proper pronunciation in German is important to me.
3. I will never be able to speak German with a good accent.
4. I believe I can improve my pronunciation skills in German.
5. I believe more emphasis should be given to proper pronunciation in class.
6. One of my personal goals is to acquire proper pronunciation skills and preferably be able to pass as a near-native speaker of the language.
7. I try to imitate German speakers as much as possible.
8. Communicating is much more important than sounding like a native speaker of German.
9. Good pronunciation skills in German are not as important as learning vocabulary and grammar.
10. I want to improve my accent when speaking German.
11. I’m concerned with my progress in my pronunciation of German.
12. Sounding like a native speaker is very important to me.
13. I do not practice a native-like accent in class because of how other students in class would perceive it.
14. Which set of statements best suits your attitude towards speaking a foreign language?
   a. I like to speak a language perfectly. I like to learn grammar and make sure that I speak correctly when I say things. I like to think about what I will say in a foreign language and how I will say it before I say it.
b. When I speak a foreign language, I aim to communicate. I don’t worry as much about speaking perfectly or getting all of the grammar right. I’m alright with making some mistakes as long as the other speakers understand what I mean.

15. What is the reason for your interest in learning German?
Appendix C: Language Background Questionnaire

Initials (use all 3): _____________________

This questionnaire is designed to give us a better understanding of your experience with other languages. We ask that you be as accurate and as thorough as possible when answering the following questions.

**Part I**
1. Gender: ___________________

2. Age: _____ years

3. Do you have any known hearing problems (either corrected or uncorrected)?
   - No
   - Yes [Please explain: _____________________________________________________________]

4. Native Language(s) (Please check all that apply.)
   - English
   - Other [Please specify: ___________________]

5. What language do you currently think is your dominant language (i.e., the language you are most comfortable using on a daily basis)? (Please check one)
   - English
   - Spanish
   - German
   - Chinese
   - Other [Please specify: ___________________]

**Part II The next section of the questionnaire deals with your second language learning experience.**

6. Please check all of the following that apply and indicate the starting age and length of study for any second language(s) learned.
   - Home/Outside of School – Language(s):
     Starting age? _______ For how long? _________________________
   - Elementary School – Language(s):
     Starting age? _______ For how long? _________________________
   - Middle School – Language(s):
     Starting age? _______ For how long? _________________________
   - High School – Language(s):
     Starting age? _______ For how long? _________________________
   - College – Language(s):
     Starting age? _______ For how long? _________________________

7. Please list the most advanced second language course(s) you have completed in college:
8. What do you consider to be your primary second language? (You may check more than one if you feel that you have multiple primary second languages.)
   - English
   - Spanish
   - German
   - Chinese
   - Other [Please specify: ____________________________]

9. Have you studied and/or lived abroad?
   - Yes
   - No

   **If YES,** where and when did you study, for how long, and what language(s) did you speak?

<table>
<thead>
<tr>
<th>Country</th>
<th>Approx. dates</th>
<th>Length of Stay</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. Are you: (Please check all that apply and indicate which language each applies to if you have studied more than one second language at college.)
   - Taking German for a requirement but interested in being a major or minor.
   - Taking German for a requirement; NOT interested in being a major or minor.
   - A German minor.
   - A German major.
   - Other [Please explain: ____________________________]

11. Do you have any native German friends or acquaintances? If yes, please list.
   a. Relationship to Person 1: ____________________________
      How close are you? ____________________________
      How often do you see this person? ____________________________
   b. Relationship to Person 2: ____________________________
      How close are you? ____________________________
      How often do you see this person? ____________________________
   c. Relationship to Person 2: ____________________________
      How close are you? ____________________________
      How often do you see this person? ____________________________

12. How much contact do you have with native German speakers?
   - None
   - Very little
   - Monthly
   - Weekly
   - Every day
13. How much time do you spend practicing each of these skills in German outside of class time?

<table>
<thead>
<tr>
<th>Skill</th>
<th>None</th>
<th>1-30 minutes</th>
<th>30-60 minutes</th>
<th>1-2 hours</th>
<th>2-4 hours</th>
<th>4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Part III. The next section of the questionnaire deals with your GERMAN language skills. Please rate yourself on each measure by circling the appropriate number. These ratings are for GERMAN.**

14. Your reading proficiency in this language. (1=not literate and 10=very literate)

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

15. Your writing proficiency in this language. (1=not literate and 10=very literate)

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

16. Your speaking ability in this language. (1=not fluent and 10=very fluent)

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

17. Your speech comprehension ability in this language. (1=unable to understand conversation and 10=perfectly able to understand)

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

**Part IV. The next section of the questionnaire deals with your ENGLISH language skills. Please rate yourself on each measure by circling the appropriate number. These ratings are for ENGLISH.**

18. Your English reading proficiency. (1=not literate and 10 = very literate)

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

19. Your English writing proficiency. (1=not literate and 10=very literate)

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

20. Your English speaking ability. (1=not fluent and 10=very fluent)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
21. Your English speech comprehension ability. (1=unable to understand conversation and 10=perfectly able to understand)

1 2 3 4 5 6 7 8 9 10

Thank you for your participation!
VITA
Ashley Roccamo

EDUCATION
2014 Ph.D. in German and Language Science, The Pennsylvania State University, University Park, PA
2007 B.A. in German, Minor in International Studies, Graduated Magna Cum Laude with Departmental Honors, Millersville University, Millersville, PA

PUBLICATIONS


CONFERENCE PRESENTATIONS
2013 “Training our students to listen better: Perceptual awareness training as part of a pronunciation intervention in German language classrooms.” Germanic Linguistics Annual Conference 19. University at Buffalo (SUNY).

GRANTS AND AWARDS
2012 Dissertation Research Grant from Language Learning (Blackwell Publishers)
2012 RGSO Dissertation Support Grant, College of the Liberal Arts, The Pennsylvania State University
2011 Adele Miccio Travel Award, Center for Language Science, The Pennsylvania State University