BRIDGING THE GAP BETWEEN L2 PRONUNCIATION RESEARCH AND TEACHING: USING iCPRS TO IMPROVE NOVICE GERMAN LEARNERS’ PRONUNCIATION IN DISTANCE AND FACE-TO-FACE CLASSROOMS

A Dissertation in
German Applied Linguistics and Language Science
by
Ines A. Martin

© 2018 Ines A. Martin

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

May 2018
The dissertation of Ines A. Martin was reviewed and approved* by the following:

B. Richard Page  
Associate Professor of German and Linguistics  
Dissertation Adviser  
Chair of Committee

Carrie N. Jackson  
Associate Professor of German and Linguistics

Michael T. Putnam  
Associate Professor of German and Linguistics

Matthew Carlson  
Assistant Professor of Spanish and Linguistics

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

Thomas O. Beebee  
Edwin Erle Sparks Professor of Comparative Literature and German  
Head of the Department of Germanic & Slavic Languages and Literatures

*Signatures are on file in the Graduate School
ABSTRACT

Communication in the target language is the main goal in most second language classrooms. Even though there is considerable evidence that pronunciation plays a crucial role in effective communication (Celce-Murcia et al., 2010), pronunciation training is frequently neglected in the foreign language classroom for reasons such as lack of classroom time or lack of instructor preparedness to teach pronunciation (Derwing, 2013; Grim & Sturm, 2016; Foote et al., 2011). The present dissertation addresses this discrepancy by investigating the effectiveness of a homework-based method of pronunciation instruction for beginner learners that does not require teacher feedback or in-class time and thus allows for pronunciation training to be included in any second language (L2) curriculum. Through this method, called innovative Cued Pronunciation Readings (iCPRs; Martin, 2015, 2017, see also Tanner & Landon, 2009), students receive pronunciation training in the form of Microsoft PowerPoint units that every learner can access from a home or library computer. Thus, this type of pronunciation training can be assigned as homework and students work on it alone. This allows for an implementation of pronunciation training in the curriculum without taking up valuable in-class time or requiring special skills from the instructor.

Since pronunciation training via iCPRs can be delivered online, the method lends itself not only to use in face-to-face but also in distance (i.e. online) learning environments. The emergence of online and hybrid courses over the past two decades has presented new challenges as well as opportunities for language instruction. In particular, the development of learners’ oral proficiency in an online learning environment has been shown to be problematic (Deutschmann et al., 2009). In light of these developments, it is important to consider language instruction not only in traditional face-to-face environments but also in online learning environments.
The present dissertation assessed the effectiveness of the iCPR method in on-campus and online learning environments by measuring novice German learners’ improvement in perception and production skills over the course of one semester. Data collection was conducted with 90 first-semester learners of German. In the face-to-face environment, each of the six sections of first semester German classes was pseudo-randomly assigned the status of treatment, comparison, or control group. Students in the treatment group received ten weeks of pronunciation training delivered through homework-based iCPR units, whereas students in the comparison group received ten weeks of ten minutes in-class pronunciation training delivered by their instructors. Finally, students in the control group followed the same standardized first-semester German lesson plans as the other two groups, but did not receive targeted pronunciation training. Since there is no in-class instruction in the online learning environment, there was no comparison group in this condition. Instead, four sections of online first-semester German classes were divided into two sections that served as a treatment group (i.e., a group that received instruction through iCPR units) and two sections that served as a control group.

The study employed a pretest / posttest design: participants’ pronunciation was recorded in the 2nd and 14th week of the semester. The assessment included a binary-choice perception task and participants’ oral productions at the word and sentence level. Following standard procedure in pronunciation training research, these speech productions were rated by native German speakers for ease of comprehensibility and strength of accent. The ratings from each group were compared in order to assess the effectiveness of the different types of pronunciation training.

For the face-to-face learning environment, results show that the learners in the treatment group, who received homework-based iCPR pronunciation training, significantly outperformed
learners in the control group on measures of phonological perception skills, comprehensibility in productions of individual words, and accentedness in productions of individual words. Moreover, results reveal no significant difference between the learners who received homework-based pronunciation training and learners who received in-class pronunciation training. Together, these results suggest that learners benefit from iCPR pronunciation training and that homework-based pronunciation training is as effective in improving learners’ German perception and production skills as in-class pronunciation training.

For the online learning environment, results show that the learners in the treatment group, who received iCPR pronunciation training as part of their online curriculum, significantly outperformed learners in the control group on all measures, that is, in phonological perception skills, in ease of comprehensibility in productions on the word- and paragraph-level, and in a diminished accent in productions on the word-and paragraph-level. These findings suggest that learners in an online environment benefit from iCPR pronunciation training.

Taken together, this dissertation shows that effective pronunciation training can be provided outside the classroom via iCPRs, which saves valuable in-class time and opens up possibilities for learners to receive pronunciation instruction even in an online learning environment or if their instructor lacks the preparation to teach this topic. As such, this dissertation suggests a method of pronunciation instruction that addresses the problems that all too often lead to the neglect of pronunciation training in traditional classrooms and in online learning environments.
# TABLE OF CONTENTS

List of Figures .................................................................................................................. ix
List of Tables ................................................................................................................... x
List of Abbreviations ....................................................................................................... xi
Acknowledgements ........................................................................................................ xiii

## CHAPTER 1: Introduction ....................................................................................... 1
  1.1 Description of the Problem ......................................................................................... 1
  1.2 Layout of the Dissertation .......................................................................................... 5

## CHAPTER 2: Review of the Literature ................................................................... 6
  2.1 Pronunciation in the L2 Classroom .......................................................................... 6
     2.1.1 Why teach L2 Pronunciation? ............................................................................ 6
     2.1.2 The Research-Practice Gap and Teacher Preparedness .................................... 12
  2.2 Prior Research in L2 Pronunciation Instruction ....................................................... 17
     2.2.1 The Goal of Pronunciation Instruction .............................................................. 17
     2.2.2 The Relationship between Perception and Production ....................................... 21
     2.2.3 The Scope of L2 Pronunciation Training: Segmentals and Suprasegmentals .... 26
     2.2.4 Ecological Validity in L2 Pronunciation Teaching: Lab-based and Classroom-based Studies .............................................................................. 28
     2.2.5 Using Technology to Teach Pronunciation ....................................................... 31
  2.3 Pronunciation Instruction in Distance Language Learning ...................................... 38
  2.4 The Present Study ................................................................................................. 44
     2.4.1 The Method of iCPRs ...................................................................................... 44
     2.4.1.1 CPRs vs. iCPRs .......................................................................................... 44
     2.4.1.2 Design of Materials for the Present Study .................................................... 47
     2.4.1.3 Pronunciation Targets Chosen for Training .................................................... 53
     2.4.2 Research Questions and Hypotheses ................................................................. 55
     2.4.2.1 Experiment 1 – Face-to-Face Learning Environment ................................ 55
     2.4.2.2 Experiment 2 – Distance Learning Environment .......................................... 58

## CHAPTER 3: Experiment 1 – Face-to-Face Environment ......................................... 61
  3.1 Methodology ............................................................................................................ 61
  3.1.1 Participants ......................................................................................................... 61
  3.1.2 Materials ........................................................................................................... 64
     3.1.2.1 Instructional Materials ................................................................................ 64
     3.1.2.2. Testing Materials ..................................................................................... 76
  3.1.4 Procedure ............................................................................................................ 81
     3.1.4.1 Overview of Experimental Procedure ......................................................... 81
     3.1.4.2 Testing Procedure ....................................................................................... 82
     3.1.4.3 Pronunciation Training Procedure ............................................................... 83
  3.1.5 Scoring and Data Analysis .................................................................................. 85
     3.1.5.1 Perception ................................................................................................. 85
CHAPTER 5: Conclusions .......................................................... 171
5.1 Overview of Findings .......................................................... 171
5.2 Differences in Experiment 1 and Experiment 2 ......................... 174
5.3 Pedagogical Implications ...................................................... 177
5.4 Limitations and Directions for Future Research ....................... 180
5.5 Final Remarks and Conclusion .............................................. 183

References .................................................................................. 185
Appendix A: List of Sound Discrimination Training Items .................. 199
Appendix B: Sample iCPR Production Unit ..................................... 200
Appendix C: Handout Perception Pretest ........................................ 204
Appendix D: Production Task Word Reading Items .......................... 206
Appendix E: Production Task Reading Paragraphs ............................ 207
Appendix F: Sample of Production Test PowerPoint Slides ................. 208
Appendix G: Language Background Questionnaire ........................... 211
Appendix H: Participant Exit Questionnaire – iCPR Group ................. 214
Appendix I: Participant Exit Questionnaire – In-Class Group .............. 217
Appendix J: Summary of Findings – Exp. 1 – p-values, effect sizes ........ 219
Appendix K: Summary of Findings – Exp. 2 – p-values, effect sizes ........ 220
List of Figures

Figure 1. Sample Slides for Accentedness Detection Task ............................................. 68
Figure 2. Sample Slides for Sound Discrimination Task .................................................. 70
Figure 3. Sample Production Practice Slides .................................................................... 73
Figure 4. Exp. 1 – Accentedness Detection Task – Between-group Results .................... 93
Figure 5. Exp. 1 – Accentedness Detection Task – Within-group Results ....................... 94
Figure 6. Exp. 1 – Sound Discrimination Task – Between-group Results ...................... 96
Figure 7. Exp. 1 – Sound Discrimination Task – Within-group Results .......................... 96
Figure 8. Exp. 1 – Comprehensibility on Word-level Productions – Between-group Results..... 98
Figure 9. Exp. 1 – Comprehensibility on Word-level Productions – Within-group Results ...... 99
Figure 10. Exp. 1 – Comprehensibility on Paragr.-level Prod. – Between-group Results....... 101
Figure 11. Exp. 1 – Comprehensibility on Paragr.-level Productions – Within-group Results . 101
Figure 12. Exp. 1 – Accentedness on Word-level Productions – Between-group Results ...... 103
Figure 13. Exp. 1 – Accentedness on Word-level Productions – Within-group Results ........ 104
Figure 14. Exp. 1 – Accentedness on Paragraph-level Productions – Between-group Results . 106
Figure 15. Exp. 1 – Accentedness on Paragraph-level Productions – Within-group Results .... 106
Figure 16. Exp. 2 – Accentedness Detection Task – Between-group Results ..................... 149
Figure 17. Exp. 2 – Accentedness Detection Task – Within-group Results ....................... 149
Figure 18. Exp. 2 – Sound Discrimination Task – Between-group Results ........................ 151
Figure 19. Exp. 2 – Sound Discrimination Task – Within-group Results .......................... 151
Figure 20. Exp. 2 – Comprehensibility on Word-level Productions – Between-group Results. 153
Figure 21. Exp. 2 – Comprehensibility on Word-level Productions – Within-group Results .... 153
Figure 22. Exp. 2 – Comprehensibility on Paragr.-level Prod. – Between-group Results ....... 155
Figure 23. Exp. 2 – Comprehensibility on Paragr.-level Productions – Within-group Results . 155
Figure 24. Exp. 2 – Accentedness on Word-level Productions – Between-group Results ....... 157
Figure 25. Exp. 2 – Accentedness on Word-level Productions – Within-group Results ....... 157
Figure 26. Exp. 2 – Accentedness on Paragraph-level Productions – Between-group Results . 159
Figure 27. Exp. 2 – Accentedness on Paragraph-level Productions – Within-group Results .... 159
List of Tables

Table 1: Exp. 1 – Learners’ Age and German Proficiency Ratings ........................................... 62
Table 2: Exp. 1 – Biographical Information of Speakers for Recordings .................................. 66
Table 3: Exp. 1 – Rater Consistency for Rated Continua by Task and Rater Group .................. 90
Table 4: Exp. 1 – Descriptive Statistics for Accentedness Detection Task .............................. 94
Table 5: Exp. 1 – Descriptive Statistics for Sound Discrimination Task ................................. 97
Table 6: Exp. 1 – Descriptive Statistics for Comprehensibility on Word-level Productions ...... 99
Table 7: Exp. 1 – Descriptive Statistics for Comprehensibility on Paragr.-level Productions ... 102
Table 8: Exp. 1 – Descriptive Statistics for Accentedness on Word-level Productions .......... 104
Table 9: Exp. 1 – Descriptive Statistics for Accentedness on Paragraph-level Productions ...... 107
Table 10: Exp. 1 – Summary of Findings .............................................................................. 107
Table 11: Exp. 1 – Mean Scores on Exit Questionnaire – iCPR Training Group .................... 110
Table 12: Exp. 1 – Mean Scores on Exit Questionnaire – In-class Training Group ............... 116
Table 13: Exp. 2 – Learners’ Age and German Proficiency Ratings ....................................... 140
Table 14: Exp. 2 – Rater Consistency (Cronbach’s α, [95% CI]) for Rated Continua by Task . 146
Table 15: Exp. 2 – Descriptive Statistics for Accentedness Detection Task ......................... 150
Table 16: Exp. 2 – Descriptive Statistics for Sound Discrimination Task ............................. 152
Table 17: Exp. 2 – Descriptive Statistics for Comprehensibility on Word-level Productions ... 154
Table 18: Exp. 2 – Descriptive Statistics for Comprehensibility on Paragr.-level Productions . 156
Table 19: Exp. 2 – Descriptive Statistics for Accentedness on Word-level Productions ....... 158
Table 20: Exp. 2 – Descriptive Statistics for Accentedness on Paragraph-level Productions .... 160
Table 21: Exp. 2 – Summary of Findings .............................................................................. 160
Table 22: Exp. 2 – Mean Scores on Exit Questionnaire – Online iCPR Training Group ....... 162
List of Abbreviations

ADT: Accentedness Detection Task
ASR: Automatic Speech Recognition
CALL: Computer Assisted Language Learning
CAPT: Computer Assisted Pronunciation Training
CI: Confidence Interval
CMC: Computer Mediated Communication
CPH: Critical Period Hypothesis
CVC: Consonant Vowel Consonant
ESL: English as Second Language
F2F: Face-to-Face
FL: Foreign Language
HVPT: High Variability Phonetic Training
iCPRs: innovative Cued Pronunciation Readings
IQR: Interquartile Range
JSLP: Journal of Second Language Pronunciation
L1: first language
L2: second language
M: Mean
Mdn: Median
PAM: Perceptual Assimilation Model
SD: Standard Deviation
SDT: Sound Discrimination Task
SLA: Second Language Acquisition
SLM: Speech Learning Model
TESOL: Teaching English to Speakers of Other Languages
VCV: Vowel Consonant Vowel
Acknowledgements

First and foremost, I would like to express my deep gratitude to the members of my doctoral committee: Richard Page, Carrie Jackson, Michael Putnam, Matt Carlson, and Mary O’Brien. Thank you for your thoughtful comments, guidance, and support during my doctoral studies. Richard, thank you for always being so generous with your time, for all your thoughtful advice—not only on my dissertation but on every aspect of my grad school career and life—, and for encouraging, guiding, and supporting me in finding my own path and following my own research interests. Your confidence in my abilities as a teacher and researcher has always allowed me to grow and I am fortunate to have had you as an advisor. Carrie, you have taught me so very much and have had an enormous influence on making me the researcher I am today: I will always be grateful for that! Learning not only the theoretical, but also the practical side of sound research practice from you was invaluable to me and played a significant role in allowing me to write this dissertation. Thank you also for giving me the privilege of being a member of your lab for the past four years! Mary, you were so kind to me from the first moment we met at my first GLAC conference very early in my graduate career: thank you for believing in me from the start, when I presented a very rough draft of a dissertation project to you. Your expertise and your willingness to share this expertise even over the 2200 miles that separate State College from Calgary have been invaluable to me. I am so grateful for all your advice in designing my experiments, creating my stimuli, and keeping my research questions manageable. Mike and Matt, thank you for so generously contributing your time and expertise to me and my dissertation. You have provided a wealth of valuable feedback to this dissertation at various stages of this project. Mike, thank you also for always being up for a conversation about non-
academic topics and for reminding me on various occasions throughout grad school that it is important to have a life outside of academia.

I would also like to express my gratitude to the numerous organizations that have funded this dissertation: the College of the Liberal Arts, Language Learning, TOEFL ETS, and The Department of Applied Linguistics. A special thank you also goes out to the Department of Germanic and Slavic Languages and Literatures at Penn State, whose faculty and staff have not only made my life so much easier on a daily basis, but which has also supported my doctoral career with very generous funding for professional development that should not go unmentioned.

I am further incredibly grateful to all my undergraduate research assistants who have been tremendously helpful with the coding of my data: O’Neill Kennedy, Daphne Koufomihalis, Larissa Gil, Lisa Cox, Katherine Muschler, Delaney Wilson, and Devin McCoy. A German Dankeschön further goes out to all my participants who I cannot name here without violating my IRB protocol but without whom this dissertation would not exist: the students in twelve sections of first semester German at Penn State, the instructors of these sections, the native speaker raters, and also the native speakers who have lend their voices in the recordings of the iCPR units for this project.

Finally, I would like to take this opportunity to thank my friends and family as well. Nora, Lara, Hyoun-A, and Courtney, I am so fortunate that I had colleagues who turned out to be such wonderful friends too! You were there for the ups and downs, made any workload bearable, and made attending conferences so much more fun. Thank you for being both my academic and my personal support system, for always understanding when people outside of academia do not, and, most importantly, for all the times you made me laugh until I cried! Thank you also to my
wonderful friend, Emilie, for providing me feedback on my dissertation from the perspective of a non-linguist reader and for proofreading the final version of this manuscript.

I am also so incredibly grateful to my parents, Frauke and Dieter, for creating such a loving and supportive home environment and for not allowing any roadblocks to get in my way to success. Mama, Papa, ohne eure Unterstützung während meiner Schul- und Studienzeit und all die Auslandsaufenthalte, die ihr mir ermöglicht habt, wäre ich nie bis zur Doktorwürde gekommen. Danke!

The biggest thank you of all goes to my husband, Mike. Without your unconditional love and support, your incredible patience, and your understanding, there would be no dissertation! Thank you for standing with me every step of the way. I am so lucky to have you!
CHAPTER 1: Introduction

1.1 Description of the Problem

Whenever we meet a nonnative speaker, a foreign accent is usually the first hint that gives away the fact that we are not talking to a native speaker of the language. Flege (1984) found that sometimes short samples of less than a second can be enough for native speakers to detect nonnative speech, and Major (2007) showed that even listeners who do not speak a particular language can detect foreign accents in it. Not passing as a native speaker does not necessarily entail a disadvantage, but research has shown that a nonnative accent can lead to stigmatization of the interlocutor (Munro, 2003). Gordon and Darcy (2016) observe that accented second language (L2) speech is usually independent from education levels and intelligence as can be seen in the case of many foreign doctors in the North American health-care system or teaching assistants at universities that often exhibit very accented pronunciation despite their high education level. Yet, Duppenthaler (1991) points to the fact that listeners often assume that second language (L2) speakers with a marked accent have a very limited language ability and even an inferior mental ability. He claims that this “is a distinct disadvantage for those engaged in activities, such as political and business negotiations […] that require a high degree of mutual respect on the part of all concerned” (1991, p. 33) (see also Lev-Ari & Keysar, 2010, 2012).

While this quote highlights the importance of good pronunciation for political and business negotiations, Celce-Murcia, Brinton, Goodwin, and Griner (2010) go one step further and point out the importance of pronunciation in everyday communication: "there is a threshold level of pronunciation for nonnative speakers of English; if they fall below this threshold level, they will have oral communication problems no matter how excellent and extensive their control of English grammar and vocabulary might be" (p. 8). This clearly shows that good L2
pronunciation—and, thus, pronunciation training—play an important role in foreign language instruction.

Despite these findings that highlight the importance of pronunciation proficiency in an L2, pronunciation instruction is often neglected in the classroom compared to the teaching of other L2 skills. Reasons for this shortcoming are widely discussed in the literature on pronunciation instruction (see Levis, 2016a for an overview), but the two most dominant reasons are that foreign language teachers often do not feel comfortable with teaching pronunciation as well as that they are lacking time to address pronunciation in the classroom. As for the former reason, this feeling usually comes from a lack of training during the language instructors’ education, whereas the lack of time problem is often explained by the fact that standardized testing focusses heavily on grammar and vocabulary knowledge (Breitkreutz, Derwing, & Rossiter, 2001; Derwing, 2013; Foote, Holtby, & Derwing, 2011; O’Brien, 2004). This in turn leads teachers to focus on these areas of L2 acquisition even though they are aware that, for successful communication, a focus on speaking and pronunciation is just as valuable and necessary.

The goal of the present dissertation is to present a solution for the discrepancy between the importance of pronunciation in L2 communication and its neglect in the classroom. To this end, I am investigating a homework-based method of pronunciation instruction, called innovative Cued Pronunciation Readings (iCPRs). The method of iCPRs is based on Cued Pronunciation Readings, which were first introduced by Tanner and Landon (2009). It differs from the original version in that Tanner and Landon addressed pronunciation on the level of sentence melody and stress, whereas iCPRs more generally target any problematic L2 sound or suprasegmental feature. In applying the method of iCPRs, learners receive pronunciation training in the form of
Microsoft PowerPoint units that every student can access from a home or library computer. Thus, this type of pronunciation training can be assigned as homework and learners work on it alone. Providing pronunciation training through homework assignments solves the problems mentioned above: it does not take up valuable in-class time, and even those teachers who do not feel adequately prepared to teach pronunciation can assign the exercises as homework. Furthermore, when practicing pronunciation as homework, learners are no longer singled out in front of the entire class, which is a valuable feature both to instructors and learners.

Due to its homework-based nature, the method of iCPRs not only lends itself for use in traditional face-to-face but also in distance (i.e., online) learning environments. The emergence of online and hybrid courses over the past two decades has presented new challenges as well as opportunities for language instruction. In particular, the development of learners’ oral proficiency in an online learning environment has shown to be problematic (Deutschmann, Panichi, & Molka-Danielsen, 2009). In light of these developments, it is important to consider language instruction not only in traditional face-to-face environments but also in distance learning environments like Penn State’s World Campus. Thus far, no study has explored the inclusion of particular methods of teaching pronunciation in online environments or their effect on oral proficiency development.

In short, the present study seeks to investigate whether pronunciation training delivered through iCPRs is as effective as in-class pronunciation training—both, in a face-to-face and an online learning environment. To address this question, two experiments were conducted, measuring novice learners’ improvement in listening and speaking skills over the course of one semester. The first experiment explored iCPRs in a face-to-face (i.e., a campus) learning

---

1 The terms distance and online language learning will be used interchangeably in this dissertation. For the reasoning behind this use of both terms refer to Section 2.3.
environment. Learners from six sections of Penn State University Park first-semester German classes were pseudo-randomly assigned the status of treatment, comparison, or control group. Students in the treatment group received ten weeks of pronunciation training delivered through homework-based iCPR units, whereas students in the comparison group received ten weeks of ten minutes in-class pronunciation training delivered by their instructors. Finally, students in the control group followed the same standardized GER 001 lesson plans as the other two groups, but did not receive targeted pronunciation training. The second experiment explored iCPRs in an online learning environment. Learners from four sections of Penn State World Campus first-semester German classes were pseudo-randomly assigned the status of treatment or control group. Since there is no in-class instruction in the online learning environment, there was no comparison group in this experiment. Learners’ pronunciation skills were assessed in the 2\textsuperscript{nd} and 14\textsuperscript{th} week of the semester by means of a perception test and native-speaker ratings of learners’ productions. These ratings constituted the data for statistical analyses that allowed the learners’ improvement to be measured.

In conclusion, the present study is directly applicable to any foreign language classroom at the K-12 and post-secondary education levels. The method of iCPRs is designed for direct implementation by language learners and can supplement existing foreign language curricula with minimal effort by the instructors. The results of this study can directly link linguistic research to L2 teaching and positively impact learners’ overall L2 proficiency and oral communication skills.
1.2 Layout of the Dissertation

This dissertation is organized as follows. Chapter 2 presents the background literature that is most relevant to the present study. This overview starts with a discussion of why L2 pronunciation should be taught and why it is often neglected in the L2 communicative classroom. I then review prior research in L2 pronunciation, highlighting major findings on the goal of pronunciation instruction and the role of ecological validity of such instruction in lab-based, classroom-based, and technology-based pronunciation studies. Moreover, I provide a short overview of the special role of pronunciation instructions in distance language learning as compared to face-to-face learning. Chapter 2 then concludes with a review of the method of iCPRs and a description of the present study, including the research questions addressed in this dissertation. Chapter 3 presents the methodology, results, and discussion of the first experiment—the investigation of iCPRs in a face-to-face learning environment. Chapter 4 presents the methodology, results, and discussion of the second experiment—the investigation of iCPRs in an online learning environment. Finally, in Chapter 5, I summarize the results from both experiments and discuss the major pedagogical implications of these findings. I further discuss some potential limitations of the present study and highlight directions for future research.
CHAPTER 2: Review of the Literature

2.1 Pronunciation in the L2 Classroom

2.1.1 Why teach L2 Pronunciation?

Up until very recently, a majority of literature review sections of L2 pronunciation training studies started by lamenting the meagre attention that pronunciation research has received compared to other fields in Second Language Acquisition (SLA), such as vocabulary or grammar learning (see e.g., Arteaga, 2000; Counselman, 2010; Couper, 2003; Couper, 2006; Derwing & Munro, 2005; Derwing & Rossiter, 2002; Ducate & Lomicka, 2009; Elliott, 1997; Lord, 2008; O’Brien, 2004; Ruellot, 2011; Saito, 2011b). Even still in 2012, in his comprehensive synthesis of 15 experimental pronunciation studies, Saito claims “what is surprising in this vein of L2 acquisition studies, [...] is the lack of research in the area of L2 pronunciation development” (p. 842). Over the past five years, however, the tide has shifted and the field of pronunciation research saw an increase in publications in peer-reviewed journals, conference proceedings, and graduate level theses (see also Thomson & Derwing, 2015 for a narrative review; J. Lee, Jang, & Plonsky, 2015 for a meta-analysis; and Derwing & Munro, 2015 for a comprehensive book summing up evidence-based pronunciation fundamentals). The recent upswing in interest in pronunciation research and instruction is also reflected in the emergence of a new journal entirely dedicated to pronunciation research (Journal of Second Language Pronunciation, first volume published in 2015) as well as in several international conferences in the field of pronunciation research: English Pronunciation: Issues and Practices (currently in its 5th year), New Sounds (currently in its 8th year), and Pronunciation in Second Language Learning and Teaching (currently in its 9th year). Moreover, Levis (2016) reports an increasing number of well-attended sessions on L2 pronunciation at established applied
linguistics conferences, and further provides interesting figures obtained in a survey conducted for the *Journal of Second Language Pronunciation*. According to this survey, the years 2010-2012 saw 55 articles on L2 pronunciation published in 31 frequently cited journals in applied linguistics, while the years 1999-2008 had only shown publications in 14 journals, and with an average of only about 3% of articles on topics related to L2 pronunciation (Levis, 2016a). Thus, we see a clear trend for increased productivity in the field of L2 pronunciation research, and increased interest in studies that relate to pronunciation teaching. Yet, since this change has been so recent, it is important to discuss the reasons why pronunciation research had been comparatively neglected before and what sparked this recent change in attitude towards pronunciation research.

In one of the classic, much-cited papers in pronunciation research, Elliott (1995) writes that “we justify the neglect of pronunciation instruction by believing that pronunciation, at least for adult FL learners, is more difficult to improve than other target language skills” (p. 531). This statement warrants a discussion of why many foreign language instructors might think so. In response to this question, SLA literature usually engages in a discussion of the *Critical Period Hypothesis* (CPH), that is, the idea that a critical period exists after which native-like acquisition of an L2, particularly the L2 phonology, seems impossible to achieve due to changes in the brain that make it difficult to access the cognitive mechanisms needed for language learning after a certain age (see Derwing & Munro, 2015 and Scovel, 2000 for a detailed summary). The CPH was first posited by the Canadian brain researcher and neurosurgeon Wilder Penfield (Penfield & Roberts, 1959) and was then made known to the applied linguistics community and quite vigorously defended in the following years by Eric Lenneberg (Lenneberg, 1967). Despite the fact that a few studies have since shown that even adult learners can attain native-like
pronunciation under certain circumstances (Bongaerts, van Summeren, Planken, & Schils, 1997; see also Moyer, 1999), the CPH was widely accepted in the 1970s-1990s. This aids in explaining why many teachers then considered it a waste of classroom time to teach pronunciation. Fortunately, a wealth of research conducted in the last few decades has furthered our understanding of the CPH. While early research like Oyama (1976) already suggested that rather than a critical (i.e., a cut-off) period, it might be better to talk about a sensitive (i.e., more fluid) period of native-like L2 attainment, the study that finally challenged the widespread acceptance of the CPH was Flege, Munro, and MacKay (1995). The authors conducted a large-scale study with 240 Italian immigrants who came to Canada between the ages of 3-25. They found that age of arrival served as a predictor of strength of foreign accent, with a perfectly linear relationship between age of arrival and degree of perceived accent. That is, there was no cut-off age before which the participants’ speech was judged as native-like or after which it was judged as non-native like. This showed that it is unlikely that improvement in L2 pronunciation is constrained by changes in the brain at a certain age, e.g. before puberty. Nevertheless, the authors, as well as most research in the following years, acknowledge that while there does not seem to be a critical period in the sense of a cut-off age for native-like L2 acquisition, the ease with which a foreign language and its phonological system can be learned still declines with age. Today, the notion of “the earlier to start learning an L2, the better” has asserted itself while acknowledging the fact that it is still possible to attain a high degree of proficiency in an L2 after puberty. The interest in topics relating to the CPH has rather shifted to questions of how age of learning is affected and mediated by factors such as motivation, aptitude, experience, and instructional methods (Derwing & Munro, 2015).
Another main reason frequently discussed in the literature for why L2 pronunciation instruction has often been neglected in the foreign language classroom, is that with the shift towards Communicative Language Teaching, starting in the mid-1980s, a focus on meaning was prioritized over form-focused instruction. This shift in language learning methodology led to a de-emphasis of pronunciation training based on the general assumption that pronunciation would improve through simple exposure to the L2 or through immersion and study-abroad experiences (Celce-Murcia et al., 2010; Chun, 2002; Derwing & Munro, 2015; Thomson & Derwing, 2015). This assumption did not, however, withstand the test of time and did not hold up when addressed in empirical research. Several studies conducted over the past twenty years were designed with a treatment group that received some form of pronunciation instruction and a control group that followed the same lesson plan but did not receive pronunciation instruction. The lack of improvement in the control groups’ pronunciation performance in these studies generally shows that, even in classes with an advanced proficiency level, learners’ pronunciation in the L2 does not automatically improve without some form of instruction (e.g. Botero, 2011; Couper, 2006; Elliott, 1997; J.-Y. Lee, 2009; Roccamo, 2014).

Moreover, recent research by Trofimovich and Baker (2006) and by Derwing and Munro (2008) suggests that not even extended stays in immersion/study-abroad environments guarantee automatic improvement of L2 pronunciation. Trofimovich and Baker (2006) investigated the effects of 3-months-, 3-years-, and 10-years-immersion stays of Korean ESL learners in the United States. While they showed that, with instruction and over time, learners’ command of some suprasegmental features can improve, their analyses also revealed that on measures of suprasegmentals (stress timing, peak alignment, speech rate, pause frequency, and pause duration), the 10-years-immersion speakers of English were still rated significantly lower than
native speakers. This suggests that even an immersion period of ten years is not sufficient to
acquire a native-like accuracy on suprasegmental features. A somewhat similar study by Munro
and Derwing in 2008 came to a similar conclusion. Munro and Derwing’s study differed from
Trofimovich and Baker’s study in several characteristics, that is, the authors focused on
segmental productions (the acquisition of vowels) and the immersion time was much shorter (up
to a year). Nevertheless, their findings were comparable in that they found that while initially
improvement in L2 vowel production occurred naturally, it stagnated after around six months of
living in the immersion environment. One recurring hypothesis that can be found in the literature
as an explanation for this lack of improvement of pronunciation without instruction is connected
to the fact that learners do not seem to be able to identify and therefore overcome their own
pronunciation errors (Derwing & Rossiter, 2002; Dlaska & Krekeler, 2008). By means of
analysis of a survey administered among 100 adult ESL learners, Derwing and Rossiter (2002)
showed that almost 40% of the participants were unable to identify which specific areas in their
speech caused communications breakdowns or problems. That is, a majority of the surveyed
learners expected segmental problems to lead to communication breakdowns, however, this was
usually not the actual reason. Similarly, in a study investigating the role of self-assessment in
pronunciation, Dlaska and Krekeler (2008) found that in their treatment group of L2 German
learners, the participants were only able to identify about half of the number of speech sounds,
which the native-speaker raters had judged as inaccurate, when they were asked to listen back to
their own recordings. In fact, the analysis showed that the learners had so much difficulty with
identifying their own mistakes in L2 German that they even rated some of their correct
productions as incorrect.
In the decades during which the field of L2 pronunciation research dealt with the consequences that Communicative Language Teaching had had on pronunciation training, and while the knowledge sunk in that pronunciation does not improve merely through exposure to the L2, more and more research was conducted which showed that actively teaching pronunciation can help improve pronunciation proficiency in an L2 (e.g. Arteaga, 2000; Cenoz & Lecumberri, 1999; Couper, 2003; Elliott, 1995, 1997; Gonzalez-Bueno, 1997; Lord, 2005, 2008). More recently, Derwing, Munro, Foote, Waugh, and Fleming (2014) showed that pronunciation instruction was effective even in putatively fossilized learners. The researchers conducted a pronunciation training study in a work-place setting with learners that had been immersed in an English-speaking environment for an average of nineteen years. While the pronunciation (and L2 skills in general) of such individuals would generally be judged as strongly fossilized, the participants showed improvement on measures of speech perception and in comprehensibility and intelligibility of their productions. Findings like these, together with clear results from large-scale narrative reviews (such as Thomson & Derwing, 2015) and meta-analyses (such as J. Lee, Jang, & Plonsky, 2015), have influenced the change of attitude towards pronunciation training in the recent past, finally putting the debate to rest whether pronunciation training is beneficial in improving learners’ L2 pronunciation. That is, in their narrative review of 75 pronunciation studies, Thomson and Derwing (2015) state that 82% of the studies report a significant improvement of learners’ pronunciation through instruction. Moreover, in their meta-analysis of the same 75 pronunciation studies, J. Lee et al. (2015) found that the learners who received instructional treatment improved by 0.89 standard deviation units in comparison with their pre-treatment performance and outperformed control groups by 0.80 standard deviation units. The authors further compare these results to meta-analytic findings in other areas of instructed SLA.
and find that these results show “that instruction on pronunciation can be just as (or more) effective as vocabulary, grammar, and pragmatics” (p. 357).

Finally, not only has pronunciation instruction shown to be beneficial in improving learners’ pronunciation proficiency, but it has also been linked to learners’ willingness to communicate (Derwing et al., 2014), their confidence (Tang, Zhang, Li, & Zhao, 2013), their feeling of belonging (Gluszek & Dovidio, 2010), and even their orthographic development (M. Wang, Park, & K. Lee, 2006) (see also Gooch, Saito, & Lyster, 2016). Taken all together, it comes as no surprise that the tide has shifted and more L2 pronunciation studies have appeared in the past few years, laying the foundation for increasingly rigorous research and solid findings.

2.1.2 The Research-Practice Gap and Teacher Preparedness

In the previous section, I discussed findings that have shown that pronunciation training is beneficial for learners in improving their pronunciation proficiency and that pronunciation instruction can be as effective as instruction in other areas of SLA. Moreover, various surveys conducted among language learners have shown that L2 learners themselves see pronunciation training as very important (Couper, 2003; Derwing & Munro, 2015). A logical consequence of these findings would be to see an increase of pronunciation instruction in L2 classrooms. Yet, unfortunately, research has not translated into widespread practice in this area.

In 2005, Derwing and Munro published a widely cited paper in which they lamented that not only pronunciation research, but particularly pronunciation teaching research and practice had been heavily neglected in the past. The authors complained that in pronunciation teaching, instructors have to rely almost entirely on their intuitions and practical experience. Throughout the article, Derwing and Munro discussed why such trial-and-error based pronunciation teaching
has serious drawbacks, and why it is detrimental to the field in general if every teacher seemingly has to re-invent the wheel when they want to teach L2 pronunciation. As main reasons for this predicament, the authors listed the fact that teachers do not receive adequate training to teach pronunciation, that research on pronunciation instruction is mostly published in academic journals focusing on speech production and perception (which are not easily accessible for teachers and are not written in a teacher-oriented fashion), and, finally, that many pronunciation studies are carried out in laboratory environments that often do not clearly relate to classroom instruction (for further discussion of the merit of lab-based studies, see Section 2.2.2). Derwing and Munro therefore called for a change in future research and insisted on greater collaboration between researchers and teachers, with the ultimate goal that more research actually relevant for the classroom is conducted.

Derwing and Munro’s (2005) article received a lot of attention in the literature, and their concerns were echoed by many other researchers. A study by Foote, Trofimovich, Collins, and Soler Urzúa (2016), for example, showed that pronunciation instruction is indeed neglected in the classroom. In this longitudinal study, the authors analyzed 40 hours of videotaped lessons of 6th grade francophone learners of English and found that pronunciation teaching occurred very infrequently, that it targeted almost exclusively individual sounds, and that it was almost never included in the lesson plans but rather only occurred spontaneously in the form of corrective feedback in response to a learner error. Importantly, the authors point out that these findings directly contradict results of published teacher surveys in which more than 75% of teachers self-reported teaching pronunciation more than once per week. It thus appears as if there is a gap between perception and reality in teaching pronunciation. Moreover, several studies echo Derwing and Munro’s (2005) findings that many teachers do not feel comfortable with or
adequately prepared to teach pronunciation. In their 2001 survey of 67 ESL programs in Canada, Breitkreutz and colleagues found that two thirds of the teachers surveyed had not received training in pronunciation instruction during their teacher education. When Foote et al. (2011) ran a follow-up study using the same survey ten years later, they found that not much had changed in terms of preparation that ESL teachers in Canada receive in the area of pronunciation instruction. This finding in the Canadian context is also echoed in Europe. In their large-scale survey of 635 respondents in seven European countries, Henderson et al. (2012) found that European teachers of English in Finland, France, Germany, Macedonia, Poland, Spain and Switzerland have little or no training in teaching pronunciation. Sicola and Darcy (2015) further discuss that even if TESOL or foreign language learning programs incorporate some pedagogical training around pronunciation, it often comes in the form of an introductory phonology course on a theoretical level, in which practical application is kept to a minimum and at best limited to activities such as transcribing speech using the IPA. They explain that courses like this rather help the future teachers to get a deeper understanding of L2 phonology, but does not help deepening their understanding of how to teach L2 pronunciation. This argument is strengthened by recent findings in a study on second language teacher education in Australia, conducted by Burri, Baker, and Chen (2017). The authors traced how fifteen pre-service and in-service teachers’ belief and knowledge vis-à-vis pronunciation teaching changed during a postgraduate course on pronunciation instruction. They report that at the end of the course, particularly the pre-service teachers who did not have previous teaching experience, still reported not knowing how to teach pronunciation. Thus, it seems that one general course is not enough to make future language teachers comfortable with teaching pronunciation in the classroom (see also, Bøhn & Hansen, 2017).
Another interesting study, conducted by Wahid and Sulong (2013), found additional reasons for the chasm between practice and research in pronunciation teaching. In their paper, the authors used a mixed-method design consisting of a survey, teaching observations and semi-structured interviews to elicit responses from 27 ESL teachers concerning the extent to which pronunciation was taught, the teachers’ rationales underlying their practices, and how these rationales relate to pronunciation teaching research. In addition to echoing the well-attested finding that most teachers simply do not know how to teach pronunciation, one trend that the authors found was that the teachers did not know about any findings in academic pronunciation research or showed any interest in learning about these findings. Similarly to Derwing and Munro’s (2005) request to see more research conducted that actually relates to the classroom, Wahid and Sulong thus demand that research should serve teachers’ practical needs and appear in outlets that are accessible to them.

A good illustration of how pronunciation teaching can go wrong if teachers are not confident and knowledgeable in teaching pronunciation and do not have access to findings in pronunciation research is exemplified by X. Wang and Munro (2004). As a motivation for their perception-based study, the authors discuss that instructors may actually mislead learners in pronunciation matters, simply because they do not know better. One example they give is that, for one of the vowel contrasts investigated in their study (i.e., /ɪ/ versus /u/ as in words like ‘seat’ and ‘sit’), Chinese learners of L2 English are often wrongly taught that the difference between these vowels is one of length. The Chinese learners in this study were thus most likely unaware that these two vowels primarily differ in quality and can both be produced as long vowels. This assumption is supported by the fact that their teachers had not focused on the difference in vowel quality and, rather than discussing the relevant cue of tongue position, had only focused on a
discussion of vowel length. After the researchers corrected this approach by means of perception-based pronunciation training and instructed the treatment group on the difference, the learners improved in their production of both vowels.

Finally, other reasons, which have not been investigated in specifically designed research studies but are repeatedly mentioned in the discussion of why instructors do not teach pronunciation, include that teachers do not like to single out students in front of the entire class for pronunciation mistakes since this is often perceived as harsher than simply correcting a learner’s syntax, as well as that many language teachers are non-native speakers and thus feel insecure about the quality of their own pronunciation which in turn makes them hesitant to teach pronunciation (Derwing & Munro, 2015; Murphy, 2014).

After discussing the various reasons for the research-practice gap in pronunciation teaching, I want to conclude by highlighting the development in this field since Derwing and Munro’s (2005) influential article. Unfortunately, it has to be pointed out that the situation did not improve as much as one would hope. Yet, there is definitely a trend towards innovation in L2 pronunciation research to design studies that better inform classroom practice. In his 2016 article on how to translate pronunciation research into practice, for instance, Levis gives a clear outline of which pronunciation research is already well represented in teaching materials and which is not (Levis, 2016a). He then provides suggestions on how to include under-represented but highly useful research like advantages of High Variability Phonetic Training (HVPT) in the classroom (for a discussion of HVPT, see Section 2.2.2). Furthermore, the fact that two issues of the *Journal of Second Language Pronunciation* (JSLP) have focused on the interaction of research and pedagogy (Volume 2, Issue 1; Volume 3, Issue 1) shows that serious attempts are being undertaken to bridge the research-practice gap. In his 2016 editorial for the *JSLP*, Levis (2016b)
explains that future pronunciation studies need to focus on finding better techniques for pronunciation instruction, and showing in what contexts, with which learner populations, and under which time conditions these techniques should be applied. One good example for a study addressing such important questions is Gordon and Darcy (2016). In this classroom-based study, the authors tested the effectiveness of explicit pronunciation instruction in short term pronunciation practice. They set out by acknowledging that teachers have very limited time in the classroom to teach pronunciation and often do not know whether to teach segmentals or suprasegmentals if they are forced to prioritize. Thus, Gordon and Darcy designed a study that investigated specifically what would help to improve comprehensibility in short-term pronunciation training and found that it is most beneficial for learners to focus explicit training on suprasegmentals if only limited time can be devoted to pronunciation training. Findings like these are directly applicable to the classroom, and can help teachers decide how to prioritize their pronunciation interventions. More studies like this one are direly needed in the field of pronunciation research and, most importantly, researchers need to find a way to communicate these research findings to L2 teachers, that is, findings like these need to be published in outlets that are being read by teachers. Thus, there is still room for improvement and further practical research.

2.2 Prior Research in L2 Pronunciation Instruction

2.2.1 The Goal of Pronunciation Instruction

The discussion above has illustrated that pronunciation training should have a role in language classrooms and that teaching pronunciation generally leads to improvement in the learners’ pronunciation proficiency. When we talk about improvement in pronunciation skills,
however, it is important to define clear goals of what constitutes “good pronunciation”. A first step then is to explore the notions of the *nativeness principle* and the *intelligibility principle* (Levis, 2005). If an instructors’ teaching goals target the nativeness principle, they believe that it is desirable and achievable for their students to attain native-like command of the L2 phonological system. Teachers that target the intelligibility principle, on the other hand, follow the goal to improve their students’ intelligibility and thus believe that learners need to be understandable, but do not need to sound native-like. Pronunciation research in the new millennium has largely adhered to the intelligibility principle, especially once research findings had shown that native-like pronunciation is rarely achieved among adult L2 learners (see discussion in Section 2.1.1). When adopting the intelligibility principle as a goal for pronunciation instruction, it is important to discuss the tripartite distinction among *accentedness*, *intelligibility*, and *comprehensibility* that was first introduced by Munro and Derwing (1995) and has become the standard in defining goals in pronunciation research. Accentedness is often still used as a measure of nativeness and is a subjective measure that refers to how much a learner’s phonology differs from that of an L1 speaker of that language. It is frequently measured by listener ratings on a 9-point Likert scale, ranging from “no accent” to “very strong accent”. Comprehensibility is also a subjective measure, measuring how easy or difficult to understand a listener perceives the learner’s speech to be. It is usually measured on a 9-point Likert scale as well, ranging from “very easy to understand” to “very difficult to understand”. Comprehensibility is often distinguished from intelligibility, which is a more objective measure of the extent to which a speaker’s utterance is actually understood. Intelligibility can, for example, be measured if the listeners transcribe or translate an utterance spoken by an L2 learner and this notation is checked against the intended message (see also Bohn & Hansen, 2017;
Derwing & Munro, 2005; Foote & McDonough, 2017; O’Brien, 2014 for further discussion of the three categories). While these concepts are of course linked to each other in some ways, they are also very much independent—on the one hand, for instance, research has found that intelligibility and comprehensibility measures are usually highly correlated, but on the other hand even in the presence of a strong accent, L2 speech can still be fully comprehensible (Derwing & Munro, 2015). Several studies have addressed the question of which linguistic dimensions of L2 pronunciation are linked to accentedness and which are associated with comprehensibility. Most of these studies have been conducted in ESL learning contexts and have shown that for ESL, comprehensibility and intelligibility are linked to stress assignment, speech rate, pitch range, and pausing (Field, 2005; Munro & Derwing, 2001; Kang, Rubin, & Pickering, 2010; Trofimovich & Isaacs, 2012) whereas accentedness is linked to segmental accuracy, syllable structure errors, pausing, and non-native like rhythm (Derwing, Munro, & Wiebe, 1998; Kang, 2010; Trofimovich & Isaacs, 2012) (see also Saito, Webb, Trofimovich, & Isaacs, 2016 for lexical correlates of comprehensibility and accentedness, and Saito, Trofimovich, & Isaacs, 2016 for the varying influence of linguistic dimensions depending on learners’ ability levels). Only very recently has this question been addressed for languages other than English (see O’Brien, 2014 for L2 German and Bergeron & Trofimovich, 2017 for L2 French).

With the above-mentioned orientation towards the intelligibility principle, the goal of pronunciation instruction in the classroom has shifted towards improving learners’ comprehensibility and intelligibility in the L2. Based on personal communication with teachers, to a majority of them, it is now most important that their students are easily understood and can communicate effectively, conveying the message they intended. In fact, the focus on comprehensibility and intelligibility has become so strong that some pronunciation researchers
have made it a point to warn teachers and other researchers in the field of pronunciation instruction not to forget about the dimension of accentedness altogether. For example, Derwing and Munro (2015) included an entire chapter entitled “Social aspects of accent” in their book, and Levis (2016c) wrote an impactful editorial on the role of accentedness in pronunciation. These authors stress the key factor that learners themselves ask for pronunciation training that also targets their accentedness. Many learners want to improve their pronunciation in general, but particularly their accentedness. Reasons for this can mainly be found in the social implications of speaking with an accent. Levis (2016c) explains that a noticeable accent triggers certain questions in the interlocutor, whether this happens consciously or unconsciously. Questions like “Is this person like me?” or “How do I classify them within my social world?” call up (possibly even unconsciously) beliefs about the speaker. He further explains that we can see the social power of accent even in L1 environments, in which younger speakers usually take on the accents of their social group around them. Given that an accent provides us with an identity and social anchor, it comes as no surprise that foreign accents can also lead to discrimination in employment. Derwing and Munro (2015) give a very powerful summary of the social role of accents when they remind us that in movies, bad guys often have a heavy foreign accent, but we rarely see good people being played with foreign accents. Thus, it is understandable that our students still want to focus on improving their accents as well—a consideration we have to keep in mind when designing studies or teaching interventions that target only intelligibility or accentedness.

Finally, in the discussion of goals in pronunciation instruction, the principle of functional load plays an important role as well. Functional load is a measure of how much work two distinctive features do to distinguish words in a language (Munro & Derwing, 2006). For
example, the phonemes /n/ and /l/ have many minimal pairs in English, while /ð/ and /θ/ do not. Thus, the /n/–/l/ contrast is seen as having a higher functional load than the /ð/–/θ/ contrast (see also Levis, 2016a). Derwing and Munro’s (2006) study shows that the notion of functional load is valid and can be effectively used when thinking about goals in pronunciation instruction.

2.2.2 The Relationship between Perception and Production

While many studies have shown improvement in L2 pronunciation skills by focusing on various forms of oral practice (Couper, 2006; Elliott, 1995, 1997; Gordon, Darcy, & Ewert, 2013; Lord, 2005; Saito, 2011a, 2011b; Saito, 2013; Saito & Lyster, 2012a, 2012b), there is also a multitude of studies that have addressed pronunciation training through perception training rather than (or sometimes in addition to) production-based training (see e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lacabex, Lecumberry, & Cooke, 2005; Lambacher, Martens, Kakehi, Marasighe, & Moldhold, 2005; Rochet, 1995; Thomson, 2011; Y. Wang, Jongman, & Sereno, 2003; X. Wang & Munro, 2004). In this section, I want to discuss why this is the case.

The perception-centered approach to pronunciation training is based on the finding that learners’ L2 perception abilities are linked to their L2 production skills. This link is explored in existing models of L2 speech learning such as the Perceptual Assimilation Model (PAM; Best, 1995; Best & Tyler, 2007) or the Speech Learning Model (SLM; Flege, 1995, 2002, 2007). Best’s (1995) PAM as well as Best and Tyler’s (2007) update of the PAM (i.e., the PAM-L2) explain adult learners’ difficulties with non-native speech sounds with the fact that adults are

---

2 The PAM was originally not developed as a model of L2 learning but rather as a model to describe ways in which adults perceive non-native phones. The PAM-L2 is an update to the PAM that accounts for L2 phonetic learning over time—making the PAM more comparable to Flege’s SLM.
likely to perceptually assimilate the non-native sounds to the most similar native sound, based on similarity in articulatory gestures in their L1 phonological system. For example, when two L2 sounds are judged as the same L1 sound, the PAM would call this a *Single Category Assimilation*, which hinders the discrimination between the two sounds. A more widely adopted model in pronunciation research, however, is Flege’s SLM, a model that was directly developed for L2 learning—as compared to the PAM that was developed as a means to explain the ways in which adults perceive speech sounds with which they had no prior experience. The SLM also takes age-related constraints into consideration and is based on the assumption that the perceptual mechanisms for learning new sounds are preserved after childhood, but that it is difficult to access these mechanisms later in life (see discussion of the CPH in Section 2.1.1; see also Thomson, 2012). According to the SLM, adult learners’ difficulties in learning L2 sounds can then be anticipated on the basis of how similar sounds are in the learners’ L1 and L2. If a sound is identical in both languages, immediate correct production of the L2 phoneme is predicted due to direct transfer from the L1. Further, if a new L2 sound is very dissimilar from any phoneme in the learner’s L1, the learner will develop a new category for this phoneme, a process that might vary in time across learners and categories, but should ultimately lead to correct production of the L2 sound. The most difficult to acquire, according to the SLM, are L2 sounds that are similar to a sound in the learner’s L1, but not identical to this sound. These phonemes cause difficulties because the similarity might lead learners to falsely associate a sound in the L2 with a similar L1 category, when really an entirely new phonological category has to be built. These sounds then take the longest to learn, if they are learned at all. The SLM further proposes that production problems often have their roots in inaccurate perception because learners might produce a target sound with the acoustic qualities of the incorrect L1 category.
(Flege, 1995, 2002, 2007). Such an error can only be corrected if the learners build new perceptual categories for the L2 phonological space, which in turn explains why several pronunciation studies focus on improving learners’ L2 perception first, in order to ultimately improve their oral production skills (see e.g. Botero, 2011; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Counselman, 2010; Lacabex, Lecumberry, & Cooke, 2005; Lambacher, Martens, Kakehi, Marasighe, & Moldhold, 2005; Rochet, 1995; Thomson, 2011; Y. Wang et al., 2003; X. Wang & Munro, 2004).

Several studies have been able to show that perception training can improve learners’ pronunciation skills even in the absence of production training (Bradlow et al., 1997; Lambacher et al., 2005; Rochet, 1995; Y. Wang et al., 2003). In Bradlow et al.’s (1997) well known study, Japanese learners of L2 English received a four-week course in perceptual training of the /r/-/l/ contrast—a contrast that has been proven to be challenging for L2 learners from many Asian L1s. After 45 sessions with multiple talker tokens and feedback, the learners showed improvement on their success rate in identification for the /r/-/l/ contrast. Moreover, native speaker raters judged the learners’ oral productions, which were recorded before and after the perception training, on whether the pretest or posttest token was clearer to understand, and they indicated in a binary-choice-task which word was being produced. The authors found that the gains that learners had made during the perception training had transferred to their oral production capabilities as well. That is, at the end of the perception training, the participants’ productions skills for the /r/-/l/ contrast had improved significantly, while the learners in the control group did not show a statistically significant gain. Similar findings were published by Rochet (1995) with a focus on improvement of voice onset time and by Y. Wang et al. (2003),
who built on Y. Wang, Spence, Jongman, and Sereno (1999) and showed that improvement in tone identification led to improvement in tone production.

After research had demonstrated that improvements in perception abilities usually lead to improvements in production abilities, the next question that researchers addressed was, what kind of perceptual training would be most effective in understanding L2 speech? Findings addressing this question suggest that training learners to perceive sounds that are produced by multiple speakers and in varying phonetic contexts results in significantly greater improvement than perception training that simply relies on a single talker or a single phonetic context. This type of training is referred to as high variability phonetic training (HVPT). HVPT input better represents the range of variation in natural speech and allows learners to be better at identifying speech sounds when they are produced by new speakers. Thomson (2011) explains that a reason for this lies in the fact that learners who are trained to identify a sound from just one single voice may not attend to the relevant cues for the sound, but might rather rely on cues that help them distinguish between the training stimuli. Oftentimes, however, these cues only pertain to an individual speaker and are not generalizable. That is, phonological categories are highly variable clusters of sounds that share several properties or cues. It then takes extensive exposure before the cues for a sound category can be correctly recognized and stored (Thomson, 2012). Logan, Lively, and Pisoni (1991) and Lively, Logan, and Pisoni (1993) already began testing HVPT in the early 1990s, and they indeed found that increasing the variability of the input resulted in greater and more generalizable gains in L2 speech perception abilities. Logan et al. (1991)

---

3 It is worth noting here, however, that perception does not always precede production. Some studies have shown that learners can produce certain phonemes quite well without being able to hear the difference between them. Such cases can, for instance, occur when learners are being taught with a focus on articulatory parameters rather than auditory cues or are being trained to focus on kinesthetic feedback regarding their speech production (Derwing & Munro, 2015; Sheldon & Strange, 1982; Smith, 2001)
trained six Japanese learners of English on perception of the /r/-/l/ contrast. After presenting 207 minimal pairs in which /r/ and /l/ contrasted in word-initial and word-final position—collected from six native English speakers—to their learners, they found “small but reliable” (p. 874) gains in performance. While Logan et al. (1991) had only shown a small effect for the advantage of varying input, in their follow-up study, Lively et al. (1993) found a significant improvement in token discrimination after using a similar training method. This time, the study design included a control group in which a group of learners received input that only came from a single talker. They found that while the learners in the control group made some gains on the trained items, they were not able to generalize these skills to tokens produced by new talkers, whereas the HVPT group was able to generalize their skills to novel tokens, thus showing the benefit of highly variable input. The benefits of HVPT have since been confirmed in many subsequent studies (for consonants: Bradlow et al., 1997; Iverson, Hazan, & Bannister, 2005; for vowels: Iverson & Evans, 2009; Lambacher et al., 2005; Nishi & Kewley-Port, 2007, 2008; Thomson, 2011; for tone: Y. Wang et al., 1999; Y. Wang et al., 2003; for syllable structure: Huensch & Tremblay, 2015; Huensch, 2016). Huensch (2016) points out that while previous studies have shown that perceptual training can lead to improvements in production and that these improvements can generalize to new words and talkers, it has not been shown whether these improvements also transfer to larger discourse contexts and free speech. She therefore investigated this question in an experiment with Korean L2 learners of English, and found that learners were indeed able to generalize production improvements to continuous speech. This finding suggests that the gains from HVPT translate to real life situations and that it is worth including HVPT in classroom instruction. It has to be pointed out that HVPT is a great addition to classroom-based instruction in general. Traditionally, classroom-based instruction only has
one voice, the voice of the instructor, but incorporating HVPT allows for learners to be trained with the input of multiple voices (see Levis, 2016a for a summary of the role of HVPT for practical pronunciation instruction, and see Section 2.2.5 for a discussion of computer assisted pronunciation training).

2.2.3 The Scope of L2 Pronunciation Training: Segmentals and Suprasegmentals

Considering the time constraints on pronunciation training in classroom settings, it is important to make pronunciation instruction as time efficient as possible, and thus teachers need to know what goals to set in their instruction. To date, for ESL, German or any other language, no study has provided a definitive “map to pronunciation instruction”, that is, a guideline for teachers of what exact sounds or prosodic feature to teach in what order, and at what proficiency level. While the above-mentioned goals of intelligibility, comprehensibility, and accentedness as well as the priority of high functional load contrasts inform this debate, it further helps teachers to have a guideline of whether to focus on segmental or suprasegmental features when designing pronunciation interventions. Training on the segmental level targets individual speech sounds—consonants, vowels, and their language-specific combination—whereas suprasegmental training focuses on prosodic elements like stress, rhythm, intonation, pausing, and tempo.

Early pronunciation studies usually investigated how to improve learners’ production of individual sounds, but starting in the 1990s, this approach began to change with a (temporary) rise in interest in suprasegmental pronunciation training. Derwing et al. (1998) was one of the first and is probably the most influential study comparing the relative effectiveness of pronunciation instruction on segmentals and suprasegmentals (but see also Anderson-Hsieh, Johnson, & Koehler, 1992). In this study, the authors tested ESL learners over a period of 12
weeks on sentence reading and free speech productions. Learners were recruited from three courses: one course focusing on segmentals, one course focusing on suprasegmentals, and a control group that did not receive explicit pronunciation training. While students who had received segmental training only improved significantly on the read-aloud task, students in the suprasegmental group showed significant improvement on comprehensibility and fluency, both in free speech and on the reading task. The authors thus concluded that the suprasegmental training was more effective since the acquired skills can be transferred to free speech. The years after 1998 then saw an increase in studies focusing on suprasegmental pronunciation training (Hahn, 2004; Hardison, 2004; O’Brien, 2004, see also Chun, 2002 for a comprehensive review). Thomson and Derwing’s (2015) narrative review of 75 pronunciation studies showed that 53% of the research focused on segmentals, 23% on suprasegmentals, and 24% on both types of features in one study. The value of instruction on suprasegmentals is further underscored by their influence on comprehensibility and accentedness (Gordon & Darcy, 2016; Isaacs & Trofimovich, 2012). It has to be pointed out, however, that in their 2015 book, Derwing and Munro write “[…] we never suggested that teachers should restrict themselves to one of the kinds of instruction that we used in the study. In fact, we expressly stated that both segmentals and suprasegmentals have a place in the L2 pronunciation classroom” (p. 9). This is in line with most recent calls in the pronunciation literature that ask for a combination of instruction in segmental and suprasegmental training. In his synthesis of 15 quasi-experimental intervention studies, Saito (2012) found that studies providing training on both segmental and suprasegmental features usually lead to gains. Finally, in her encyclopedia article on pronunciation instruction, Derwing (2013) recommends addressing both types of features in pronunciation interventions. She explains that suprasegmental features seem most helpful in improving intelligibility in
spontaneous speech production, whereas focusing on segmental features can be valuable when communication breaks down, allowing L2 learners to focus on repairing their strongest pronunciation problems in a targeted fashion.

### 2.2.4 Ecological Validity in L2 Pronunciation Teaching: Lab-based and Classroom-based Studies

In their narrative review of the effectiveness of L2 pronunciation instruction, Thomson and Derwing (2015) state: “To address concerns regarding ecological validity⁴, the ideal study should be conducted in a classroom—although laboratory research can be extremely informative.” (p. 327). This leads to questions about the relationship between classroom-based and lab-based studies and how each of these contribute to the field of L2 pronunciation research and teaching.

Many of the studies in the first wave of L2 pronunciation studies in the 1980s and 1990s were lab-based studies. That is, learners for these studies were specifically recruited for the purpose of these studies and trained under laboratory conditions, rather than researchers using intact L2 classes and carrying out a training in the classroom (e.g., Flege, 1988; Flege, 1989; Logan et al., 1991; Lively et al., 1993). Conducting a lab-based study allows researchers to address research questions in a much more homogenous group because learners can be recruited based on a specific constellation of features—e.g., a homogenous L1 background, a certain level of motivation, a certain proficiency level, or even a certain age, gender, learning style, etc. This is usually not possible when conducting research with intact classes where students are often of various L1 backgrounds, have different levels of motivation and commitment, show unreliable attendance, etc. Moreover, lab-based studies sometimes allow the use of methods that would not

---

⁴ Ecological validity refers to the extent to which the findings of a research study are able to be generalized to real-life settings (“Ecological Validity in Psychology”, 2017)
be possible in a classroom setting. For instance, Logan et al.’s (1991) and Lively et al.’s (1993) studies, which have been discussed above in the context of HVPT, used a computer to deliver the perception training with multiple voices to the learners, which would not have been feasible in a classroom in the early 1990s, when the researchers collected their data. Lab-based research, however, did not only occur in early pronunciation research, but is still used today to address research questions that cannot be addressed in a classroom setting. Such studies have contributed important insights to the field of L2 pronunciation over the past decades (e.g., Iverson & Evans, 2009; Iverson et al., 2005; Guion & Pederson, 2007; Hardison, 2004; Saito, 2011b; Thomson, 2011). Iverson et al. (2005), for instance, set out to compare the effectiveness of different training techniques within HVPT. That is, they investigated the role of altering cue weightings in perceptual input by comparing four conditions in HVPT: 1. natural voices, 2. synthetically altered stimuli in which F3 values on English /r/ as well as closure duration was enhanced, 3. stimuli that were altered as described in 2. but in which the enhancement faded a little more every day over the course of the ten-day training period, and 4. synthetically altered stimuli in which all secondary cues had been set to median values, rendering these secondary cues unreliable to learners. Recruiting a homogenous group of Japanese adult EFL learners and randomly assigning them to one of the four conditions allowed the researchers to investigate whether there was a learning benefit in synthetically enhancing HVPT stimuli. Their results showed that all four techniques improved learners’ abilities to distinguish between /r/ and /l/—with no statistically significant differences in the outcomes across the techniques. This suggested that it is not necessary to alter input in HVPT but rather that natural voices can be used. It would have been difficult to implement the materials and methodology used in this study in classroom-
based research, which shows that lab-based research still contributes unique insights to the field of L2 pronunciation instruction.

Nevertheless, Thomson and Derwing (2015) are not alone in voicing their concern about the ecological validity of lab-based research. Many other researchers also insist that the field of L2 pronunciation should work on conducting research in more ecologically valid contexts with the thought in mind that teachers are much more likely to use findings from classroom-based studies in their own teaching because the results seem most authentic. That is, if a pronunciation intervention showed success in a real classroom setting, teachers might be more willing to try it in their own classrooms (Derwing & Munro, 2015; Gooch et al., 2016; Gordon & Darcy, 2016). And, in fact, a multitude of classroom-based studies have been carried out in the past 25 years. These studies investigated many different L2s and instructional goals and were carried out in classrooms all over the world (targeted languages: Spanish: Botero, 2011; Counselman, 2010; Elliott, 1995, 1997; Kissling, 2013; Lord, 2005; 2008; English: Couper, 2003, 2006; Derwing et al., 1998; Gooch, Saito, & Lyster, 2016; Gordon & Darcy, 2016; Neri et al., 2008; Saito & Lyster, 2012b; German: Ducate & Lomicka, 2009; Roccamo, 2014; French: Ducate & Lomicka, 2009; focus of pronunciation instruction: segmental: Botero, 2011; Counselman, 2010; Elliott, 1995, 1997; Gooch et al., 2016; Kissling, 2013; Lord, 2005; Saito & Lyster, 2012b; both segmental and suprasegmental: Couper, 2003, 2006; Derwing et al., 1998; Ducate & Lomicka, 2009; Gordon & Darcy, 2016; Lord, 2008; Neri et al., 2008; Roccamo, 2014).

The discussion above has shown that both lab-based and classroom-based pronunciation studies have a place in L2 pronunciation research. Thomson and Derwing (2015) report that out of the 75 studies included in their comprehensive review, a majority of 61% were classroom-
based, which might be a consequence of the fact that classroom-based research appears to be more ecologically valid than lab-based research. When looking at effect sizes for these studies, however, the picture is not entirely clear. That is, for within-group contrasts, J. Lee et al.’s (2015) meta-analysis reports an effect size of $d = 0.95$ for interventions in classroom-based studies and an effect size of $d = 0.84$ for interventions in lab-based studies, but for between-group contrasts, lab-based studies with an effect size of $d = 0.95$ seem to be more effective than classroom-based studies with an effect size of $d = 0.79$. This suggests that, for the comparison of pre- to posttest, learners make bigger gains in classroom-based pronunciation instruction than in lab-based interventions, but usually the more important comparison—the comparison between a treatment and control group—suggests that learners make greater gains in a lab-based environment. Nevertheless, it is important to keep in mind that the research question should determine the methodological choice of whether a lab-based or a classroom-based environment is the better place to carry out research. That is, not only classroom-based research can have implication for classroom learning. While studies with intact L2 classes might seem to provide the highest degree of ecological validity, even classroom-based studies are usually not designed to be exactly replicable, but rather largely depend of factors like teaching context (secondary vs. post-secondary context, foreign language vs. second language context), target language, and the curriculum (course materials, time frame etc.). In sum, the field of L2 pronunciation research has certainly benefitted from both types of studies.

2.2.5 Using Technology to Teach Pronunciation

The fast-paced development of technology over the past two decades has changed the landscape of language learning and teaching (Hsu, 2016). The wide range of technological
applications—such as virtual learning environments, internet-based media like YouTube videos, instant messengers, VOIP tools (e.g., Skype), blogs and wikis, online reference tools, online translators, social networking spaces (e.g., Facebook and Twitter), as well as mobile devices with their wide array of apps for purchase—allow a completely new approach to teaching foreign languages in general, and teaching pronunciation in particular. With the advent of such a multitude of possibilities in teaching with technology, researchers and teachers have been faced with questions of if and how these new technologies may be applicable in the L2 classroom. Thus, computer-assisted language learning (CALL) and computer-assisted pronunciation training (CAPT) have been a primary or secondary focus of many recent studies (Botero, 2011; Ducate & Lomicka, 2009; Eskenazi, 1999a, 1999b; Hardison, 2004; Hincks, 2003; Hincks & Edlund, 2009; Hirata, 2004; Hsu, 2016; Kawai & Hirose, 2000; Lord, 2008; Menzel et al., 2001; Neri et al., 2008; Olson, 2014; Patil & Rao, 2016; Ruellot, 2011; Seferoglu, 2005).

Before discussing the different findings in some of these studies in more detail, I want to address advantages and disadvantages of CAPT more generally. The reason for the rise in interest in CAPT research is largely based on the fact that CAPT seemingly solves several problems that occur in traditional learning environments. First of all, CAPT can serve as a solution to the constant time restrictions on pronunciation training that teachers face in L2 curricula: CAPT offers a great possibility to move pronunciation training outside of the classroom since it usually only requires access to a computer, but not the presence of a teacher. Taking pronunciation instruction outside the classroom further creates a stress-free environment in which learners can work through the pronunciation exercises at their own pace, and with nearly unlimited input. Not only does such self-paced pronunciation training outside the classroom promote learner autonomy, it also allows for a much more individualized instruction.
That is, even learners from a shared L1 background do not always exhibit the same pronunciation problems. In a classroom, teachers usually have to take a one-size-fits-all-approach by focusing training on the most common pronunciation problems, but they do not have a chance to provide individualized training for each learner’s problem area in L2 phonology. CAPT then allows for training that can address individual learners’ problems in a more targeted fashion because learners can choose to spend more time on a sound or suprasegmental aspect that they personally have problems with, and skip over other areas that they have already mastered. This, however, also entails the first possible obstacle inherent to CAPT. As shown by Dlaska and Krekeler (2008), learners are often unable to identify which specific areas in their speech cause communication breakdowns or problems (see discussion in Section 2.1.1). Thus, CAPT still needs to be guided by a knowledgeable teacher—otherwise it is possible that students zoom in on the wrong features and, despite training, will not improve their pronunciation. Moreover, several researchers warn that some CAPT provides incorrect feedback because the technology is not good enough to pick up on the learners’ productions—a problem that is particularly relevant in the case of automatic speech recognition software (Derwing & Munro, 2015; Hsu, 2016; Neri et al., 2002; O’Brien, 2011). This can be problematic in two ways: either learners receive feedback that their incorrect productions are correct and they therefore put no further effort into improving their pronunciation, or learners receive the feedback that their correct productions are incorrect, which is frustrating and might lead to students altering their correct productions until they are less intelligible. Finally, it also has to be pointed out that a majority of CAPT software is programmed for commercial use and is not always affordable for schools or learners, or even pedagogically sound.
When thinking about advantages and disadvantages of CAPT, however, it is important to keep in mind that there are considerable differences between the different types of programs, applications and software, and not all pros and cons apply to each of these types equally. In her comprehensive review of CAPT literature, O’Brien (2011) distinguishes between three types of CAPT software: basic pronunciation training, automatic speech recognition (ASR) software, and visualization software. The first category, basic pronunciation training, is similar to classic language laboratories. That is, learners often hear native speaker recordings, repeat what they hear, and finally record their own productions. While this type of training is much easier to realize nowadays with the use of computers and free digital recording software (e.g., Audacity), this type of pronunciation training is usually not very communicative. Nevertheless, the findings on HVPT discussed in the previous sections show that basic pronunciation training definitely has its place in modern pronunciation instruction and can be a great addition to traditional in-class pronunciation training (for an example of what HVPT could look like as an easily accessible app, see http://www.englishaccentcoach.com/). This is also true for several of the multitude of pronunciation apps that are available for mobile devices and personal computers. Some of these apps have been addressed in research (see Hsu, 2016 for a partial review) and others have simply held up in everyday classroom experience (for German, e.g.: the Sounds of Speech app from the University of Iowa that provides two-dimensional representations of the vocal tract, http://soundsofspeech.uiowa.edu, as well as the Fluent Forever Pronunciation Trainer that teaches German sounds with the use of interactive flashcards, https://fluent-forever.com/pronunciation-trainers/german/).

While programs and apps that fall into the category of basic pronunciation training can be beneficial for learners in improving specific, but rather small-scale, pronunciation problems, the
development of modern technology has enabled CAPT to go one step further, that is, finding a way for communication between a computer and the learner to take place. The development of automatic speech recognition (ASR), which is “software that captures, recognizes, and reacts in some way to human speech” (O’Brien, 2011, p. 384), has made it possible for learners to engage with simulated native speakers. A good example for ASR software is Auralog’s *Tell me More* (*Auralog, 2008*). In *Tell me More*, the computer produces a sentence prompt to which the learners have several possible answers. If the computer recognizes the learner’s response, the program gives feedback on the learner’s production and continues the simulated dialogue with an appropriate answer and follow-up question. If the computer does not recognize the learner’s utterance as a meaningful production in the L2, the program provides clues for improving the pronunciation of the target sentence. While this sounds like a very promising technological advancement that could play a big role in language acquisition and pronunciation training, the reality is that, unfortunately, ASR exhibits many problems. First and foremost, ASR often provides false automatic feedback when used in L2 contexts. An explanation for this lies in the fact that ASR was originally invented and programmed to convey a command to a mobile device or for dictation purposes, and not for foreign language acquisition. Put simply, ASR works by employing algorithms that decode speech into word and phone sequences and provide acoustic likelihood scores that are matched with speech models based on native speakers (Patil & Rao, 2016). Given that our goal in pronunciation training is improving learners’ intelligibility and comprehensibility rather than achieving a native-like control of the L2, however, these algorithms that are based on native speaker models often fall short in L2 ASR. One might think that a solution would be to base the algorithms on models of L2 speakers rather than native speakers, but Patil and Rao (2016) explain that there are two problems with this approach: 1.
non-native speaker input is too variable to serve as reliable input for speech models (particularly accounting for multiple L1s) and 2. since phone errors in the L2 typically involve substitutions by very similar phones borrowed from the learner’s L2, the required phone discrimination from the acoustic signal is even harder to be picked up by the traditional algorithms. As mentioned above, false automatic feedback can be detrimental to L2 learners. Therefore, most researchers who investigated ASR software in L2 learning contexts come to the conclusion that it usually fails to meet pedagogical requirements (Derwing & Munro, 2015; O’Brien, 2011; Olson, 2014). Nevertheless, Derwing and Munro (2015) point out that ASR has been used successfully in the past and is valuable for the classroom as long as the system has been carefully calibrated to meet the needs of the learners. One study that has demonstrated a successful implementation of ASR in the classroom is Neri et al. (2008). Neri and her colleagues investigated an ASR software called PARLING, which was designed for the purpose of English instruction in Italian elementary school classrooms. This software analyzes recordings in real time and gives ASR feedback. The authors compared pronunciation gains in a group that received teacher-fronted instruction with the gains in a group that was instructed using PARLING, and found that both groups made significant improvement on pronouncing isolated words. They therefore conclude that PARLING was as helpful as in-class pronunciation training, but point out how much time can be saved by using PARLING rather than in-class training (for additional ASR studies, see also Eskenazi, 1999a, 1999b; Hincks, 2003, Kawai & Hirose, 2000; Menzel et al., 2001).

Finally, the third type of CAPT—visualization software—provides a visual representation of speech, most often in the form of a spectrogram or simply with programs that plot pitch accent (e.g., Praat, Boersma & Weenink, 2017). These programs usually allow learners to see how their pronunciation deviates from native speaker norms. Tell me More, for
instance, not only includes an ASR component, but it also creates spectrograms and pitch contours of learners’ utterances. Learners can then compare their productions with the native speaker target and repeat their recordings until they approach the native norm. Many studies have found pronunciation interventions based on some form of visualization to be successful (Hincks & Edlund, 2009; Hardison, 2004; Hirata, 2004; Offerman & Olson, 2016; Olson, 2014; Ruellot, 2011; Seferoglu, 2005). Most of these studies have targeted instruction of prosodic features. Olson (2014) justifies this by explaining that prosodic features represent an ideal starting point for the pedagogical use of visualization software as even early speech software was able to produce reliable representations of intonation contours and interpretation of these contours proved to be intuitive and easy even for untrained learners. He correctly observes, however, that most of the research on the use of visualization software has been carried out in lab-based studies. Thus, Olson (2014) and Offerman and Olson (2016) set out to show how pronunciation training with the use of visualization can be implemented in L2 classrooms. They further show that this type of training also works for segmentals (not only for suprasegmentals like prosody), that it works for lower-level classrooms as well as for more advanced language learners, and that benefits extend to continuous and spontaneous speech. Nevertheless, Derwing and Munro (2015) point out that correctly interpreting spectrograms requires considerable expertise in phonetics. Thus, if visualization software is used in language classrooms, it usually comes in the form of a commercial visualization software that helps with the interpretation of the spectrograms or visualization of pitch contours. It then has to be kept in mind that these programs are usually not inexpensive.

To sum up the overview of the role of technology in pronunciation instruction, it can be said that the benefits of technological advances for pronunciation learning are exciting, but that
instructors have to be aware of possible flaws in each type of CAPT. This is also reflected in J. Lee et al.’s (2015) meta-analysis, which actually shows lower effect-sizes for the pronunciation intervention when technology was used ($d_{tech} = 0.76$ vs. $d_{no\ tech} = 0.96$ for the within-group contrast and $d_{tech} = 0.53$ vs. $d_{no\ tech} = 0.87$ for the between-group contrast). Derwing and Munro (2015) provide a perfect summary when they say that “it is crucial for teachers to treat technology as one tool among others to enhance their learners’ L2 pronunciation. To do this effectively, the teacher must have a good understanding of the foundations for pronunciation research and the pedagogical knowledge to exploit the benefits that technology has to offer” (p. 130).

2.3 Pronunciation Instruction in Distance Language Learning

While some educators still seem to question the viability of distance language learning (Cheng, 2015) or simply fear that technological developments in online learning will replace their role in the classroom, leaving them without employment (R. Blake, 2014), it is an undeniable fact that online learning is on the rise. According to the Online Report Card—the 13th annual report on the state of online learning in U.S. higher education—distance education enrollment exhibited a growth of 7% between 2012-2014 (Allen & Seaman, 2016; all data in the 2016 report were based on federal data from fall 2014, i.e., the most recent available year). Federal data further show that campus-based enrollments are declining, while many institutions are continuing to add online learning programs. Allen and Seaman (2016) report that more than 2.8 million students in the US are taking all of their higher education instruction at a distance, representing 14% of all higher education students. Furthermore, an additional 2.97 million are currently taking some of their classes at a distance, which means that more than a quarter of all
students in U.S. higher education have online learning experience by the time they graduate.
Enrollment numbers from The Pennsylvania State University’s World Campus reflect the trends reported in the *Online Report Card*: enrollment in Fall 2016 totaled 13,411 students, up from 12,242 the year before. As a comparison: University Park enrollment totaled 47,261 students with an increase of only 413 students from fall 2015 (“University releases annual enrollment snapshot”, 2016). This growth in online learning programs and enrollments comes as no surprise when considering the many advantages online learning provides, such as the availability of classes for people who need flexible access to instruction (e.g., because they work full-time or live in a remote area of the U.S. or the world) or, in the context of online language learning, the opportunity to offer classes in less commonly taught languages and even classes in more common languages but in rural regions, in which there is a shortage of language teachers (R. Blake & Delforge, 2007; R. Blake, Wilson, Cetto, & Pardo-Ballester, 2008; Cheng, 2015). Yet, despite the growing interest in online education and the obvious advantages of distance education, almost every single publication on this subject stresses that further research is needed. Moreover, according to R. Blake, Wilson, Cetto, and Pardo-Ballester (2008) and Cheng (2015), only a very small fraction of research on online education has investigated online language learning. R. Blake et al. (2008) explain this dearth of studies on distance language learning with the fact that there are simply fewer students to sample from. They further observe that it is a lot more difficult to convince online students to join a research study as students cannot be “cajoled in person to complete research questions” (p. 116). Another reason why online language learning has been relatively under-researched in the past is that findings cannot be generalized as easily as findings in face-to-face classroom studies, because online learning exhibits larger differences in curriculum design, as well as choice and use of technology. In fact, even the definitions of online
learning and distance learning exhibit a certain degree of variability, which makes it hard to replicate findings or to design studies based on previous findings. Therefore, in the following, I will provide working definitions that will be used in the present study.

While some previous studies operationalized the term *online language learning* as referring to a variety of learning contexts such as web-facilitated classes, blended or hybrid classes, or fully virtual online classes (see R. Blake 2011, for a review), the present study will adopt the definition provided in *Keeping Pace*—the annual review of policy and practice in K-12 digital learning compiled by the Evergreen Education Group—and only refer to fully virtual classes as online learning (Dobrovolny, Edwards, Friend, & Harrington, 2015). Online language learning is then synonymous with *distance language learning*, which White (2006), in her thorough review of more than three decades of research in distance language learning, defines as having the following features: “the separation of the teacher and learners, the use of technical media, provision of two-way communication, and the influence of an education organization, distinguishing it from private study” (p. 248). While it is true that many years ago, distance language learning was not necessarily web-based but relied on cassette-tapes and books sent via standard mail, I am not aware of any study in the past twenty years that investigated a type of distance language learning that was delivered without making at least some use of the internet. Since none of the studies discussed for the present dissertation focus on such an archaic type of distance language learning, in the following, distance language learning and online language learning will thus be used interchangeably. Moreover, studies reporting results from *blended language learning*/*technology enhanced language learning* (i.e., using technology for certain exercises as a supplement to classroom instruction, e.g. in homework activities) or *hybrid language learning* (i.e., providing instruction both in class and online) will not be discussed here.
since their results are not generalizable to fully virtual online learning as investigated in the present study. Finally, in the field of online language learning, it is important to define the terms of *synchronous* and *asynchronous communication*. In synchronous online learning, the learners communicate with each other or the instructor in real time (e.g., in real-time chat environments or voice/video-based communication such as Skype). In asynchronous online learning, the learners do not interact in real time with each other or with the instructor but rather by deferred-time communication such as via email or message boards (R. Blake, 2014).

Early research in online learning focused mostly on comparison studies, trying to answer the question whether online learning could be as effective as the more traditional face-to-face (F2F) learning. For the time being, this question is considered to be settled after a meta-analysis of more than 500 online learning studies published between 1996 and 2007, commissioned by the U.S. Department of Education, found that students in online learning environments performed modestly better than those learning the same materials in F2F learning environments (Means, Toyama, Murphy, Bakia, & Jones, 2010). Such a meta-analysis, however, has not yet been published specifically for the context of online language learning. In the field of online language learning, early research has mostly addressed the roles of written communication, grammatical development, or literary skills (Cahill & Catanzaro, 1997; Chenoweth & Murday, 2003; Lamy, 2004). Only more recently has research started to address oral proficiency development as well, even though Deutschmann et al. (2009) point out that addressing oral skills is one of the biggest challenges in online language learning. They explain that even audio files cannot replace the need for interaction that leads to oral proficiency in the L2. R. Blake (2008), however, argues that “research on computer-mediated communication (CMC) has answered this concern by demonstrating that electronic interactions offer benefits much like face-to-face
classroom exchanges” (p. 365). CMC can vary in being based on simple text exchanges, text and audio exchanges, or even text and video exchanges (R. Blake, 2014), it can be synchronous or asynchronous, but either way, it usually replaces the face-to-face communication that takes places in traditional classrooms. Lin (2014) explains the success of CMC technologies with the fact that this type of communication creates a social context in the virtual learning environment that appears to be beneficial for the cognitive development that underlies L2 learning. This is in line with Chenoweth and Murday (2003), who argue that synchronous written chat, for example, serves as a low speed conversation that allows students time to review the conversation visually and to notice their own mistakes (see also Abrams, 2003). Similarly, Payne and Whitney (2002) conducted a study in which they tested the hypothesis that synchronous CMC would indirectly improve learners’ oral proficiency by developing the same cognitive mechanisms underlying spontaneous speech. Addressing this hypothesis in a study with third-semester students of L2 Spanish, they indeed found that their treatment group, which received two of four contact hours per week in a synchronous text-based chat room, significantly outperformed a control group in measures of oral proficiency. The authors suggest that this finding might be explained by the fact that the decreased speed of conversation reduces the memory load for the learners and allows more time for pre-task planning. Furthermore, they argue that the non-ephemeral nature of a text-based chat might differently affect learners’ cognitive development in the L2. In later years, the benefits of CMC have been explored for entirely virtual online language learning. R. Blake and his colleagues (2008), for instance, explored the effectiveness of an online Spanish course in improving learners’ oral proficiency. Their study compared the level or oral proficiency of a F2F, a hybrid, and an online class at the end of a first semester Spanish course. All participants took a 20-minute Versant for Spanish test as an evaluation of their oral proficiency, with the goal
of showing that the learners in the distance education environment did not fall behind their classroom counterparts. Results showed that all learners performed at a comparable level of oral proficiency at the end of the semester. Thus, the authors concluded that the online learners’ oral proficiency developed on par with the F2F and hybrid learners’ proficiency (see also C. Blake, 2009 and Isenberg, 2010). Finally, a meta-analysis of 25 CMC studies published between 2000 and 2012 revealed that CMC produced a small positive effect compared to F2F learners (g = 0.40)—17 studies showed a positive effect and eight studies showed a negative effect (Lin, 2014). The author further reports that, among the 25 studies looking at effects of CMC interventions on the acquisition of oral proficiency, five assessed oral competences holistically, five at the lexical level, eight at the syntactic level, seven assessed either accuracy or fluency of oral performance, and only four studies targeted pronunciation as a measure for oral proficiency. This finding emphasizes the dearth of research that addresses pronunciation development in online language learning. In fact, it has to be pointed out that these studies focused on the role of CMC in oral proficiency and only looked at pronunciation as a component of oral proficiency, but they did not investigate ways of teaching pronunciation in distance language learning. To my best knowledge, currently there is no study that explored the inclusion of particular methods to teach pronunciation in online environments. That is, considering that previous research has shown that overall oral proficiency can be improved through CMC and that oral competencies of online learners are comparable to those of F2F learners, future research should focus on exploring ways to help students improve a crucial component of oral proficiency—pronunciation—as effectively as possible.
2.4 The Present Study

2.4.1 The Method of iCPRs

2.4.1.1 CPRs vs. iCPRs

The method of Cued Pronunciation Readings (CPRs) was first investigated by McCardell Landon (2007) in an M.A. thesis. The results of this study were published in Language Learning & Technology (Tanner & Landon, 2009) and inspired two additional M.A. theses in the following years (Mueller, 2010; Jolley, 2014). The original CPRs were developed as a self-directed, teacher- and course-independent method of delivering pronunciation instruction with the purpose of suprasegmental instruction and practice through oral reading. Tanner and Landon’s (2009) study focused on the prosodic features of pausing, word stress, and sentence-final intonation patterns. Seventy-five intermediate ESL learners were divided into a treatment and a control group. Learners in the treatment group received computer-assisted pronunciation training, which they completed voluntarily as extra credit in an on-campus computer lab over the course of 11 weeks. During this training, learners were presented with a short reading passage in English and were instructed to mark suprasegmental features in each passage while listening to the passage being read by native speakers. They could listen to the passage as often as they needed to in order to mark it appropriately. Learners were then given an answer key with visual solutions to the features of pausing, word stress, and sentence-final pitch contours and were prompted to practice reading the passage with a focus on correctly placing these suprasegmental features. Finally, they were prompted to record themselves and submit their recordings to their

---

5 The M.A. theses by Mueller (2010) and Jolley (2014) will not be discussed here. While they did technically use the method of Cued Pronunciation Readings, the training was implemented in an entirely different context and under conditions that do not apply to the present study. That is, both authors adapted CPRs for the use in Missionary Training Centers in Utah—this adaptation made it necessary to modify all materials to work within the curriculum and computer systems of the Provo Utah Missionary Training Center under very different conditions than were investigated in Tanner and Landon (2009) or in the present study.
instructor—they did not, however, receive feedback on their productions. The authors reported significant improvements for perception of pause and stress assignment and significantly fewer instances of incorrectly placed stress in the production task. However, they did not find significant improvement on the level of comprehensibility.

As I will show in the following, the pronunciation training in the present study differs in many aspects from Tanner and Landon’s (2009) original CPRs and will therefore be referred to as innovative Cued Pronunciation Readings (iCPRs). In the discussion of the limitations of their study, Tanner and Landon note that it appears as if two factors had a negative influence on the effectiveness of CPRs: the fact that the training was voluntary and assigned as extra-credit and the fact that CPRs could only be completed in computer labs on campus. The authors report that less than half of the learners completed even 80% of the work (i.e., 9 of the 11 possible CPR units). In a brief follow-up survey administered in Tanner and Landon’s study, learners report that they did not complete all the CPR units because it was too inconvenient to go to the computer lab or because the extra credit was not enough of an incentive to go there. The iCPR units used in the present study are therefore designed to be easily accessible for the learners. They were created in Microsoft PowerPoint and can be accessed and used with Microsoft PowerPoint, Apple’s Pages, or free alternatives such as Google Slides or LibreOffice. This makes them easily accessible to any learner with access to a home or school computer. Moreover, iCPR units were implemented into the standard curriculum and assigned as homework rather than making them an extra-credit option. This still leaves the limitation that not all students complete every single homework assignment, but at least learners see the pronunciation as an integral part of their classwork and their homework completion grade, which usually leads to a much higher completion rate than extra-credit work. Another way in which the
method of iCPRs differs from the original CPRs lies in the learners’ proficiency level and L1 background. Tanner and Landon’s study investigated the efficacy of CPRs in a group of ESL intermediate-level language learners who had learned English for a time period of two months to 17 years (median = 4 years). In a study comparing gains in pronunciation proficiency for first-semester learners and fourth semester learners, however, Roccamo (2014) showed that novice learners seem to benefit more from pronunciation training. This finding is in line with Thomson and Derwing (2015) and Kissling (2013), who also call for an early implementation of pronunciation training in the classroom (see also Flege, 2007 who showed that in immersion settings, learners’ phonological system is most easily influenced during the first year). Therefore, in the present study, iCPRs were implemented in first-semester German courses, that is, classes for absolute beginners. Moreover, it has to be pointed out that German classes in the US differ from ESL classes in that the group of learners is usually more homogenous in terms of L1: while there are sometimes international students with L1s other than English, the majority of learners has English as an L1. For an ESL context, however, Tanner and Landon reported that their participants were recruited from learners of nine different L1s. The more homogenous L1 background in the present study facilitated addressing typical learner errors, which are often similar between learners of the same L1 (for additional discussion see Section 2.4.1.3). Finally, Tanner and Landon decided to only focus on suprasegmental features in their study for practical reasons. They do not claim, however, that the method cannot work for features on the segmental level as well. Given the findings discussed in Section 2.2.3, which showed that both segmentals and suprasegmentals should be addressed in pronunciation instruction, the iCPRs in the present study thus address both types of features.
Previous research on iCPR pronunciation training has been promising. That is, results from a pilot study with 22 first-semester German learners revealed that even though there were no statistically significant differences between the iCPR and the control groups, the participants who received homework-based pronunciation instruction descriptively outperformed their peers in the control groups (Martin, 2015; for a practical teacher guide on how to create and use iCPR units, see also Martin, 2017).

Ultimately, I want to investigate the extent to which the method of iCPRs can offer a solution for the discrepancy between the importance of pronunciation in L2 communication and its neglect in the classroom. Providing pronunciation training through homework assignments solves the problems discussed in Section 2.1.2: it does not take up valuable in-class time and even those teachers who do not feel adequately prepared to teach pronunciation can assign the exercises as homework. While Derwing (2013) calls for pronunciation instruction courses in L2 teacher education to address the issue that many L2 teacher do not feel adequately prepared to teach pronunciation in the classroom, it is important to find a more timely solution for current teachers who are not able to go back to university. The use of iCPRs offers such a solution. This method does not cause additional work for the teacher as iCPR units do not require feedback and the teacher does not need particular skills in teaching pronunciation when using previously created iCPR units.

2.4.1.2 Design of Materials for the Present Study

There are two different types of pronunciation training employed in the present study: homework-based, computer-delivered iCPR units, and teacher-directed, in-class pronunciation training. In this section, I will show how the literature reviewed in this chapter has influenced the
design of materials for both types of pronunciation training. I will start by discussing those findings that have informed the design of both types of pronunciation materials.

The literature reviewed in Section 2.2.2. has emphasized the relationship between perception and production in L2 pronunciation, showing that improvements in perception abilities often lead to improvements in production abilities in the L2. To account for this finding, pronunciation training in the present study always begins with a perception training component: both the iCPR and the in-class pronunciation training are designed to spend three training units on one pronunciation target (e.g., a vowel, a consonant, or stress assignment), and the first of these three training units always targets perception skills. Moreover, this perception training is designed to account for findings in HVPT: each sound is presented in various phonetic contexts and was recorded by several native speakers of German in order to give learners an ideal opportunity to build new sound categories for the targeted features. The learners also receive immediate feedback on which was the correct form during the perception training, just as is the case in lab-based HVPT. In order to ensure that both the homework-based and the in-class groups can benefit from HVPT, the recordings collected for the material design for the iCPRs are being used in the classroom as well and are played to the entire class by the course instructor. This offers a unique way to bring the benefits of a multiple-voice-based perception training into a classroom in which learners traditionally only hear the voice of their teacher, or at best, some additional audio-materials.

Another factor that is relevant for both types of pronunciation training included in the present study is the component of explicit instruction. Previous research has shown that explicit instruction on phonological form helps to improve learners’ pronunciation (Couper, 2006; Garcia, 2005; Gordon, Darcy, & Ewert, 2013; Lord, 2005; Saito, 2011b; for a review see
Thomson & Derwing, 2015, but see also Kissling, 2013). A possible explanation for this is that orienting learners’ attention to the phonetic input and its features provides students with declarative knowledge that can accelerate the creation of new linguistic categories and thus leads to greater gains in pronunciation proficiency (DeKeyser, 2003). Therefore, explicit instruction was included in the pronunciation materials for the present study and is part of both the iCPR units and the in-class pronunciation training.

Finally, in Section 2.2.1, I discussed the goals of pronunciation training and showed that the main goal of pronunciation instruction in the classroom has shifted from native like attainment towards improving learners’ comprehensibility and intelligibility in the L2. However, while it is true that the majority of teachers now focusses on intelligibility and comprehensibility, the review of the literature also showed that accentedness still plays a crucial role due to the social stigma it carries and that learners themselves express the desire to improve their accentedness in L2 pronunciation training. The materials in the present study therefore aim to address all three concepts—intelligibility, comprehensibility, and accentedness—however, only comprehensibility and accentedness are included in the testing. This concession was methodologically necessary because testing for intelligibility as a measure would have entailed to ask raters to transcribe the words and sentences they heard. In order to keep these transcriptions fair and reliable, raters could not have heard the same word or sentence more than once, which means that every learner would have had to produce different tokens. In a large-scale classroom study like the present study, this was not feasible—particularly not with novice learners who have very limited L2 skills to begin with (see Gordon & Darcy, 2016 for the same methodological approach).
In addition to the above-mentioned findings that influenced both the computer-based and the teacher-directed pronunciation training, there are some findings that have primarily influenced the design of the computer-based iCPR units. First and foremost, iCPRs use the method of learners modelling their productions after native-speaker input. While this is a common practice (e.g., Kissling, 2013; Weinberg & Knoerr, 2003), it differs somewhat from the practice of shadowing (see Foote & McDonough, 2017 for a review) in that the technique of shadowing is usually used for longer sentences or entire paragraphs, which the learners imitate as closely as possible. Findings by Guion and Pederson (2007), however, suggest that shadowing longer samples of speech might not be the most effective, especially at a beginner’s level. In their study, Guion and Pederson manipulated attention during training of unfamiliar phonetic categories from Hindi. English monolingual participants were randomly divided into two groups—one meaning-attending and one sound-attending group. They administered a pretest consisting of a sound discrimination task and a semantic task that was identical for both groups. The training instructions for both groups then differed in that the meaning-attending group was told that they would learn the English meaning for the Hindi words that were taught, whereas the sound-attending group was told that they would learn the differences between Hindi sounds. Both groups received identical training, only the instructions concerning the learning focus differed. Results from a posttest then showed that the sound-attending group were better able to discriminate between Hindi dental and retroflex consonants than the meaning-attending group. This not only shows that orienting L2 learners’ attention to phonetic information before a perceptual training task might result in significantly greater gains for learners, but also that learners can benefit the most from phonetic input and training when stimuli are presented in contexts where competing demands for attention are kept to a minimum (see also Thomson,
2011). Presenting learners with individual words as in iCPR units thus maximizes the learners’ attention on the targeted sound, and allows learners to focus their attention on producing that sound and word rather than using too much attention to focus on communication. Furthermore, Thomson and Isaacs (2009) found that learners’ pronunciation of individual tokens was the best when they produced them after simultaneously hearing the word modelled by a native speaker and seeing its written form at the same time. Thus, iCPR units were designed in a way so that learners can always see the word while they listen to the native-speaker recording and while they practice their own production of the word.

Finally, I want to address a few issues that relate to the differences between the computer-delivered and in-class pronunciation training. As illustrated by the many studies discussed in this chapter, teacher-directed, in-class pronunciation training is the most common form of pronunciation training in L2 classrooms, and its effectiveness has been repeatedly confirmed (Counselman, 2010; Couper, 2003, 2006; Derwing et al., 1998; Elliott, 1995, 1997; Lord, 2005; Neri et al., 2008; Roccamo, 2014; see J. Lee et al., 2015 for results of a meta-analysis). There are two ways in which homework-based iCPR units and teacher-directed instruction differ. The first is the absence of corrective feedback in iCPRs. Given findings from studies such as Saito and Lyster (2012a), which showed that corrective feedback plays an important role in improving learners’ pronunciation skills, it might be a substantial advantage of teacher-directed training that teachers can provide corrective feedback while practicing pronunciation in the classroom, whereas corrective feedback is not a component in the design of iCPR units. However, more recent research by Saito (2015) has shown that corrective feedback (here in form-focused instruction) is not always necessary for improving pronunciation skills. This is further confirmed in J. Lee et al.’s (2015) meta-analysis in which the authors found only a
small difference in effect sizes based on whether feedback was provided or not (for instance, $d_{\text{feedback}} = 0.92$ vs. $d_{\text{no feedback}} = 0.89$ for the within-group contrast). In fact, Saito (2015) argues that particularly at the lower level of language proficiency, corrective feedback does not seem necessary in order for learners to improve their pronunciation, whereas it becomes more important at higher proficiency levels. Since the present study only targets novice learners of L2 German, no learning disadvantage should be predicted for learners who receive pronunciation training through iCPR units instead of in-class pronunciation instruction, which allows for corrective feedback through the instructor. Finally, the second apparent way in which homework-based iCPR units and teacher-directed instruction differ is that iCPR units were designed taking findings in HVPT into consideration (i.e., introducing sounds in various phonetic contexts and spoken by several native speakers), whereas input in in-class pronunciation instruction is usually limited to the voice of the instructor and, at best, offers some additional voices through audio or video materials. This shortcoming is addressed in the present study, however, by using the recordings collected for the iCPR perception units in the classroom instruction as well. That is, instead of learners clicking through the slides and listening to the recordings as a homework assignment, the instructor displays the iCPR unit in front of the class and all students receive two colored cards, allowing them to actively participate by raising the correctly colored card depending on which sound constitutes the correct answer (for details see Section 3.1.2.1.2). Overall, serious attempts were made to even out the advantages and disadvantages of each type of pronunciation instruction.
2.4.1.3 Pronunciation Targets Chosen for Training

Ten pronunciation targets were chosen for training in the present study:

1. allophones of /r/: consonantal [ʁ] and vocalic [ɐ]
2. ich- and ach-sounds [ç] and [x]
3. monophthong [e:]
4. monophthong [o:]
5. front rounded vowel [y:]
6. front rounded vowel [ø:]
7. orthographic-phoneme-correspondence <z> → /ts/
8. orthographic-phoneme-correspondence <ie> → /i:/ and <ei> → /ai/
9. orthographic-phoneme-correspondence <v> → /f/
10. lexical stress in German-English cognates

In Section 2.2.3, I showed that previous research suggests that learners benefit from pronunciation training the most when instruction targets segmental and suprasegmental features. Therefore, both types of features were included as pronunciation targets in the present study. The targets presented above were chosen because they all present difficulties in pronunciation for American learners of German and, when mispronounced, impede learners’ intelligibility and readily give the L2 speaker away as being a native speaker of American English (Hall, 2003). Unfortunately, to the best of my knowledge, there is no study investigating specifically which sounds are most difficult for L1 English learners of L2 German and which features influence intelligibility, comprehensibility and accentedness the most. Therefore, I chose these targets based on experience in teaching German in American classrooms, and previous research findings (Dieling & Hirschfeld, 2000; Hall, 2003; Moyer, 1999; O’Brien, 2003; Roccamo, 2015). Most of
these sounds might be difficult to acquire correctly due to their similarity with sounds in the English phonetic inventory, which often leads to sound substitutions by L1 English learners of German. That is, English speakers commonly substitute English /u:/ for German /y:/—possibly due to the fact that English /u:/ is more fronted than German /u:/.

Moreover, English speakers often substitute English [ɜ] (the vowel in bird) for realizations of German /ø:/, such that schön ‘pretty’ rhymes with burn. Furthermore, [ɐ] (i.e., the vocalic allophone of German /r/) is perceptually close to [ə], which causes English L1 speakers trouble perceiving and producing the difference between the two vowels, particularly in unstressed positions in words like bitte ‘please’ versus bitter ‘bitter’. Similarly, German [ç] is similar to English [ʃ], thus it can be difficult to perceive or produce these sounds in contrasting words like Kirche ‘church’ and Kirsche ‘cherry’. The German monophthongs /o:/ and /e:/ seem particularly problematic because they are so similar to the realization of English tense mid-vowels, but when prompted to produce a long vowel here, learners often substitute the English diphthongs [ou] as in boat and [ei] as in main for the German monophthongs. Finally, while the velar fricative [x] seems to be perceptually distinct from realizations of American English /k/, and the German uvular fricative [ʁ] seems to be perceptually distinct from realizations of English /ʃ/, it is clear that L1 English speakers routinely substitute [k] for [x] and [ʁ] for [x]—the latter being a particularly strong shibboleth for an American accent in German. According to Flege’s (1995) SLM (see Section 2.2.2), it should be possible to correctly acquire these sounds, but it might be necessary to train learners to perceive the differences between the commonly substituted and the targeted sounds first. Furthermore, three of the features targeted in this study fall into the category of “orthographic-phoneme-correspondence”. Classroom experience frequently shows that students struggle with mapping sounds to orthographic representations. The targeted phonemes /ts/, /i:/,
/ai/, and /f/ all have counterparts in the English phonological system and, according to the SLM, should pose no problems to acquire. Yet, the challenge for American learners of German does not lie in learning to perceive and produce these sounds but rather to acquire the German orthographic-phonemic mapping where it differs from the mapping in English. Therefore, these features will be addressed in the pronunciation training of the present study as well. Finally, it has been shown that correct placement of lexical stress is important for intelligibility (Hirschfeld, 1994). While both German and English mark lexical stress in similar ways, it often poses a difficulty for learners to place the German stress assignment on the correct syllable (Roccamo, 2015). Stress assignment in German is largely systematic and predictable, but varies depending on the word’s structure and syllable count (O’Brien & Fagan, 2016). Difficulties then often arise for English learners of German when cognate pairs follow opposite stress patterns—for instance, ‘stu.dent vs. Stu.’dent (Maczuga, O’Brien, & Knaus, 2017). Therefore, the suprasegmental feature of stress in German-English cognates was included as a pronunciation target in the present study.

2.4.2 Research Questions and Hypotheses

2.4.2.1 Experiment 1 – Face-to-Face Learning Environment

The aim of Experiment 1 was to explore the effectiveness of homework-based, computer-delivered iCPR pronunciation training in a F2F learning environment. The previous literature reviewed in this chapter has shown that learners’ L2 pronunciation generally improves through pronunciation instruction. Yet, pronunciation instruction is often neglected in the classroom because instructors do not feel adequately prepared to teach pronunciation or do not have time to include pronunciation training in a tight curriculum. Thus, the present study addresses this chasm
by exploring a method of pronunciation instruction that can be assigned as homework—allowing instructors to save valuable in-class time and serving as a tool that instructors can easily fall back on if they do not feel confident in teaching pronunciation themselves. In order to examine the effectiveness of the method of iCPR pronunciation training in F2F learning environments, the present study explores two levels of comparison: (1) learners who received iCPR pronunciation training (i.e., the treatment group) were compared to learners who did not receive targeted pronunciation training, but otherwise followed the same GER 001 curriculum (i.e., the control group), and (2) learners who received iCPR pronunciation training were compared to learners who received teacher-directed in-class pronunciation training (i.e., the comparison group). The pronunciation training in the treatment and comparison groups focused on the same pronunciation targets and the time spent on task was kept constant among both groups. The control group received additional grammar and vocabulary exercises instead of the pronunciation training to ensure that the time learning German was kept constant among all three groups. The training was administered over the course of an entire semester and gains in learners’ L2 pronunciation skills were measured through a pretest in the second week of the semester and a posttest in the 14th week of the semester. Each test consisted of a perception and a production task. The perception task was a binary-forced-choice task, and learners’ pre- and posttest perception skills were measured based on the total score learners reached in this task. The production task consisted of word- and paragraph-readings. Learners’ pre- and posttest productions were rated for comprehensibility and accentedness by native speakers of German. The perception scores and the native speaker ratings of learners’ productions constituted the data for statistical analyses that were performed to determine whether the differences between the three groups and between each group at the two testing times were significant. Finally, a
questionnaire was administered in the treatment and comparison groups at the end of the semester in order to explore learners’ experience with both types of pronunciation training. As such, the following research questions were addressed in Experiment 1:

RQ1: Is pronunciation training delivered through iCPR units in a face-to-face learning environment effective in significantly improving novice learners’ L2 perception skills when compared to a control group?

RQ2: Is pronunciation training delivered through iCPR units in a face-to-face learning environment effective in significantly improving novice learners’ L2 production skills when compared to a control group?

RQ3: Is pronunciation training delivered through iCPR units in a face-to-face learning environment as effective as in-class pronunciation training in improving novice learners’ L2 perception skills?

RQ4: Is pronunciation training delivered through iCPR units in a face-to-face learning environment as effective as in-class pronunciation training in improving novice learners’ L2 production skills?

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

Previous research suggests that learners who receive pronunciation training in a F2F learning environment are more likely to improve their pronunciation skills than learners who do not receive pronunciation training (see Thomson & Derwing, 2015 for a review). Moreover, the design of the present pronunciation training was informed by most recent findings in pronunciation research. Therefore, it is expected that the learners receiving pronunciation training through iCPR units will make greater gains in perception and production skills over the course of the semester than learners who do not receive pronunciation training.
Furthermore, given that L2 learners have demonstrated improvement in pronunciation skills as a result of training in classroom-based pronunciation studies (see Thomson & Derwing, 2015 for a review) and that Tanner and Landon (2009) have found ESL learners to improve their pronunciation through the use of Cued Pronunciation Readings, it is anticipated that both groups of learners—that is, learners receiving pronunciation training through iCPR units and learners receiving in-class pronunciation training—will make comparable gains in perception and production skills over the course of the semester.

2.4.2.2 Experiment 2 – Distance Learning Environment

The aim of Experiment 2 was to explore the effectiveness of computer-delivered iCPR pronunciation training in a distance learning environment. Previous research has shown that novice learners’ oral proficiency in German and Spanish online learning environments developed on par with F2F learners (R. Blake et al., 2008; Isenberg, 2010). While the development of learners’ oral proficiency in an online learning environment has been shown to be problematic (Deutschmann et al., 2009), to my best knowledge, no previous study has explored the effectiveness of particular methods of L2 instruction to teach pronunciation in online environments. The present study therefore aims to close this gap in the literature by exploring homework-based, computer-delivered iCPR pronunciation training. This method of pronunciation training lends itself ideally to pronunciation training in distance learning environments since it can be added on to any existing online curriculum and does not require synchronous communication or F2F interaction.

In order to examine the effectiveness of the method of iCPR pronunciation training in distance learning environments, the present study compares gains in pronunciation skills from
learners who received iCPR pronunciation training (i.e., the treatment group) and learners who did not receive targeted pronunciation training, but otherwise followed the same GER 001 curriculum (i.e., the control group). Both groups were instructed in an online learning environment. Contrary to the set-up in Experiment 1, there was no comparison group in Experiment 2 since the comparison group in Experiment 1 received in-class pronunciation training but there are no in-class meetings in the online learning environment. All other methodological choices and analyses were the same as in Experiment 1 and will therefore not be repeated here. As such, the following research questions were addressed in Experiment 2:

*RQ5:* Is pronunciation training delivered through iCPR units in a distance learning environment effective in significantly improving novice learners’ L2 *perception* skills when compared to a control group?

*RQ6:* Is pronunciation training delivered through iCPR units in a distance learning environment effective in significantly improving novice learners’ L2 *production* skills when compared to a control group?

The prediction and hypothesis for this research question is as follows:

Previous research has shown that novice learners’ oral proficiency in an online environment develops on par with their F2F counterparts (R. Blake et al., 2008; Isenberg, 2010). In combination with findings that have shown that learners who receive pronunciation training in a F2F learning environment are more likely to improve their pronunciation skills than learners who do not receive pronunciation training, these findings suggest that learners who receive iCPR pronunciation training in an online learning environment are more likely to improve their pronunciation skills than learners who do not receive pronunciation training. Therefore, it is expected that the learners receiving pronunciation training through iCPR units in an online
environment will make greater gains in perception and production skills over the course of the semester than learners who do not receive pronunciation training.
CHAPTER 3: Experiment 1 – Face-to-Face Environment

3.1 Methodology

3.1.1 Participants

A total of 136 learners from six sections of first-semester German classes at Penn State University were eligible to participate in the face-to-face environment part of this study. One-hundred and twenty-two of these learners gave consent to participate. Of these 122 learners, 66 had to be excluded for one of the following reasons: a) they were not L1 speakers of English \( n = 34 \), b) they missed more than 20% of either the iCPR homeworks or in-class meetings \( n = 15 \), c) they dropped the class before the final round of data collection \( n = 7 \), d) they had spent more than four weeks in Germany or Austria and reported speaking German during their time abroad \( n = 5 \), or e) the quality of their spoken recordings was so poor that their data could not be rated by the native speaker raters \( n = 5 \). After these exclusions, there were 56 participants who met all of the criteria for inclusion in the study. However, only a subset of 39 of these participants had completed both pre- and posttest for perception and production. To avoid additional participant loss, learners were therefore included in the analyses if they had completed at least one set of the pre- and posttest: either the perception or the production tests. That is, learners were included in the perception analyses if they had completed the perception pre- and posttest, even if the production data collection for the learner was incomplete, and learners were included in the production analyses as long as they had completed the production pre- and posttest. This led to a total of 46 participants in the perception analyses and 49 participants in the production analyses.

---

6 An additional two sections of GER 001 classes were originally recruited. The study was introduced to the learners and the learners took the pretest and received some pronunciation training. After the instructor fell ill and the classes had multiple substitute teachers, I could no longer be sure that the learners in these two sections received the same training as learners in the other sections. Therefore, I excluded the sections from the analyses.
analyses. There were seven learners that were included in the perception analyses but not in the production analyses and ten learners that were included in the production analyses but not in the perception analyses.

Two sections each were assigned to the treatment group ($N = 19$, $n_{\text{perception}} = 16$, $n_{\text{production}} = 19$), comparison group ($N = 16$, $n_{\text{perception}} = 14$, $n_{\text{production}} = 14$), and control group ($N = 21$, $n_{\text{perception}} = 16$, $n_{\text{production}} = 16$). Table 1 shows learners’ average age and self-rated proficiency for German reading, writing, speaking, and listening on a scale from 1 (very poor) to 10 (very proficient) by group. A one-way ANOVA revealed that the learners from the three groups did not differ significantly in age ($F(2,53) = 2.26$, $p = .114$) or their self-rated proficiency in German reading ($F(2,53) = 0.27$, $p = .768$), writing ($F(2,53) = 0.59$, $p = .560$), speaking ($F(2,53) = 0.99$, $p = .378$), and listening ($F(2,53) = 0.36$, $p = .702$).

**Table 1: Experiment 1 – Learners’ Age and German Proficiency Ratings**

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group ($n = 19$)</th>
<th>Comparison Group ($n = 16$)</th>
<th>Control Group ($n = 21$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f = 10$; $m = 9$</td>
<td>$f = 5$; $m = 11$</td>
<td>$f = 6$; $m = 15$</td>
</tr>
<tr>
<td>Age (years)</td>
<td>19.3 (1.4)</td>
<td>21.4 (5.3)</td>
<td>19.5 (1.9)</td>
</tr>
<tr>
<td>Self-ratings (max. 10)</td>
<td>Reading 2.8 (1.9)</td>
<td>2.4 (1.3)</td>
<td>2.8 (1.4)</td>
</tr>
<tr>
<td></td>
<td>Writing 2.4 (1.5)</td>
<td>2.1 (1.1)</td>
<td>2.5 (1.3)</td>
</tr>
<tr>
<td></td>
<td>Speaking 2.3 (1.0)</td>
<td>2.4 (1.3)</td>
<td>2.9 (2.0)</td>
</tr>
<tr>
<td></td>
<td>Listening 2.3 (1.1)</td>
<td>2.4 (1.4)</td>
<td>2.7 (1.5)</td>
</tr>
</tbody>
</table>

Furthermore, three German instructors participated in this part of the study. A native speaker of German with one year of teaching experience taught both sections of the control group. Another native speaker of German with two years of teaching experience taught one section of the treatment group and one section of the comparison group. A native speaker of Swedish with
native-like German proficiency, who had lived nine years in Germany and had more than 20 years of German teaching experience, taught the other section of the treatment and the comparison group. The author of this dissertation did not act as an instructor in this study.

Finally, eight native speakers of German (3 female, 5 male) were recruited as raters for Experiment 1 by means of a mass email sent through the international office at Penn State. In line with previous L2 pronunciation research (Oh, Jun, Knightly, & Au, 2003; Saito, Webb, et al., 2016; Yeni-Komshian, Flege, & Liu, 2000), these native speakers ($M_{age} = 31$) had lived in the US for no more than five years ($M_{length\_of\_stay} = 32$ months), had never taught German, and had no hearing impairment. All but one rater, who was a visitor to Penn State, were Master’s or PhD students at Penn State majoring in Business Administration, Spanish, Physics, Mathematics, Econometrics, Political Science, and Computer Science. All raters grew up in Germany or Austria with German as their first language. They self-reported their German reading, speaking, writing, and listening proficiency at either 9 or 10 out of 10. Additional information on inter-rater reliability will be provided in Section 3.1.5.2.1.

---

7 One rater grew up hearing Persian as an additional first language. The intraclass correlation analyses presented in Section 3.1.5.2.1, however, show that this rater exhibited excellent inter-rater reliability with the other raters. Additional inter-item correlation analyses showed that inter-rater reliability for the group would not improve if the rater was removed, suggesting that his native German skills were not influenced by the fact that he also heard Persian from birth.
3.1.2 Materials

3.1.2.1 Instructional Materials

3.1.2.1.1 iCPR Units

General Structure of iCPR units

The homework-based pronunciation training materials were delivered in the form of Microsoft PowerPoint presentations and consisted of 30 iCPR units, split into ten sets of three units – each set covering one of ten conditions (see Section 2.4.1.3 for a discussion of the pronunciation conditions chosen for investigation in this study). Each set of three units was further split such that the first unit always addressed perception training while the last two units addressed production training. This pattern was only broken in the last week focusing on segmental instruction (Week 9 of 10), in which the first unit addressed both perception and production of the orthographic-phoneme-correspondence <v>-[f] in German and the last two units served as a review of the previous eight conditions treated in the first eight weeks of the study (for the special role of the <v> condition and an explanation of why it was addressed in only one unit, refer to Section 3.1.2.2.2).

All words included in the training materials were taken from the first four chapters of the learners’ textbook, *Sag mal* (Anton, Barske, Grabowski, McKinstry, 2014). In a few instances, however, it was necessary to take items from later chapters of the textbook or even from external sources. This was only the case if the first four chapters of *Sag mal* did not provide enough words to satisfy the training conditions, that is, for the [ø]-condition as well as for cognates with differing stress patterns in German and English.

The ten sets of iCPR units were presented in the following order:

1. allophones of /r/: consonantal [ʁ] and vocalic [œ]
2. orthographic-phoneme-correspondence <z> \rightarrow /ts/

3. ich- and ach-sounds [ç] and [x]

4. monophthong [e:]

5. front rounded vowel [y:]

6. orthographic-phoneme-correspondence <ie> \rightarrow /i:/ and <ei> \rightarrow /au/

7. front rounded vowel [ø:]

8. monophthong [o:]

9. orthographic-phoneme-correspondence <v> \rightarrow /f/ and review of previous conditions

10. lexical stress in German-English cognates

This order was determined by the number of available words per condition in the first chapters of
Sag mal, as well as with the goal to mix up the two conditions of orthographic-phoneme-
correspondences and acquisition of new sounds. That is, the basic vocabulary in the first four
chapters of the learners’ textbook included the most words containing an <r>, thus, starting the
pronunciation training with the [ʁ]-[ɐ]-condition allowed for the largest number of known
vocabulary items to be included. On the other hand, there were fewer cognates with differing
stress in the first chapters of the book, therefore, this condition came last so that all cognates in
the first four chapters could be included and were expected to be known by the learners at the
time of this training set. Overall, an effort was made to match the words included in the training
in each particular week to the vocabulary list of the book chapter that the learners were working
on at the time of the training: for example, if the learners worked on Chapter 3 in the sixth week
of the semester and that week’s pronunciation focus was on the ich- and ach-sound, all words
from Chapter 3 were prioritized in the design of the pronunciation materials. This was done so
that the learners would have a secondary benefit from the pronunciation exercises in reviewing
current vocabulary items. If the current chapter did not provide sufficient words, words from previous chapters were taken before words from future chapters. Additionally, every single German word included in iCPR units appeared with its English translation, so that, even if a learner forgot the meaning of a previously learned vocabulary item, they would always know its translation while working on the pronunciation of the word.

Each training item in the perception and production units was recorded by a native speaker of German (or English if applicable, see Perception Units section below). A total of nine speakers participated in the recordings of words. Table 2 provides an overview of all speakers and their biographical information. Only a subset of these speakers, however, participated in each subset of recording conditions (i.e., in the perception units, production units, and testing items). In the following sections, I will list the participants for each of these tasks, but I will not repeatedly mention their biographical information as provided in the table below.

Table 2: Experiment 1 – Biographical Information of Speakers for Recordings

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age</th>
<th>Time participant studied, taught, or used German (years; months)</th>
<th>Time participant lived in a German speaking country (years; months)</th>
<th>Time participant spent living in the US (years; months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>f</td>
<td>30</td>
<td></td>
<td></td>
<td>2;4</td>
</tr>
<tr>
<td>G2</td>
<td>f</td>
<td>36</td>
<td></td>
<td></td>
<td>3;5</td>
</tr>
<tr>
<td>G3</td>
<td>f</td>
<td>55</td>
<td></td>
<td></td>
<td>1;0</td>
</tr>
<tr>
<td>G4</td>
<td>f</td>
<td>30</td>
<td></td>
<td></td>
<td>4;3</td>
</tr>
<tr>
<td>G5</td>
<td>m</td>
<td>25</td>
<td></td>
<td></td>
<td>0;5</td>
</tr>
<tr>
<td>G6</td>
<td>m</td>
<td>57</td>
<td></td>
<td></td>
<td>1;0</td>
</tr>
<tr>
<td>E1</td>
<td>f</td>
<td>33</td>
<td>20;0</td>
<td>6;0</td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>m</td>
<td>30</td>
<td>8;10</td>
<td>3;7</td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>m</td>
<td>37</td>
<td>12;0</td>
<td>3;2</td>
<td></td>
</tr>
</tbody>
</table>

Note: “G” marks a native speaker of German, “E” marks a native speaker of US English. Fields that do not apply are shaded grey.
In the following, the design of the two types of iCPR units—perception and production units—will be discussed further.

**Perception Units**

The iCPR perception units consisted of two types of exercises: an accentedness detection task and a sound discrimination task. The accentedness detection task was modelled after classroom-based perception treatments as used in Botero (2011) and Roccamo (2014), and had the goal of training learners’ ability to discern native from accented productions of problematic L1 speech sounds in real words. This type of training is aimed at raising learners’ awareness of differences between a native and nonnative accent and at encouraging learners to self-monitor their own productions for these accent markers. Each perception unit contained ten accentedness detection practice items. These items were presented to the learners in Microsoft PowerPoint. On each slide, the learners saw the written version of a German word and were given two recordings for the same word. One recording was always spoken with a Standard German pronunciation, while the other recording was intentionally rendered with an American English accent of the German word. On the next slide, the learners received feedback on which recording of the word was spoken with Standard German pronunciation (see Figure 1 for an illustration of this type of exercise). The order (i.e., whether the accented or Standard German production of a word appeared on the left or right side of the slide) was semi-randomized. Five native German speakers (G1, G2, G3, G5, G6) and three native English speakers (E1, E2, E3) recorded the practice items for this task. The Standard German and the accented versions of the same word were produced by two different speakers. To ensure that the learners did not simply rely on individual voices when judging the recordings as accented or not, the native Germans were asked
to record 40% of the accented stimuli and the native English speakers were asked to record 40% of the Standard German stimuli. The native Germans were all proficient speakers of English who had started to learn English in elementary school and had lived in the US for an average of 1 year and 8 months (range: 5 months to 3 years and 5 months). The Americans were all proficient speakers of German who had spent an average of 14 years studying, teaching, or speaking German (range: 8 years and 10 months to 20 years) and had lived in a German speaking country for an average of 4 years and 3 months (range: 3 years and 2 months to 6 years). When producing the accented stimuli, speakers were instructed to only alter the targeted sound and to produce the rest of the word in Standard German. Speakers repeated all recordings until they produced a token that was deemed acceptable to the author of this dissertation and to themselves.

Figure 1. Sample Slides for Accentedness Detection Task

Listening Practice

- This exercise is designed to help you hear the differences between English and German pronunciation.
- In this exercise, you will listen to pairs of German words.
  - One word is pronounced with a Standard German pronunciation, and the other one is pronounced with an American English accent.

⇒ Decide whether the first word or the second word sounds more like a Standard German native speaker.

What to do:

- Read the German word at the top of the slide.
- Click each sound symbol once to play the sound.
- Decide which word sounds more like Standard German.
- Listen to each word as many times as you need to to make a decision.

⇒ The answer will always appear on the following slide.

Solution

⇒ #2 sounds more like a Standard German pronunciation.
- Listen to it again:
The sound discrimination task was modelled after lab-based perception treatments (Guion & Pederson, 2007; Iverson & Evans, 2009; Thomson, 2011) and focused on assessing the perception of problematic German speech sounds in contrast to similar sounds in English, which could easily be perceived as the targeted sound by an English speaker who is unfamiliar with the phonetic inventory of German. For this exercise, the L2 speech sounds were embedded in nonsense words—that is, words that follow phonotactic rules of German and could bear meaning but that do not. Just as in the accentedness detection task, learners were presented with ten nonsense words one by one on a Microsoft PowerPoint slide. This time, however, not the entire nonsense word was spelled out but only the orthographic correspondence of the targeted L2 speech sound (e.g. the letter <ö> for the German sound [øː]). Two recordings were still provided on each slide, and the learners had to listen to them and decide which of the two recordings contained the targeted German speech sound. On the following slide, the learners received automated feedback on which word was spoken with Standard German pronunciation (see Figure 2 for an illustration of this type of exercise). Three native German speakers (G1, G2, G6) and two native English speakers (E1 and E3) recorded the practice items for this task. As opposed to the accentedness detection task in which the accented and the Standard German version of the same word were produced by two different speakers, in the sound discrimination task, the same speaker produced both words that formed a minimal pair. This was decided upon after piloting two units with 40 learners in two GER 001 classes in the semester prior to the beginning of the study. The pilot study had shown that it was more difficult for the learners to identify the Standard German sound in nonsense words if the two minimal pairs were produced by the same speaker. Given the instructional goal of preparing learners for real-life conditions in which they also have to discern individual sounds produced by the same speaker in combination with the
research goal of avoiding ceiling effects at the time of the pretest, it was deemed appropriate to train the learners on the condition that was more difficult, thus this design was used in the present study. Each of the five speakers then produced an equal number of tokens in each perception unit. The targeted sound in each nonsense word was contrasted with two to four phonetically similar sounds, for example, [œ:] was contrasted with [y:], [u:], [e:] and [o:] (for a list of all nonsense words stimuli and their minimal pairs, see Appendix A). For the consonantal target sounds [ʁ], <z>, [ç] and [x], the minimal pair nonsense words followed a VCV (vowel-consonant-vowel) pattern. Front and back vowels were altered in all positions. For the vocalic target sounds [y:], <ie>, <ei>, [œ:], and [o:], the minimal pair nonsense words followed a CVC (consonant-vowel-consonant) pattern, only allowing the obstruents [b], [p], [t], [d], [k], [g], [s] and [z] in the consonantal positions. There was no sound-discrimination training for the <v> and cognate stress condition as both of these conditions make no sense in nonsense words.

*Figure 2. Sample Slides for Sound Discrimination Task*
The perception units for [o:] and [e:] slightly differed from the other perception units, only in that they started with explicit information on the monophthongal nature of German [o:] and [e:] before the accentedness detection and sound discrimination tasks followed. This was deemed necessary because English native speakers often think that they do not need training on those two vowels as these vowels are so similar to the realization of English tense mid-vowels, and American learners of German often do not even perceive the difference between the long German monophthongs [o:] and [e:] and the English diphthongs [oo] and [ei]. Since it is a strong accent marker to produce the English diphthongal realizations instead of the clear German monophthongs, it was decided that it would be helpful to make the learners aware of this misconception before they started the pronunciation practice.

Production Units

For each week of pronunciation training, there were two iCPR production units. The first unit always started with explicit information about the week’s pronunciation focus. These explanations differed, depending on whether the focus was to learn a new sound as for [ʁ], [ŋ], [ç], [x], [e:], [y:], [ø:], and [ɔ:], to learn about an orthographic-phoneme-correspondence as for <z>, <ie>, <ei>, and <v>, or to learn about stress placement. When the goal was to learn the production of a new sound, metalinguistic instructions were provided, for example, lip and tongue placement were explained. When the goal was to learn about an orthographic-phoneme-correspondence, the difference between the American way to pronounce a written letter and the German way to do so were explained. The learners were also made aware of the accentedness that stems from an American way of producing that letter when speaking German. In the units

---

8 All perception and production units that were used can be accessed at: https://www.dropbox.com/sh/c22epjgij12gpu4c/AADPfHzTQc1VnSkd7WwJimUua?dl=0.
addressing stress placement, explanations were given about the nature of lexical stress, why it matters, and how to detect it. All types of explanations were supported by examples in the form of native speaker recordings. All German examples were recorded by speaker G4, the author of this dissertation. All English examples (e.g., when used to contrast cognates or orthographic-phoneme-correspondences) were recorded by speakers E1 and E2. For almost every German native speaker recording that was part of the explicit information, the learners were encouraged to imitate the recording and practice saying it out loud.

After the introductory explanations, the learners were encouraged to start practicing production. Production practice started with a prompt to produce the targeted sound in isolation by imitating a native speaker recording. Learners then had to imitate a native speaker recording of a total of 12 words containing the targeted sound. They were instructed to repeat these words out loud until they were satisfied with their own performance. Whenever possible (i.e., when phonotactic constraints and the textbook vocabulary list allowed for it) these practice items started with words in which the targeted sound appeared in initial position, then in medial or final position in short words, and finally in longer words. For all practice items, the German word appeared in large print on the left side of a PowerPoint slide, while the English translation appeared in smaller font underneath, and an icon for the recording appeared in the middle of the slide. At the end of the practice items, learners saw a “Practice Review” slide on which all practice items were listed again, together with each item’s native speaker recording. Learners received instructions to practice all items again until they were satisfied with their own performance. On the following slide, they were then prompted to open their Audacity® program (Audacity Team, 2014) and record themselves saying all of the practice items out loud (see Figure 3 for illustration; see Appendix B for an entire sample iCPR production unit). The
practice items were recorded by speakers G1, G2, G3, G5, and G6 to account for findings in HVPT—that is, multiple speakers were recruited to better represent the range of variation in natural speech and therefore allow learners to be better at identifying speech sounds when they are produced by new speakers. All recordings were done using Microsoft PowerPoint’s recording feature, recording on the internal microphone of a MacBook Pro 2010. Using Microsoft PowerPoint’s internal recording feature reduced the file size by about 90% compared to importing external sound files, thus making the average file size of a production unit about 10MB—a reasonable file size for learners to download and to be stored on the course management system.

*Figure 3. Sample Production Practice Slides*

The second production unit of each week’s pronunciation focus always started with a short review of the explicit information introduced in the first unit: metalinguistic explanations for sound formation or rules for orthographic-phoneme-correspondences were reiterated. Then, 12
more practice items were provided in the same pattern as in the first unit. In the units on [e:], [y:], [ø:], and [o:], initially, only six practice items were provided and then a second block of explicit information followed, targeting the difference between the tense and lax versions of these four vowels. After this additional block of explicit information, an additional six practice items were given and at the end of the unit, all 12 items were again listed for practice on one slide.

3.1.2.1.2 In-class Pronunciation Exercises

General Structure of In-class Exercises

The in-class pronunciation training materials consisted of 30 days of instruction materials, split into ten sets of three days of instruction—just as the homework-based iCPR units. The ten sets addressed the same conditions as introduced in Section 2.4.1.3 and used the same vocabulary items from Chapter 1-4 of Sag mal (Anton, Barske, Grabowski, McKinstry, 2014). It has to be pointed out, however, that a 100% overlap between vocabulary items practiced in the iCPR and the in-class pronunciation training was not possible due to the different nature of the two training modalities.

The pronunciation materials were designed for a 10 minutes pronunciation intervention at the beginning of the lesson on three days a week. The lesson plans for this intervention, which were provided for the instructors by the author of this dissertation, followed the same outline for perception and production training as the homework-based instruction method: the first day of instruction focused on perception exercises, and the following two days of instruction targeted production exercises. This design was only altered for instruction on [e:] and [o:], which started

---

9 Again, all perception and production iCPR units can be accessed at: https://www.dropbox.com/sh/c22epjgj12gpu4c/AADPfHzTQc1VnSkd7WwJjmUua?dl=0.
with a day of production training, followed by a day of perception training and another day of production training—this approach was in line with the slightly altered approach in the iCPR training as explained above\textsuperscript{10}.

Perception Exercises

The in-class perception materials were identical to the homework-based perception materials to ensure that the in-class learners could also benefit from a perception training approach based on findings in HVPT building on the input of multiple speakers. The Microsoft PowerPoint units including the accentedness detection and sound discrimination exercises recorded by several different speakers were brought to the classroom and presented in front of the class by the instructor. The material design only differed in that the learners received a blue and a green card and were instructed to raise the blue card when the first recording played by the instructor was the correct answer and to raise the green card when the second recording was the correct answer. The instructor was instructed to play each recording at least twice. The next slide always provided the correct answer for the learners.

Production Exercises

The in-class production materials were inspired by the materials used in Roccamo (2014). Instructors were given detailed instructions on how to introduce a new sound and how to practice this sound with their students in the classroom. First, instructors usually introduced the targeted sound of the day and then explained the importance of its focus, e.g. when the focus was on an

\textsuperscript{10} All lesson plans, perception, and production exercises can be accessed at: https://www.dropbox.com/sh/dkufsimmh7lhgr8c/AAARBWo5EA6Nyn2BbWMuGLvEa?dl=0.
orthographic-phoneme-correspondence, they explained why it was important to learn to map the German sound to the particular letter. If the focus was on learning a new sound or on stress placement, they provided metalinguistic instructions. All these materials were designed to parallel the information given in the iCPR units as closely as possible. This first step was always followed by a handout that the learners worked on to practice their pronunciation of the targeted sound. All of these exercises were designed to draw the learners’ attention to the importance of pronunciation for comprehensibility, meaning, and accentedness. That is, many exercises focused on meaningful minimal pairs (e.g. *Nacht* – *nackt* ‘night’ – ‘naked’ when practicing [x]) or differences between German and English (saying German *Poster* with a monophthongized [ɑː] vs. English *poster*). Most exercises were designed as partner work, allowing for interactive practice. The learners were constantly reminded to provide peer-feedback throughout these partner exercises and were given the chance to practice the targeted sound while speaking out loud. The instructor was encouraged to walk around the room and provide additional pronunciation feedback when necessary\(^{11}\).

3.1.2.2. Testing Materials

3.1.2.2.1 Perception Test

All items included in the perception test were unknown to the learners at the time of the pre- and posttest and were chosen to satisfy the following conditions: a) words did not appear in any of the iCPR units or in-class training materials, b) words were not part of the vocabulary in *Sag mal* chapters 1-5 (which are the chapters covered in GER 001), c) words were not English-

\(^{11}\) Again, all production exercise handouts are accessible at https://www.dropbox.com/sh/dkufsimh7lhqr8c/AAARBWo5EA6Nyn2BbWMuGLvEa?dl=0.
German cognates (except for the stress items in the cognate condition, see below), d) words could only satisfy exactly one testing condition (e.g., Tür ‘door’ could not serve as an item because it contains two conditions: [y:] and [g], but Tüte ‘bag’ could serve as an item because it only contains one condition: [y:]), and e) words appearing in chapters 6-10 of the textbook (i.e., words that will be treated in GER 002) were preferred in choosing testing items. These rules, however, did not apply for items in the cognate stress condition. The instructional goal for this condition was different in that the learners were supposed to train stress placement in certain high frequency English-German cognates, rather than to acquire a new generally transferable skill as was the instructional goal on the segmental level. Thus, all items tested in the cognate stress condition were included in the pronunciation training on purpose. Moreover, it was not possible to find high-frequency cognates that did not include any of the sounds targeted in the segmental conditions, therefore, rule d) did not apply to items in the cognate stress condition.

Just like the perception training outlined above, the perception test contained two types of exercises: an accentedness detection task and a sound discrimination task. The accentedness detection task consisted of 24 stimuli, two stimuli for each of the following conditions: [ʁ], [v], [ç], [x], [e:], [y:], [o:], [œ:], [œ:], <z>, <ie>, <ei>, and cognate stress. The <v> condition was not tested in the perception task (see Section 3.1.2.1.1 for an explanation). The sound discrimination task consisted of 20 stimuli, two stimuli for each of the above-mentioned conditions, minus the stimuli on cognate stress and [v]. Testing of cognate stress or [v] does not make sense in nonsense words as a nonsense word cannot be a cognate and the appearance of [v] depends on phonotactic constraints that cannot be reliably fulfilled in simple CVC or VCV nonsense words. All stimuli were recorded by speakers G1, G4, G5, G6, E1 and E3 to strike a balance between male and female speakers. The same speaker always produced both tokens for one item (both
the accented and standard German version of a real word and the nonsense word). Speakers were
instructed by the investigator on how to produce the accented version of each item. These
recordings were completed after the speakers had completed all the training items and were
already experienced at producing accented items. Just as in the perception training, all recordings
were delivered in a Microsoft PowerPoint file. The accentedness perception task preceded the
sound discrimination task. Within each task, all items appeared in randomized order (see
Appendix C for the handout for the perception task, which also serves as a list of all stimuli).

The posttest contained the same items as the pretest, but they appeared in a new semi-
randomized order in both tasks.

3.1.2.2 Production Test

The production pretest consisted of two tasks: a word reading and a paragraph reading
task. The items in the word reading task were chosen according to the same constraints as
outlined above for the perception test items. The word reading task consisted of 75 words: seven
words each for [ʁ], [œ], [x], [ɛ:], [y:], [ø:], [o:], as well as for the cognate stress condition,
and four words each for <z>, <ie>, and <ei>. Since no follow-up analyses are planned for the
items in the orthographic-phoneme-correspondence condition and the overall number of items
was to be kept as small as possible to minimize testing fatigue in the learners, it was decided to
only include four instead of seven items in this condition. The 75 items were split into three
blocks of 25 words. Each block contained two items from all conditions that had seven test items
and one item from the conditions that had four test items. The remaining test items were equally
distributed among the three blocks (see Appendix D for a list of the three blocks). The paragraph
reading task consisted of six paragraphs of approximately 45 words each of coherent text. Each
paragraph contained unknown items for at least ten out of the thirteen targeted segmental and suprasegmental conditions. In three paragraphs, inclusion of a targeted sound was achieved by using common German names containing the targeted sound. Paragraphs were designed to strike a balance between inclusion of high-frequency German words containing the targeted conditions, and creating an authentic, plausible short reading text. While <v> was not included in testing on the word reading level, it was tested on the paragraph reading level (see Appendix E for an overview of all reading paragraphs).

The word and paragraph reading tasks were delivered in the form of a Microsoft PowerPoint presentation. This presentation started with instructions for the tasks and included two practice items. These items were followed by three blocks of 25 words to read out loud in the carrier phrase Ich habe das Wort XX gesagt ‘I said the word XX’. Since all words were unknown to the learners, they saw a picture of the word’s meaning on each slide. Each block of 25 words was followed by two of the six reading paragraphs. Since these paragraphs contained several words unknown to the learners, they were first presented with an English version of the paragraph, illustrated by several pictures showing the content of the paragraph. The same illustration was then also provided on the slide containing the German reading paragraph to ensure face validity (see Appendix F for a shortened version of the production testing PowerPoint slides, which illustrates this method). The reasons that the production task was broken into three blocks containing both word and paragraph readings were on the one hand to fight testing fatigue, but also to allow for three separate recordings in order to keep file size reasonable.
3.1.2.3 Language Background Questionnaire

The Language Background Questionnaire was adapted from Henry (2015) and Roccamo (2014). It was a standard three-page document with questions concerning biographical information like gender and age, language learning history, and study abroad experience. Additionally, learners were asked to rate their German and English language proficiency in reading, writing, speaking, and comprehension on a scale from 1 to 10, with 10 being the best possible rating (see Appendix G for the entire Language Background Questionnaire).

3.1.2.4 Participant Exit Questionnaire

There were two versions of the Participant Exit Questionnaire, one version for learners who had received the homework-based iCPR training and one version for learners who had received in-class pronunciation instruction. The questionnaire for the iCPR group was three pages long. The first page was comprised of seven statements about the perceived usefulness of the iCPR pronunciation training to the learners. These statements had to be rated on a scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”. The second page was comprised of seven questions targeting the learners’ behavior in working through the pronunciation units, for example, time spent on each unit or how often they practiced saying each word. These seven questions had a limited number of possible responses per question. The third page was comprised of five open ended questions asking what the learners liked best or did not like about the pronunciation homework, and what suggestions they had for improving the units (see Appendix H for the entire questionnaire).

The questionnaire for the in-class pronunciation instruction group was slightly shorter since it did not serve the additional goal of assessing specific elements of the new method of
iCPR pronunciation training. Thus, this questionnaire was a two-page subset of the previously described questionnaire, comprising of only five statements to be rated, three questions with a limited number of possible answers and five open ended questions (see Appendix I for the entire questionnaire).

3.1.4 Procedure

3.1.4.1 Overview of Experimental Procedure

A total of six sections of GER 001 participated in this part of the study. These sections were divided as follows: two sections served as a treatment group and received homework-based pronunciation instruction via iCPR units, two sections served as a comparison group and received in-class pronunciation training, and two sections served as a control group and did not receive a particular pronunciation training, but followed the standard GER 001 curriculum.

All groups received the same amount of total German instruction. It was controlled that both the treatment and comparison groups had the same amount of pronunciation as well as overall German instruction: learners in the comparison group, who completed pronunciation training in the classroom, received an additional ten minutes of grammar exercises as homework as compared to the treatment group. Learners in the treatment group, who completed pronunciation training as homework, worked on the grammar exercises that were assigned to the comparison group as homework in class, so that they freed up ten minutes of homework time to complete the iCPR units. The control group, who did not receive targeted pronunciation instruction, received additional grammar or vocabulary practice as homework, ensuring that, overall, all groups had the same amount of in-class time and spent a comparable amount of time on homework assignments.
Three instructors participated in this part of the study. In order to control for the effect of instructor, the sections were divided such that one instructor each taught one section of the treatment group and one section of the comparison group. The third instructor taught both sections of the control group (see also Section 3.1.1).

Data collection took place in Spring and Fall 2016. In each semester, pretest data collection took place in Week 2 of the semester, pronunciation treatments were administered in Weeks 3-13 with one week pause for midterm exams, and posttest data collection took place in Week 14 of the semester.

All learners who were registered for one of the six sections participating in this study had to partake in all parts of this study as relevant for their respective groups. However, their data were only used if they gave consent to participate.

3.1.4.2 Testing Procedure

The pre- and posttest procedure was identical for all three groups. During Week 2 of the semester, the author of this dissertation went to all participating sections of GER 001 to introduce the study, collect consent for participation, and administer the language background questionnaire. The next day, still during Week 2 of the semester, each instructor completed the perception pretest with the learners in their class sections. The perception pretest was identical for all sections and groups. The instructors opened the PowerPoint file and distributed the handouts to the learners. They then administered the accentedness perception task followed by the sound discrimination task by moving through the PowerPoint file and playing each embedded sound file twice for the learners. The learners had about 15 seconds to note down their answers for each item. The perception task took a total of 15 minutes. At the end of that day’s
lesson, the instructors assigned the production task as homework in all groups. Learners had to download the PowerPoint file from their course management system, complete the word- and paragraph readings while recording themselves using Audacity software, and upload their recordings to the course management system. Completing the production task took a total of 15-20 minutes. As described in the materials section, the production task consisted of three blocks. These blocks were counter-balanced between sections and instructors so that different sections within one group (e.g., the treatment group) would receive them either in the order 123 or 312.

In Week 14 of the semester, the instructors followed the same procedure to administer the perception and production posttests. The perception posttest was identical to the pretest, but the tokens appeared in a new randomized order. The production posttest was identical to the pretest, but each section that received the blocks in the order 123 at the time of the pretest, now received them in the order 312 for the posttest. After all posttest data were collected, the author of this dissertation went to all participating sections, explained the goal and purpose of the study, answered remaining questions, and brought candy to thank the learners for their participation.

3.1.4.3 Pronunciation Training Procedure

The treatment group received the pronunciation training in the form of iCPR units, whereas the comparison group received the pronunciation training in the form of in-class pronunciation exercises. Training for all conditions took place on the same days of the semester across both groups. Both instruction modalities were designed to take about ten minutes a day, for a total of 30 minutes of pronunciation instruction per week. As explained above, it was controlled that both groups had the same amount of overall German instruction.
The daily procedure for the homework-based pronunciation instruction was as follows: the instructors assigned the iCPR units as part of the learners’ homework. Learners then downloaded the current iCPR unit from their respective course management systems (Angel or Canvas) and worked through the unit. Both perception and production units were designed to take the learners approximately ten minutes, however, individual differences might have occurred. At the end of the production units, learners were prompted to record a list of words that they had practiced and to upload these recordings to Angel/Canvas. They did not receive feedback on their recordings, but uploading these recordings allowed the investigator of this study to track which learner completed their homework.

The daily procedure for the in-class pronunciation training was as follows: pronunciation instruction took place at the beginning of class on three days a week. The instructor of each course carried out the detailed instructions provided to them in the lesson plans. On the first day of the week, this usually meant that the instructors administered the perception training. They downloaded the PowerPoint file provided for them on Angel/Canvas, opened it and played each recording twice. Meanwhile the learners were prompted to participate in the perception training by using colored response cards. Solutions were given in the PowerPoint file, it was not the role of the instructors to provide solutions. On the remaining two days, the instructors administered the production training. Often, they were prompted to give a short explanation at the beginning of this training, then they handed out the exercise sheets and walked around to provide help if needed, while the learners worked on the pronunciation exercises with a partner.
3.1.5 Scoring and Data Analysis

3.1.5.1 Perception

The perception task was a binary forced-choice task. Learners had to choose between option #1 and option #2 for each item in both exercises on a handout, and received one point for choosing the correct answer and zero points for choosing the incorrect answer or skipping the item. A total of 44 points was possible: 24 points in exercise 1, and 20 points in exercise 2. The answers were coded by research assistants, and 25% of the data were re-coded by the author of this dissertation, with an interrater reliability of 98.9%. Ceiling performance was determined at 23 of 24 points on exercise 1 and 19 of 20 points on exercise 2. No learner performed at ceiling. Therefore, no learner was excluded from the analyses.

As not all of the data were normally distributed (according to Shapiro-Wilks tests as recommended by Larson-Hall, 2016: treatment group dataset \( p = .028 \)) and variances were not equal, I present the results of nonparametric tests and provide standardized estimates of the effect (Cohen’s \( d \)), and an estimate of the error associated with the comparison (95% confidence intervals) whenever possible (Norris, Plonsky, Ross & Schoonen, 2015)\(^{12} \). All analyses were carried out separately for exercises 1 and 2. Since my principal interest was to see whether the type of treatment would lead to differences in perception skills between the three groups at the time of the posttest, my primary analyses target between-group differences. I first compared pretest scores between the treatment, comparison and control groups using Kruskal-Wallis tests to confirm that there were no significant differences between the groups prior to the treatment. I then compared posttest scores in the same manner and followed-up the Kruskal-Wallis tests with Mann-Whitney tests to compare group pairs wherever the Kruskal-Wallis tests indicated a

\(^{12}\) For parity with other studies, I carried out parametric tests (one-way ANOVAs and repeated measures ANOVAs with planned contrasts) and found the same patterns of findings as for nonparametric tests.
statistically significant difference in mean scores. Since I was also interested in gains over time for each group, I finally compared pretest and posttest scores for each group using Wilcoxon signed-rank tests.

As recommended by Plonsky and Oswald (2014), Cohen’s $d$ field-specific benchmarks were used for interpretation and effect sizes for between-group comparisons (e.g., treatment group vs. control group at posttest) were considered large when $d = 1.00$, medium when $d = 0.70$, and small when $d = 0.40$. Effect sizes for within-group comparisons (e.g. treatment group at pretest vs. at posttest) were considered large when $d = 1.40$, medium when $d = 1.00$, and small when $d = 0.60$. Since the effect sizes between the three groups at the pretest were very similar, and therefore there were no baseline differences that had to be accounted for, no adjusted effect sizes will be reported here.

3.1.5.2 Production

3.1.5.2.1 Rating Procedure

Materials

The production pre- and posttest yielded three audio files per person for each testing time. These six audio files per person were spliced into shorter parts using Audacity 2.0. For the word-level production, words were spliced from the carrier phrase and each word was extracted and saved as an individual file. Following the conventions of previous studies (Derwing & Munro, 2013; Foote & McDonough, 2017; O’Brien, 2014), for the paragraph-level productions, the first 20 seconds of each paragraph reading were extracted and initial disfluencies were removed. After the splicing, all audio files were normalized by scaling them to a peak intensity of 70db and by inserting 500ms of silence, using a Praat script (Boersma & Weenink, 2014).
Twenty-four words (two from each of the 12 targeted conditions outlined in Section 3.1.2.2.2) and three paragraphs per person were randomly chosen and included in the pool of final ratings. These words and paragraphs were consistent between the pre- and posttest such that the raters heard the same word produced by the same learner twice—once taken from the pretest and once taken from the posttest recording. The total productions per learner that were included in the final analyses thus were comprised of 48 individual words and six 20-seconds excerpts from the paragraph readings. Additionally, 48 words and six paragraphs recorded by four native speakers and four near-native speakers were included in the ratings. They served as a check on the raters’ use of the rating scales—that is, failure to assign high scores to the native speakers might have indicated unreliable ratings.

Procedure

The rating procedure was the same as that followed in numerous other studies investigating gains in L2 pronunciation (Crowther, Trofimovich, & Isaacs, 2016; Derwing & Munro, 2013). Eight native-speaker raters were recruited to rate the learners’ productions (for details on the raters see Section 3.1.1). Raters were paid an hourly rate for their participation. Each rater rated data from 27 learners plus half of the native and near-native speaker data as a control measure. Thus, each rater rated a total of 1320 word productions, and 165 paragraph productions for a total of approximately 360 minutes or six hours of rating per rater. Since six hours of ratings were deemed as the maximum amount of time that raters could be recruited for, not all eight raters rated all 49 learners. Instead, the 49 learners were randomly distributed to two lists with an overlap of five learners. That is, each list contained word- and paragraph-level productions of 22 learners, plus five learners that appeared on both lists and were included to
check for inter-rater consistency between the raters in the two groups. Each list was then assigned to four raters, yielding two groups of raters that each rated data from 27 learners plus half of the native speaker productions.

Multiple rating sessions were conducted in a quiet room to accommodate the raters’ schedules. Each of these rating sessions began with a training to familiarize and re-familiarize the learners with the rating scales and the procedure. The training was comprised of instructions on how to rate the speech samples as well as five practice items each for the word-level ratings and for the paragraph-level ratings. The training also contained an explanation of the speech continua targeted in this study: comprehensibility and accentedness. While it would have also been useful to assess gains in intelligibility, this could not be a principal measure in this study because raters rated the same 75 words and six paragraphs for all learners. To assess intelligibility, however, raters are asked to transcribe what they hear, which is not a reliable measure when hearing the same items over and over again. Since intelligibility and comprehensibility are highly correlated (Derwing & Munro, 2015), this study therefore followed the methodology of Gordon and Darcy (2016) and Bergeron and Trofimovich (2017), and only collected comprehensibility ratings to serve as a general metric of understanding in a broad sense (see also, Levis, 2005). The raters were then asked to judge the comprehensibility and accentedness of each word or paragraph they had listened to on a 9-point Likert scale. The 9-point rating scale was selected because it has been used successfully in other studies and has been shown to result in high inter-rater reliability ratings (Derwing, Munro, & Wiebe, 1998; Gordon & Darcy, 2016; Kennedy & Trofimovich, 2008, O’Brien, 2016). The raters heard each item only once before making their judgments. Based on findings by O’Brien (2016) that showed no significant differences for the comparison between rating a speech sample along three
continua after it is heard once or rating a speech sample along one continuum at a time (i.e.,
listening to it three times, once for each rating step), raters heard each item once and then had to
enter their rating for comprehensibility and for accentedness on two consecutive screens. Each
screen displayed a question targeting the speech continuum to be rated (Wie einfach war es für
Sie, den Sprecher zu verstehen? ‘How easy was it for you to understand the speaker?’ and Wie
gut war der Akzent des Sprechers? ‘How good was the speaker’s accent?’) and a short reminder
of the values of 1 and 9, but no elaborate descriptors in between (1 = sehr schwierig/unmöglich
‘very difficult/impossible’, 9 = ganz einfach ‘very easy’; 1 = der Akzent war extrem stark ‘the
accent was very heavy’, 9 = der Sprecher hatte eigentlich gar keinen Akzent, er könnte
Muttersprachler sein ‘the speaker had almost no accent, he could be a native speaker’).

The 1320 individual words and 165 paragraph excerpts that were presented to each rater
were randomly distributed to 12 blocks of word ratings and eight blocks of paragraph ratings.
Each block took the raters between 15 and 20 minutes to complete. Raters were asked to
alternate between word and paragraph blocks and were encouraged to take pauses between each
block. The presentation of all blocks was counterbalanced between the four raters in each group.
Raters came to the language lab between three and five times to complete the ratings of all
blocks. All audio files were presented to the raters using high-quality headphones, and raters
performed the task individually on a Mac laptop computer that employed PsychoPy software
(Peirce, 2016) to present the audio stimuli in randomized order within each block.

Intraclass correlation coefficients were used to calculate the inter-rater reliability of the
raters. As shown in Table 3, high inter-rater consistency within each group of four raters was
observed for all raters on both continua of ratings (Cronbach’s \( \alpha = .82-.91 \)), in all cases
exceeding the recommended .70-.80 benchmark (Larson-Hall, 2016). Reliability values were
thus deemed sufficiently high to calculate mean comprehensibility and accentedness scores for each speaker, averaging across all raters, separately for each speech continuum. Moreover, since there was excellent inter-rater reliability for the five learners that were rated by all eight raters (Cronbach’s $\alpha = .90-.95$), we can assume that the raters in both groups rendered similar ratings overall. The average values for each learner then constituted the data points for the production analyses for comprehensibility and accentedness, presented in the following section.

Table 3: Experiment 1 – Rater Consistency (Cronbach’s $\alpha$, [95% CI]) for Rated Continua by Task and Rater Group

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (raters 1-4)</th>
<th>Group 2 (raters 5-8)</th>
<th>Group 1&amp;2 overlap (5 learners rated by raters 1-8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.89</td>
<td>.91</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>[.88, .90]</td>
<td>[.90, .92]</td>
<td>[.94, .96]</td>
</tr>
<tr>
<td>Accentedness</td>
<td>.85</td>
<td>.86</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>[.84, .86]</td>
<td>[.85, .87]</td>
<td>[.90, .94]</td>
</tr>
<tr>
<td><strong>Paragraphs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.82</td>
<td>.83</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>[.78, .86]</td>
<td>[.78, .87]</td>
<td>[.91, .97]</td>
</tr>
<tr>
<td>Accentedness</td>
<td>.83</td>
<td>.83</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>[.78, .87]</td>
<td>[.78, .87]</td>
<td>[.83, .95]</td>
</tr>
</tbody>
</table>

3.1.5.2.2 Statistical Analyses

Standard practice before analyzing rating data is to exclude learners who performed at ceiling at the time of the pretest. Ceiling performance for this task was set at 8 with 9 being the highest possible rating. No learner received an average comprehensibility or accentedness score higher than 6.79. Therefore, no data were excluded. Moreover, it was confirmed that all raters recognized native speakers and near-native speakers as such and used the entire rating scale.
As not all of my data were normally distributed (according to Shapiro-Wilks tests as recommended by Larson-Hall, 2016: words accentedness control group dataset \( p = .043 \) and paragraph accentedness control group dataset \( p = .018 \)) and variances were not equal, I present the results of nonparametric tests and provide standardized estimates of the effect (Cohen’s \( d \)) and an estimate of the error associated with the comparison (95% confidence intervals) whenever possible (Norris, Plonsky, Ross & Schoonen, 2015). All analyses were carried out separately for measures of comprehensibility and accentedness, as well as for the word- and paragraph-level productions. Since my principal interest was to see whether the type of treatment would lead to differences in comprehensibility and in accentedness between the three groups at the time of the posttest, my primary analyses target between-group differences. I first compared pretest scores between the treatment, comparison and control groups using Kruskal-Wallis tests to confirm that there were no significant differences between the groups prior to the treatment. I then compared posttest scores in the same manner and followed-up the Kruskal-Wallis tests with Mann-Whitney tests to compare group pairs wherever the Kruskal-Wallis tests indicated a statistically significant difference in mean scores. Since I was also interested in gains over time for each group, I finally compared pretest and posttest scores for each group using Wilcoxon signed-rank tests.

As recommended by Plonsky and Oswald (2014), effect sizes for between-group comparisons (e.g., treatment group vs. control group at posttest) were considered large when \( d = 1.00 \), medium when \( d = 0.70 \), and small when \( d = 0.40 \). Effect sizes for within-group comparisons (e.g. treatment group at pretest vs. posttest) were considered large when \( d = 1.40 \), medium when \( d = 1.00 \), and small when \( d = 0.60 \). Since the effect sizes between the three groups

---

13 For parity with other studies, I carried out parametric tests (one-way ANOVAs and repeated measures ANOVAs with planned contrasts) and found the same patterns of findings as for nonparametric tests.
at the pretest were very similar and therefore there were no baseline differences that had to be accounted for, no adjusted effect sizes will be reported here.

3.2 Results

3.2.1 Perception

3.2.1.1 Accentedness Detection Task

For a visual representation of the results, see Figure 4 and Figure 5. For descriptive statistics beyond the visual representation, see Table 4. A Kruskal-Wallis test did not show a significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.22, p = .896$), but it revealed a significant difference at the time of the posttest ($X^2(2) = 8.18, p = .017$). Follow-up Mann-Whitney tests revealed a significant difference between the treatment and control group ($U = 58.00, z = -2.68, p = .007, d = 1.10$), but not between the comparison and control group ($U = 76.50, z = -1.50, p = .142, d = 0.59$) and not between the treatment and comparison group ($U = 73.50, z = -1.62, p = .110, d = 0.64$). With a between-group effect size of $d = 1.10$, the difference in posttest scores between the treatment group and the control group would be considered a large effect. The difference between the comparison group and the control group ($d = 0.59$) would be considered a small effect. The difference between the treatment group and the comparison group ($d = 0.64$) approached the benchmark for a medium-sized effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -3.12, p = .002, d = 1.33$) but not for the comparison group ($z = -1.69, p = .091, d = 0.57$) and for the control group ($z = -0.47, p = .641, d = 0.10$). With a within-group effect size of $d = 1.33$, the gains over time in the treatment group approached the benchmark for a large effect. The gains over time in
the comparison group approached the benchmark for a small effect \((d = 0.57)\) and the gains in the control group \((d = 0.10)\) would be considered negligible.

*Figure 4. Experiment 1 – Accentedness Detection Task – Between-group Results*

Note: y-axis: percentage correct; \(n.s. = p > .05\); \(* = p \leq .05; ** = p \leq .01; *** = p \leq .001; \) – = negligible effect size \((d < 0.4)\); \# = small effect size \((0.4 \leq d < 0.7)\); ### = medium effect size \((0.7 \leq d < 1.0)\); #### = large effect size \((d \geq 1.0)\)
Figure 5. Experiment 1 – Accentedness Detection Task – Within-group Results

Table 4: Experiment 1 – Descriptive Statistics for Accentedness Detection Task

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
</tr>
<tr>
<td>Treatment</td>
<td>82.81 (6.06)</td>
<td>[79.58, 86.04]</td>
<td>83.33 (11.5)</td>
<td>91.93 (7.53)</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>82.44 (10.86)</td>
<td>[76.17, 88.71]</td>
<td>85.42 (20.8)</td>
<td>87.50 (6.12)</td>
</tr>
<tr>
<td>(n = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>82.29 (6.72)</td>
<td>[78.71, 85.87]</td>
<td>83.33 (10.2)</td>
<td>83.08 (8.53)</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.1.2 Sound Discrimination Task

For a visual representation of the results, see Figure 6 and Figure 7. For descriptive statistics beyond the visual representation, see Table 5. A Kruskal-Wallis test did not show a
significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.77, p = .682$) and at the time of the posttest ($X^2(2) = 3.07, p = .215$). Due to the lack of significance in the posttest analyses, no follow-up Mann-Whitney tests were performed, but effect sizes show the following: With a between-group effect size of $d = 0.67$, the difference in posttest scores between the treatment group and the control group approached the benchmark for a medium-sized effect. The difference between the comparison group and the control group ($d = 0.47$) would be considered a small effect. The difference between the treatment group and the comparison group ($d = 0.22$) would be considered a negligible effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for all three groups: treatment group ($z = -3.15, p = .002, d = 1.56$), comparison group ($z = -2.47, p = .014, d = 0.83$), and control group ($z = -2.36, p = .018, d = 0.37$). With a within-group effect size of $d = 1.56$, the gains over time in the treatment group would be considered a large effect. The gains over time in the comparison group approached the benchmark for a medium effect size ($d = 0.83$) and the gains in the control group ($d = 0.37$) would be considered negligible.
Figure 6. Experiment 1 – Sound Discrimination Task – Between-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; -- = negligible effect size (d < 0.4); # = small effect size (0.4 ≤ d < 0.7); ### = medium effect size (0.7 ≤ d < 1.0); #### = large effect size (d ≥ 1.0)

Figure 7. Experiment 1 – Sound Discrimination Task – Within-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; -- = negligible effect size (d < 0.6); # = small effect size (0.6 ≤ d < 1.0); ### = medium effect size (1.0 ≤ d < 1.4); #### = large effect size (d ≥ 1.4)
Table 5: Experiment 1 – Descriptive Statistics for Sound Discrimination Task

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 16)</td>
<td>72.50</td>
<td>[67.64, 77.37]</td>
</tr>
<tr>
<td>Comparison</td>
<td>74.29</td>
<td>[66.62, 81.95]</td>
</tr>
<tr>
<td>(n = 14)</td>
<td>(9.13)</td>
<td>(7.23)</td>
</tr>
<tr>
<td>Control</td>
<td>75.31</td>
<td>[69.77, 79.08]</td>
</tr>
<tr>
<td>(n = 16)</td>
<td>(10.40)</td>
<td>(10.06)</td>
</tr>
</tbody>
</table>

3.2.2 Production

3.2.2.1 Comprehensibility

3.2.2.1.1 Word-level Productions

For a visual representation of the results, see Figure 8 and Figure 9. For descriptive statistics beyond the visual representation, see Table 6. A Kruskal-Wallis test did not show a significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.06, p = .972$), but it revealed a significant difference at the time of the posttest ($X^2(2) = 7.89, p = .019$). Follow-up Mann-Whitney tests revealed a significant difference between the treatment and control group ($U = 80.00, z = -2.38, p = .017, d = 0.88$), between the comparison and control group ($U = 52.50, z = -2.47, p = .012, d = 0.93$), but not between the treatment and comparison group ($U = 129.50, z = -0.13, p = .900, d = 0.05$). With a between-group effect size of $d = 0.88$, the difference in posttest scores between the treatment group and the control group would be considered a medium-to-large effect. The difference between the comparison group and the control group ($d = 0.93$) approached the benchmark for a large effect. The difference between the treatment group and the comparison group ($d = 0.05$) would be considered a negligible effect. Wilcoxon signed-rank tests revealed a significant difference
between mean pretest and posttest scores for all three groups: treatment group \((z = -3.68; p < .001, d = 1.64)\), comparison group \((z = -3.17, p = .002, d = 1.75)\), and control group \((z = -2.69, p = .007, d = 0.52)\). With a within-group effect size of \(d = 1.64\) in the treatment group and \(d = 1.75\) in the comparison group, the gains over time in these two groups would be considered a large effect. The gains over time in the control group \((d = 0.52)\) almost reached the benchmark to be considered a small effect.

*Figure 8. Experiment 1 – Comprehensibility on Word-level Productions – Between-group Results*

Note: \(y\)-axis: percentage correct; \(n.s. = p > .05; * = p \leq .05; ** = p \leq .01; *** = p \leq .001; ~ = negligible effect size \((d < 0.4)\); # = small effect size \((0.4 \leq d < 0.7)\); ## = medium effect size \((0.7 \leq d < 1.0)\); ### = large effect size \((d \geq 1.0)\)
Figure 9. Experiment 1 – Comprehensibility on Word-level Productions – Within-group Results

![Figure 9](image)

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.6$); # = small effect size ($0.6 \leq d < 1.0$); ### = medium effect size ($1.0 \leq d < 1.4$); #### = large effect size ($d \geq 1.4$)

Table 6: Experiment 1 – Descriptive Statistics for Comprehensibility on Word-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th></th>
<th>Posttest</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>$IQR$</td>
<td>$M$ (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Treatment</td>
<td>5.65 (.53)</td>
<td>[5.39, 5.90]</td>
<td>5.70 (.92)</td>
<td>6.65 [.68, 6.98]</td>
<td>6.65 [6.32, 6.58]</td>
<td>1.16</td>
</tr>
<tr>
<td>(n = 19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>5.62 (.56)</td>
<td>[5.30, 5.94]</td>
<td>5.61 (.84)</td>
<td>6.68 [6.30, 7.06]</td>
<td>6.82 [6.30, 7.06]</td>
<td>1.00</td>
</tr>
<tr>
<td>(n = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>5.63 (.65)</td>
<td>[5.29, 5.98]</td>
<td>5.56 (1.27)</td>
<td>6.01 [5.59, 6.42]</td>
<td>5.88 [5.59, 6.42]</td>
<td>1.42</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.2.1.2 Paragraph-level Productions

For a visual representation of the results, see Figure 10 and Figure 11. For descriptive statistics beyond the visual representation, see Table 7. A Kruskal-Wallis test did not show a
significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.19, p = .908$) and at the time of the posttest ($X^2(2) = 3.37, p = .185$). Due to lack of significance in the posttest analyses, no follow-up Mann-Whitney tests were performed. With a between-group effect size of $d = 0.51$, the difference in posttest scores between the treatment group and the control group would be considered a small effect. The difference between the comparison group and the control group ($d = 0.39$) approached the benchmark for a small effect. The difference between the treatment group and the comparison group ($d = 0.08$) would be considered a negligible effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for all three groups: treatment group ($z = -3.79; p < .001, d = 1.06$), comparison group ($z = -2.99, p = .003, d = 0.79$), and control group ($z = -3.47, p = .001, d = 0.71$). With a within-group effect size of $d = 1.06$, the gains over time in the treatment group would be considered a medium-sized effect. The gains over time in the comparison group ($d = 0.79$) would be considered a small-to-medium sized effect and the gains in the control group ($d = 0.52$) would be considered a negligible effect.
Figure 10. Experiment 1 – Comprehensibility on Paragraph-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.4); # = small effect size (0.4 ≤ d < 0.7); ## = medium effect size (0.7 ≤ d < 1.0); ### = large effect size (d ≥ 1.0)

Figure 11. Experiment 1 – Comprehensibility on Paragraph-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.6); # = small effect size (0.6 ≤ d < 1.0); ## = medium effect size (1.0 ≤ d < 1.4); ### = large effect size (d ≥ 1.4)
Table 7: Experiment 1 – Descriptive Statistics for Comprehensibility on Paragraph-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mdn</td>
<td>IQR</td>
<td>M (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td>5.94 (.90)</td>
<td>[5.51, 6.38]</td>
<td>6.08 1.25</td>
<td>6.85 (.80)</td>
</tr>
<tr>
<td>(n = 19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td></td>
<td>5.98 (1.08)</td>
<td>[5.36, 6.60]</td>
<td>6.21 1.50</td>
<td>6.78 (.96)</td>
</tr>
<tr>
<td>(n = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>5.97 (.72)</td>
<td>[5.59, 6.35]</td>
<td>5.79 1.18</td>
<td>6.47 (.68)</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.2.2 Accentedness

3.2.2.2.1 Word-level Productions

For a visual representation of the results, see Figure 12 and Figure 13. For descriptive statistics beyond the visual representation, see Table 8. A Kruskal-Wallis test did not show a significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.72, p = .696$), but it revealed a significant difference at the time of the posttest ($X^2(2) = 12.49, p = .002$). Follow-up Mann-Whitney tests revealed a significant difference between the treatment and control group ($U = 59.00, z = -3.08, p = .002, d = 0.86$), between the comparison and control group ($U = 39.50, z = -3.02, p = .002, d = 0.88$), but not between the treatment and comparison group ($U = 127.50, z = -0.20, p = .843, d = 0.03$). The difference in posttest scores between the treatment group and the control group ($d = 0.86$) and between the comparison group and the control group ($d = 0.88$) would be considered a medium-to-large effect. The difference between the treatment group and the comparison group ($d = 0.03$) would be considered a negligible effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for all three groups: treatment group ($z = -3.82, p <$
.001, \( d = 1.73 \)), comparison group \( (z = -2.73, p = .006, d = 1.42) \), and control group \( (z = -2.84, p = .004, d = 0.55) \). With a within-group effect size of \( d = 1.73 \) in the treatment group and \( d = 1.42 \) in the comparison group, the gains over time in these two groups would be considered a large effect. The gains over time in the control group \( (d = 0.55) \) would be considered a negligible effect.

*Figure 12. Experiment 1 – Accentedness on Word-level Productions – Between-group Results*

Note: \( y \)-axis: percentage correct; \( n.s. = p > .05; * = p \leq .05; ** = p \leq .01; *** = p \leq .001; -- = negligible effect size \( (d < 0.4) \); # = small effect size \( (0.4 \leq d < 0.7) \); ## = medium effect size \( (0.7 \leq d < 1.0) \); ### = large effect size \( (d \geq 1.0) \)
Figure 13. Experiment 1 – Accentedness on Word-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.6$); # = small effect size ($0.6 \leq d < 1.0$); ## = medium effect size ($1.0 \leq d < 1.4$); ### = large effect size ($d \geq 1.4$)

Table 8: Experiment 1 – Descriptive Statistics for Accentedness on Word-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$95% CI$</td>
</tr>
<tr>
<td>Treatment</td>
<td>4.22 (.42)</td>
<td>[4.02, 4.42]</td>
</tr>
<tr>
<td>($n = 19$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>4.30 (.63)</td>
<td>[3.94, 4.67]</td>
</tr>
<tr>
<td>($n = 14$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>4.24 (.62)</td>
<td>[3.91, 4.57]</td>
</tr>
</tbody>
</table>
3.2.2.2 Paragraph-level Productions

For a visual representation, see Figure 14 and Figure 15. For descriptive statistics beyond the visual representation, see Table 9. A Kruskal-Wallis test did not show a significant difference between the treatment, comparison, and control groups at the time of the pretest ($X^2(2) = 0.75, p = .686$) and at the time of the posttest ($X^2(2) = 2.86, p = .239$). Due to lack of significance in the posttest analyses, no follow-up Mann-Whitney tests were performed. The difference in posttest scores between the treatment group and the control group ($d = 0.56$) and between the comparison group and the control group ($d = 0.59$) approached the benchmark for a medium-sized effect.

The difference between the treatment group and the comparison group ($d = 0.16$) would be considered a negligible effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for all three groups: treatment group ($z = -3.79; p < .001, d = 1.52$), comparison group ($z = -2.80, p = .005, d = 1.00$), and control group ($z = -2.43, p = .015, d = 0.79$). With a within-group effect size of $d = 1.59$ in the treatment group, the gains over time in this group would be considered a large effect. The gains over time in the comparison group ($d = 1.00$) would be considered a medium-sized effect and the gains over time in the control group ($d = 0.79$) would be considered a small effect.
Figure 14. Experiment 1 – Accentedness on Paragraph-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; − = negligible effect size ($d < 0.4$); # = small effect size (0.4 ≤ $d < 0.7$); ### = medium effect size (0.7 ≤ $d < 1.0$); #### = large effect size ($d \geq 1.0$)

Figure 15. Experiment 1 – Accentedness on Paragraph-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; − = negligible effect size ($d < 0.6$); # = small effect size (0.6 ≤ $d < 1.0$); ### = medium effect size (1.0 ≤ $d < 1.4$); #### = large effect size ($d \geq 1.4$)
Table 9: Experiment 1 – Descriptive Statistics for Accentedness on Paragraph-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest M (SD)</th>
<th>95% CI</th>
<th>Mdn</th>
<th>IQR</th>
<th>Posttest M (SD)</th>
<th>95% CI</th>
<th>Mdn</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (n = 19)</td>
<td>3.96 (.64)</td>
<td>[3.65, 4.27]</td>
<td>3.88</td>
<td>1.08</td>
<td>4.97 [4.64, 4.83]</td>
<td>1.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison (n = 14)</td>
<td>4.18 (.74)</td>
<td>[3.75, 4.61]</td>
<td>4.25</td>
<td>1.40</td>
<td>5.11 [4.48, 5.46]</td>
<td>1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control (n = 16)</td>
<td>4.07 (.68)</td>
<td>[3.71, 4.44]</td>
<td>3.88</td>
<td>1.23</td>
<td>4.60 [4.26, 4.67]</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.3 Summary of Findings

Table 10 summarizes the findings of Experiment 1 (for the same table with exact p-values and d-scores, see Appendix J).

Table 10: Experiment 1 – Summary of Findings

<table>
<thead>
<tr>
<th></th>
<th>Between-group comparison (at posttest)</th>
<th>Within-group comparison (pretest to posttest)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment &gt; Control</td>
<td>Comparison &gt; Control</td>
</tr>
<tr>
<td>Perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness</td>
<td>** n.s  n.s.</td>
<td>** n.s. n.s.</td>
</tr>
<tr>
<td>Detection Task</td>
<td>### # #</td>
<td>## – –</td>
</tr>
<tr>
<td>Sound Discrimination Task</td>
<td>n.s. n.s. n.s.</td>
<td>** * *</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>* * n.s.</td>
<td>*** ** **</td>
</tr>
<tr>
<td></td>
<td>## ## –</td>
<td>### ### –</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>n.s. n.s. n.s.</td>
<td>** * ***</td>
</tr>
<tr>
<td></td>
<td># – –</td>
<td>## # #</td>
</tr>
</tbody>
</table>
The between-group comparisons—that is, the comparison of posttest scores between the treatment, comparison, and control groups—revealed the following: on the perception task, learners in the treatment group significantly outperformed learners in the control group on measures of detecting accented productions of German words. No other comparisons were significant. These findings were further reflected in the effect sizes which showed a large effect for the difference between posttest scores of the treatment and control group on the accentedness detection task, but only small effect sizes and one negligible effect for all other comparisons.

On the production task, there were no significant differences between the treatment group and the comparison group, on measures of comprehensibility or accentedness on both the word- and paragraph level, and effect sizes for these comparisons were all negligible. Furthermore, for productions on the word-level, learners in both the treatment group and the comparison group significantly outperformed learners in the control group in measures of comprehensibility and accentedness. Both comparisons showed a medium-sized effect. For productions on the paragraph-level, there were no significant differences between learners in the treatment or comparison group and learners in the control group in measures of comprehensibility and accentedness. All effect sizes for these comparisons were either small or negligible.

<table>
<thead>
<tr>
<th></th>
<th>**</th>
<th>**</th>
<th>n.s.</th>
<th>***</th>
<th>**</th>
<th>**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>##</td>
<td>##</td>
<td>–</td>
<td>###</td>
<td>###</td>
<td>–</td>
</tr>
<tr>
<td>Word-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>***</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>#</td>
<td>–</td>
<td>###</td>
<td>##</td>
<td>#</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$

– = negligible effect size; # = small effect size; ## = medium effect size; ### = large effect size
The within-group comparisons—that is, the comparison from pretest to posttest within each of the three groups—revealed the following: findings on the perception task contrasted between the two types of exercises in that, on the accentedness detection task, only the treatment group improved significantly from pre- to posttest (which was also reflected in a medium-sized effect for this comparison), while on the sound discrimination task, all three groups improved significantly from pre- to posttest. In the latter comparison, data showed a large effect for the pronunciation training over time in the treatment group, a small effect in the comparison group, and a negligible effect in the control group.

On the production task, data revealed significant improvement over time in all conditions for all groups. Effect sizes, however, differed. On the word-level, the gains over time in the treatment and comparison group for both comprehensibility and accentedness showed large effects, while the gain in the control group showed a negligible effect. On the paragraph-level, the gain over time in the treatment group showed a medium-sized (comprehensibility) and a large (accentedness) effect in the treatment group, but only small effects (comprehensibility) and small-to-medium sized effects (accentedness) in the comparison and control groups.

3.2.4 Exit Questionnaire Responses

3.2.4.1 iCPR Training Group

Eighteen of the 19 learners in the treatment group completed the Exit Questionnaire. Table 11 provides an overview of the mean scores for the seven statements on the perceived usefulness of the iCPR pronunciation training in the first part of the debriefing questionnaire. These statements had to be rated on a scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”. Results showed very high values (means above 6 out of 7) for
questions 2 and 3, indicating that the learners strongly agreed that they learned something from the pronunciation homework and that they felt that their pronunciation had improved. Results showed high values (means above 5 out of 7) for questions 5 and 6, indicating that learners agreed that the benefits of the training outweighed any possible technological difficulties and that there was an added benefit in reviewing vocabulary. Results showed medium-high values (means above 4 out of 7) for questions 1 and 7, showing that learners somewhat agreed on having enjoyed working on the iCPR units and that they might have benefited from receiving teacher feedback throughout the semester. Finally, with a mean score of 2.8 on question 4, learners disagreed with the statement that they would have felt more comfortable practicing pronunciation in class rather than through homework assignments.

Table 11: Experiment 1 – Mean Scores on Exit Questionnaire – iCPR Training Group

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Statement To Be Rated By Learners</th>
<th>Mean Response (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I enjoyed working on the pronunciation homework.</td>
<td>4.5 (2-7)</td>
</tr>
<tr>
<td>2.</td>
<td>I feel like I learned something from the pronunciation homework.</td>
<td>6.2 (3-7)</td>
</tr>
<tr>
<td>3.</td>
<td>I feel like my pronunciation of German improved through the homework exercises.</td>
<td>6.4 (5-7)</td>
</tr>
<tr>
<td>4.</td>
<td>I would have felt more comfortable to practice pronunciation in-class rather than as homework.</td>
<td>2.8 (1-5)</td>
</tr>
<tr>
<td>5.</td>
<td>Any technological problems that I encountered were worth the overall benefit.</td>
<td>5.2 (2-7)</td>
</tr>
<tr>
<td>6.</td>
<td>Not only did the pronunciation homework improve my pronunciation but it also helped review some vocabulary.</td>
<td>5.5 (2-7)</td>
</tr>
<tr>
<td>7.</td>
<td>I would have taken the pronunciation homework more seriously and would have benefited more from it if I had received teacher feedback throughout the semester.</td>
<td>4.9 (2-7)</td>
</tr>
</tbody>
</table>

Note: Statements were rated on scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”.

110
The second part of the questionnaire was comprised of seven questions targeting the learners’ behavior in working through the pronunciation units, with a limited number of possible responses per question. For Question #1 (How much time, on average, did you spend on each pronunciation homework?), the majority of the learners reported spending 10-15 minutes on each iCPR unit (5-10 min: 6 learners; 10-15 min: 11 learners; 15-20 min: 1 learner; 20-25 min: 0 learners). For Question #2 (How often, on average, did you listen to the native speaker recording of each word?), half of the learners reported listening to each word 2-3 times and the other half of the learners reported listening either 1-2 times (5 learners) or 3-5 times (4 learners). No learners reported listening to the words more than 5 times. For Question #3 (How often, on average, did you practice saying each word before you recorded yourself on all the words?), the answers were distributed almost evenly between 1 and 5 times (1-2 times: 6 learners; 2-3 times: 5 learners; 3-5 times: 6 learners; 5-7 times: 1 learner; more than 7 times: 0 learners). For Question #4 (Did you always practice saying each word until you were completely happy with your own pronunciation before you recorded yourself?), the majority of learners answered “Usually yes” (always: 1 learner; usually yes: 11 learners; about half the time: 3 learners; usually not: 2 learners; never: 1 learner). For Question #5 (There was no control of whether you completed the listening exercises. Honestly, how many of the 10 listening exercise homework units did you complete?), there were responses to all options from “0/1 of 10” to “all 10” with a mode on “4/5 of 10” (all 10: 1 learner; 9 of 10: 3 learners; 8 of 10: 1 learner; 7/6 of 10: 1 learner; 5/4 of 10: 8 learners; 3/2 of 10: 2 learners; 1/0 of 10: 2 learners). For Question #6 (Did you enjoy the listening exercises or the speaking exercises more?), the majority of learners answered that they enjoyed the speaking exercises more (listening: 1 learner; speaking: 15 learners; I enjoyed both equally: 2 learners). For Question #7 (Do you think that three days of homework practice on
each pronunciation problem was enough to help you or would you have preferred more or less practice?), all learners answered that three units were a good amount (more practice: 0 learners; it was a good amount: 18 learners; less practice: 0 learners).

The third page of the questionnaire was comprised of five open-ended questions. In response to the question “What did you like about the pronunciation homework exercises?”, three learners mentioned that they liked the format of pronunciation training taking place at home:

**EXCERPT 1**
*The fact that we could be comfortable speaking and practicing without others around.* (Participant 106)

**EXCERPT 2**
*A chance to privately improve my accent and pronunciation.* (Participant 407)

Two other learners highlighted the fact of receiving pronunciation training at all:

**EXCERPT 3**
*I liked that the exercises clarified how to pronounce German sounds, as I feel this is something language classes rarely do.* (Participant 103)

Two more learners also pointed out that the exercises did not take up much time:

**EXCERPT 4**
*They didn’t take long and we could do them on our own time mostly and they were helpful.* (Participant 405)

In response to the question “What did you not like about the pronunciation homework exercises?”, five learners mentioned that the exercises were easy to forget about and that all three parts of one week’s homework should have all been assigned on a weekly, rather than on a daily basis:
They were kind of easy to forget about, especially because they were due almost every night. (Participant 112)

I often forgot about it because of other homework we did for class. (Participant 403)

Two learners complained about the length of the pre- and posttest:

I didn’t like the pre-test and post-test, as these took a very long time. (Participant 103)

Only one learner mentioned the lack of feedback as something they did not like about the pronunciation homework exercises:

There was no feedback. The sounds in made up words weren’t very helpful either. (Participant 401)

Finally, it has to be mentioned that a total of four learners reported not having anything to complain about, even when directly asked about what they did not like about the iCPR units:

I had no serious grievances about the homework. (Participant 110)

In response to the question “Do you have any suggestions for future improvements of the pronunciation exercises?” five learners suggested to not only focus on the pronunciation of words, but to embed words in sentences or even paragraphs:

To include German sentences instead of just individual words. (Participant 104)

Two learners further suggested including individualized feedback:

Feedback on areas of improvement and what we did well/poorly. (Participant 401)

Two other learners suggested making due dates more flexible:
Finally, two learners addressed the problem that students did not complete many of the listening exercises:

**Excerpt 13**
Add something to 1st exercise of week so more people do them all. (Participant 405)

**Excerpt 14**
Make the listening exercises mandatory, like put a quiz at the end or something, or else nobody will do. (Participant 402)

In response to the question “What, if any, are the areas of pronunciation not treated in the homework exercises that you would have liked to work on and improve?”, a majority of the learners did not respond. Instead, if learners chose to write something at all, most learners responded that all areas were sufficiently addressed:

**Excerpt 15**
The content was very informational and perfectly formatted. (Participant 409)

Two learners, however, mentioned that they would have wished for more practice on the Umlauts:

**Excerpt 16**
More practice with ö, ä, ü. (Participant 407)

In response to the last question “Any final comments?”, six learners chose not to respond at all, and the remaining 12 learners wrote positive comments such as:

**Excerpt 17**
Overall, I am glad to have worked on my speech that I otherwise wouldn’t have. (Participant 106)
Overall, I could tell that I personally benefitted and my German speaking is better now because of it. (Participant 110)

It helped me a lot... in classroom I tend to be nervous so helped build confidence. (Participant 404)

Leave the exercises on Angel so we can go back and review words. (Participant 408)

3.2.4.2 In-class Pronunciation Training Group

Fifteen of the 16 learners in the comparison group completed the Exit Questionnaire. Table 12 provides an overview of the mean scores for the five statements on the perceived usefulness of the in-class pronunciation training. These statements had to be rated on scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”. Results showed very high values (means above 6 out of 7) for questions 2 and 3, indicating that the learners strongly agreed that they learned something from the pronunciation exercises they completed in the classroom and that they felt that their pronunciation had improved. Results showed a high value (mean above 5 out of 7) for question 1, indicating agreement with the statement that learners enjoyed working on the in-class pronunciation exercises. Results showed a medium-high value (mean above 4 out of 7) for question 5, showing that learners somewhat agreed that the exercises also helped them review some vocabulary. Finally, with a mean score of 2.1 on question 4, learners disagreed with the statement that they would have felt more comfortable practicing pronunciation as homework assignments. To the follow-up question (“If yes: Why?”) there were two responses. One learner answered that they could have taken more time to focus on the exercises if these had been assigned as homework (Participant 209) and the other learner simply stated “I’m graduating and don’t care.” (Participant 206).
Table 12: Experiment 1 – Mean Scores on Exit Questionnaire – In-class Training Group

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Statement To Be Rated By Learners</th>
<th>Mean Response (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I enjoyed working on the pronunciation exercises in class.</td>
<td>5.3 (2-7)</td>
</tr>
<tr>
<td>2.</td>
<td>I feel like I learned something from the pronunciation exercises in class.</td>
<td>6.3 (3-7)</td>
</tr>
<tr>
<td>3.</td>
<td>I feel like my pronunciation of German improved through the exercises.</td>
<td>6.1 (4-7)</td>
</tr>
<tr>
<td>4.</td>
<td>I would have felt more comfortable to practice pronunciation as homework assignments.</td>
<td>2.1 (1-3)</td>
</tr>
<tr>
<td>5.</td>
<td>Not only did the pronunciation homework improve my pronunciation but it also helped review some vocabulary.</td>
<td>4.8 (3-7)</td>
</tr>
</tbody>
</table>

Note: Statements were rated on scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”.

The second part of the debrief was comprised of three questions targeting the learners’ behavior in working through the pronunciation exercises, with a limited number of possible responses per question. For Question #1 (How much time, on average, did you spend on the pronunciation exercises in class per day?), the majority of the learners reported spending 10-15 minutes on the exercises (5-10 min: 7 learners; 10-15 min: 8 learners; 15-20 min: 0 learners; 20-25 min: 0 learners). For Question #2 (Did you enjoy the listening exercises or the speaking exercises more?), the responses were almost equally distributed: 6 learners reported enjoying the listening exercises more, 5 learners reported enjoying the speaking exercises more and 4 learners stated that they enjoyed both equally. In response to Question #3 (Do you think that 3 days of in-class homework practice on each pronunciation problem was enough to help you or would you have preferred more or less practice?), the majority of learners answered that it was a good amount (more practice: 3 learners; it was a good amount: 8 learners; less practice: 2 learners; no answer: 2 learners).
The third part of the questionnaire was comprised of five open-ended questions. In response to the question “What did you like about the pronunciation exercises?”, seven learners mentioned that they enjoyed receiving pronunciation instruction in general:

**Excerpt 21**
I thought that they did a good job in helping change our poor pronunciation and make it better. (Participant 301)

**Excerpt 22**
It was helpful to know how letters and words sound so I was able to apply that knowledge to new words. (Participant 208)

Two learners commented on the fact that the exercises helped them review vocabulary as well:

**Excerpt 23**
Nice for review of vocab as well as learning the sounds that are not present in English. (Participant 306)

All responses except one was positive, but one learner expressed their frustration with pronunciation training in general:

**Excerpt 24**
When they were over. What good is pronunciation if you don’t know the language? (Participant 206)

In response to the second question “What did you not like about the pronunciation exercises?”, four learners commented on issues regarding the length of the exercises. They did not, however, agree in their feedback. Two learners mentioned that the exercises were too long and two learners mentioned that there was not enough time for the exercises and that they should have been longer:

**Excerpt 25**
I didn’t like the length of the exercises. (Participant 205)
EXCERPT 26
*I feel like the exercises were rushed, more time could have been spent on doing the actual exercises.* (Participant 209)

Similarly, four learners commented on their attitudes towards the listening and speaking exercises, but again, they did not agree in their opinions. Some mentioned that the listening exercises were too repetitive, some found the speaking exercises too repetitive, and some specifically pointed out that they liked both equally:

EXCERPT 27
*Very repetitive listening exercises.* (Participant 306)

EXCERPT 28
*The speaking exercises were not as organized as the listening. I got less out of those – when the correct answer was revealed we didn’t review why it was correct or why it wasn’t.* (Participant 304)

In response to the question “Do you have any suggestions for future improvement of the pronunciation exercises?”, there was no clear trend in the responses. Two learners mentioned that they would have liked more one-on-one time or individualized feedback:

EXCERPT 29
*Could we submit audio files and get feedback instead of the oral part of the exams? Then we could listen to ourselves...* (Participant 304)

Two other students commented on the role of speaking versus listening exercises:

EXCERPT 30
*Add more speaking exercises to help speech as we hear the words more than we speak them.* (Participant 306)

Other comments suggested ensuring that the partner exercises were carried out more smoothly and to add more review before starting with the practice:

EXCERPT 31
*Just more forceful group interaction.* (Participant 202)
Maybe more of a review of the sounds before practicing them. (Participant 303)

In response to the question “What, if any, are the areas of pronunciation not treated in the pronunciation exercises that you would have liked to work on and improve?”, five students responded that they felt that everything had been covered:

We practiced most of the difficult pronunciations. I can’t think of any more areas not treated. (Participant 303)

Five other students named areas of pronunciation that had in fact been covered in class:

Where to put the emphasis on words (stress). (Participant 306)

Oomlauts could have been improved. (Participant 212)

One student suggested practicing on longer chunks of speech:

Practicing whole words or words with a couple of vowels next to each other. (Participant 307)

In response to the last question “Any final comments?”, eight learners chose not to respond at all and six learners wrote positive comments such as:

This was a great study that I truly believe bettered my pronunciation. (Participant 309)

I really thought this was a great idea! I think it is very important to know how to correctly pronounce words. It has been extremely helpful. (Participant 208)

One learner, however, expressed his frustration with the training and commented:

No, glad it’s over. (Participant 206)
3.3 Discussion

3.3.1 Findings

Experiment 1 investigated the effectiveness of homework-based, computer-delivered iCPR pronunciation training in a F2F learning environment by exploring two levels of comparison: (1) learners who received iCPR pronunciation training (i.e., the treatment group) were compared to learners who did not receive targeted pronunciation training (i.e., the control group), and (2) learners who received iCPR pronunciation training were compared to learners who received teacher-directed in-class pronunciation training (i.e., the comparison group). Assessment of learners’ perceptual skills was determined by their performance on two tasks, an accentedness detection task and a sound discrimination task. Data revealed that the treatment group significantly outperformed the control group on the accentedness detection task (ADT) at the posttest and that the difference in posttest scores showed a large effect. These findings suggest that the iCPR pronunciation training helped learners detect accented German. Being able to tell Standard German and accented German apart can raise learners’ awareness for common mispronunciations of German words in American English. These findings are further supported by the fact that only the treatment group improved significantly over time on the ADT, suggesting that following the standard first-semester German curriculum alone was not enough to aid learners acquire skills to identify accented German by the end of the semester. For the sound discrimination task (SDT), on the other hand, data did not reveal a significant difference between the treatment and the control group at the posttest. Moreover, results showed that all three groups (i.e., treatment, control, and comparison group) improved significantly over time. The fact that learners in the control group also made significant gains in their ability to discriminate between similar sounds in nonsense words suggests that no targeted pronunciation
training was necessary to help learners improve in this condition. Looking at the effect sizes for the gains in all three groups, however, allows for a more precise understanding of the effect of each form of treatment: we would expect learners to make significant gains in their perception skills over the course of one semester, seeing that novice learners usually start out with no knowledge of the L2 and are expected to acquire basic knowledge of the L2 over the course of the semester. More precisely, however, data for the effect of treatment over time revealed a large effect for the iCPR treatment, a small effect for the in-class treatment, and a negligible effect for treatment in the control group that did not receive targeted pronunciation training. That is, the difference between pretest and posttest mean scores in the treatment group differed by 1.56 Standard Deviations (SD), while the means in the comparison group differed by 0.83 SD, and the means in the control group only differed by 0.37 SD. This suggests that learners who received iCPR training still benefited the most, which is consistent with the descriptive findings that also showed that learners in the treatment group outperformed learners in the control group. It is possible that the design of the assessment tool for the SDT was not fine-tuned enough to allow this contrast to be reflected in measures of statistical significance only. A larger number of items on this test might have raised the statistical power and would have allowed for the descriptive trends to translate into statistical significance (see Section 3.3.3 for further discussion).

The second level of comparison—comparing learners who received iCPR pronunciation training to learners who received in-class pronunciation training—showed no significant differences between the two groups on the ADT or on the SDT, and only small and negligible effect sizes for the difference in posttest scores. This suggests that both types of treatment were comparable in their effectiveness of improving learners’ L2 pronunciation skills, which allows for the conclusion that it is not always necessary to teach pronunciation in class only, but that
assigning pronunciation training as homework can also lead to comparable gains in perception skills among novice learners (see Section 3.3.4 for further discussion).

The same levels of comparison were employed for the production task. The production task consisted of word- and paragraph-readings and data were analyzed separately for comprehensibility and accentedness. Analyses revealed the same pattern of results for both comprehensibility and accentedness on the word-level and the same pattern for both speech continua on the paragraph-level. That is, data showed that both the treatment and the comparison group significantly outperformed the control group for both speech continua on the word-level, which was further reflected in medium-sized effects for the difference in posttest scores for all four statistically significant results. These findings suggest that both the iCPR training and the in-class training helped learners to become significantly easier to understand and significantly less accented when reading out novel German words. For the improvement over time in all three groups, we see the same pattern that we saw for the results on the SDT: all three groups improved significantly from the beginning to the end of the semester on word-level productions, but again, we see differences when considering effect sizes for each type of treatment. The treatment and the comparison groups exhibited large effects of treatment for the iCPR and the in-class training over time, whereas the effect in the control group was negligible on both comprehensibility and accentedness. Therefore, even though we see a statistically significant improvement over time in all three groups—again, this is a desirable outcome in any German classroom—the difference in effect sizes suggests that the learners who received iCPR or in-class pronunciation training still benefited more from the pronunciation training than the group that just followed the standard GER 001 curriculum.
On the paragraph-level, data did not reveal that the treatment or the comparison group significantly outperformed the control group, which is further reflected in findings of only small and negligible effect sizes for these comparisons. This suggests that the benefits of pronunciation training shown on the word level did not translate to reading out a whole paragraph of German text. Moreover, comparisons over time on the paragraph-level also did not show big differences between the gains in the three groups. Again, all three groups improved significantly on measures of ease of understanding and accentedness from the beginning to the end of the semester, however, the pattern of effect sizes does not paint as clear a picture as we saw for the perception task and the word-level reading task above. On measures of comprehensibility, the iCPR training exhibited a medium-sized effect while the effect in the other two groups was only small. On measures of accentedness, the iCPR training exhibited a large effect, in-class pronunciation training exhibited a medium-sized effect, and no pronunciation training exhibited a small effect. While this trend is consistent with the descriptive findings in which both the treatment and the comparison group outperformed the control group, the contrast is not as strong as seen above for the word-level productions. Thus, overall, the findings on the production task suggest that both iCPR and in-class pronunciation training helped learners improve their speaking skills on the level of individual words significantly more than if they had just followed the standard GER 001 curriculum. Yet, while we can see a trend that the training also helped learners improve on longer productions, we cannot say that the training had a statistically significant impact in improving learners’ productions on this level as well (see Section 3.3.2 for further discussion).
3.3.2 Word- vs. Paragraph-level Productions

As outlined in Section 3.3.1, findings suggest that learners in the iCPR- and in-class groups improved significantly in productions on the word-level, but only descriptively on the paragraph-level. While the ultimate goal of pronunciation instruction is of course improvement on all levels of speech, and, thus, these findings are somewhat discouraging, they are in line with previous research (Derwing et al., 1998; J. Lee et al., 2015; Liu, 2011). Derwing et al. (1998) explain the trend that pronunciation improvement is the most likely to be seen on short productions with the idea that L2 learners can devote more time and mental resources to self-monitoring their speech productions when they are focusing on words in isolation (or in short repeated carrier phrases as was the case in the present study). That is, when learners have to produce utterances within a longer context, their mental resources are more consumed with lexical, morphological, and syntactic concerns and less mental resources are available to focus on pronunciation. Liu’s (2011) findings support these claims. In this study, the author administered pronunciation training to Chinese learners of English, and measured improvement at three discourse levels: on the word-level, on the sentence-level, and in free speech. Results revealed that pronunciation improvement was the largest on the word-level and then gradually decreased as the learners progressed to the sentence-level and free speech tasks. These individual findings are further reflected in the mean effect sizes for item length that J. Lee et al. (2015) found in their meta-analysis of 59 pronunciation studies. For the between-group contrast, they report a mean Cohen’s $d = 1.16$ for assessment on the word-level, a mean $d = 0.87$ for assessment on the sentence-level and a mean $d = 0.23$ for assessment on the discourse level. These $d$-scores suggest a decrease in effect sizes for an increase in task demands. The results of the present study are therefore in line with findings from the meta-analysis, and it is likely that
the lack of improvement on the paragraph-level productions can be explained with the increased level of difficulty and allocation of mental resources to other aspects of speech in this task. It further has to be pointed out that all training (both iCPR and in-class practice) was carried out on the word-level. Thus, learners would have had to use even more mental resources to transfer the practiced skills from the word-level to the paragraph-level—in addition to the increase in task demand that producing longer utterances already places on learners’ mental resources. Finally, another possible explanation lies in the fact that suprasegmentals and discourse factors, such as sentence stress or intonation, are more important in connected speech—such as productions on the paragraph-level—than in productions of individual words. These discourse factors, however, were not part of the present pronunciation training. Nevertheless, we would have ideally liked to see a significant, not just descriptive, increase in pronunciation skills on the paragraph-level as well. It is likely that training on such suprasegmental discourse features would have helped to further improve learners’ productions on the paragraph-level. Moreover, it is possible that additional improvement on the paragraph-level would have occurred if the pronunciation treatment had been longer. In their study investigating how the technique of shadowing can lead to improvements that are noticeable to untrained listeners’ perception, Foote and McDonough (2017) found that while their learners improved from pre- to posttest, there was not always improvement from one testing time to the next for the tests in between the pre- and the posttest. They explain this by suggesting that there may be a minimal threshold for noticeable improvement to be identified by phonetically untrained listeners, meaning that it is possible that learners show improvements, but that the improvements in between testing times were too small to be detected by the raters or to be reflected on a Likert scale with its limited sensitivity. Since the learners’ productions in the present study were rated by phonetically untrained listeners as
well, it is possible that these raters also did not detect the improvements because the threshold for noticeable improvement was not reached or could not be reflected on a 9-point Likert scale, but that there was still improvement that could turn into noticeable improvement over a longer period of time. This assumption is further supported by the finding that, descriptively, the iCPR and in-class learners already outperformed the control group at the time of the posttest, even though the differences did not reach statistical significance.

Interestingly, the pattern of findings for the word- and paragraph-level productions were consistent across the rated continua of comprehensibility and accentedness. This, too, is in line with previous research that has found strong correlations between ease of understanding and strength of accent (Roccamo, 2014; Saito, Webb, et al., 2016; but see Derwing et al., 2014). Saito, Webb, et al. (2016) investigated the extent to which aspects of L2 speech—such as fluency, variation, sophistication, or abstractness—interact to influence listeners’ judgments of the two speech continua. While they were able to determine which aspects were primarily linked to comprehensibility (e.g., lexical accuracy, complexity) and accentedness (e.g., lexical content, form), they conclude that both dimensions are strongly interrelated. Therefore, it comes as no surprise that the findings for comprehensibility and accentedness in the present study show the same pattern of statistical significance and effect sizes across both speech continua. Given the importance for learners to improve both comprehensibility and accentedness (see discussion in Section 2.2.1), it is encouraging to see that both iCPR and in-class pronunciation training helped learners improve similarly on both dimensions.

Finally, it has to be pointed out that no analyses were performed to investigate whether the improvements on word-level productions in the iCPR and in-class group stemmed from improvements on the segmental level, on the suprasegmental level, or both. The present study
was not designed to investigate this comparison and the number of rated tokens for each individual feature was not large enough to address this question. Yet, the overall improvement in perception and production skills in both groups that received pronunciation training supports previous findings that both segmentals and suprasegmentals should be targeted when designing a pronunciation intervention (Derwing, 2013; J. Lee et al., 2015; Saito, 2012; Thomson & Derwing, 2015).

### 3.3.3 Perception vs. Production

As outlined above for the perception task, data for the accentedness detection task (ADT) revealed that only the treatment group significantly outperformed the control group at the time of the posttest, whereas the comparison group did not. Data for the sound discrimination task (SDT) showed that while the treatment and the comparison group descriptively outperformed the control group at the time of the posttest, none of these comparisons reached statistical significance. Yet, all three groups improved significantly from pretest to posttest. It therefore appears as if receiving pronunciation training was not necessary for learners to improve their abilities to discriminate between similar German sounds in nonsense words. However, seeing the descriptive trends in the between-group comparison paired with the effect sizes for gains over time in the within-group comparison in each group (a large effect in the treatment group, a small effect in the comparison group, and a negligible effect in the control group), it is possible that the number of test items simply did not suffice to reach significance for measuring the differences in gains among the three groups. Another possibility is that there was no significant difference between the treatment and the control group because the treatment group actually did not complete the listening component of their training. This becomes apparent when looking at the
results from the exit questionnaire. As outlined above, learners in the iCPR group were asked how many of the 10 listening units they completed. This question was included in the questionnaire because—as opposed to the production units—there was no control mechanism for whether learners completed the perception units or not. After every production unit, learners had to upload their own recorded speech, but no such control mechanism for the instructor to check on homework completion was built into the perception exercises. In the exit questionnaires, 12 of 18 learners self-reported having completed less than five of the listening units. These numbers suggest that only 33% of the learners completed at least half of the listening exercises, which is a very low number to expect learners to exhibit gains from the treatment. Yet, it has to be pointed out that despite the low rate of completion of the listening exercises, the iCPR learners still improved significantly on the ADT and outperformed learners in the control group. Moreover, post-hoc Pearson’s $r$ correlations did not reveal a significant correlation between the number of listening units each learner completed and the learners’ gain scores on the perception ($r = -0.069, p = 0.799, N = 16$) or on the production task ($r = 0.011, p = 0.966, N = 16$). This suggests that the lack of significant differences between the treatment and control group in the SDT as well as the significant difference in the same comparison in the ADT might not primarily stem from the listening training at all, but rather from the production training or most likely from a combination of both types of training. This line of reasoning then raises the following question: if the pattern of results found for the perception task did not primarily stem from the perception training, why did the comparison group not significantly outperform the control group on the ADT, even though the treatment group did? It seems that the answer lies in the differences in training modalities between the iCPR and the in-class group. The training in both groups differed in the

---

14 Pearson’s $r$ correlations also did not reveal a significant correlation between reported time learners spent on each iCPR unit and their pre-to posttest production gains ($r = -0.018, p = 0.941, N = 18$).
fact that even beyond the perception training—which was almost identical for both groups—the iCPR learners received input from multiple speakers in the production training as well. That is, the practice items in the production units that learners imitated were recorded by six different native speakers. Thus, learners in the iCPR group were continuously exposed to input from multiple voices whereas learners in the in-class group—beyond the perception training—were only exposed to the voice of their instructor. Previous studies on high variability phonetic training (HVPT) have found that training based on input from multiple speakers and productions of sounds in varying phonetic contexts results in greater improvement than training that simply relies on a single talker or single phonetic context. Therefore, it is likely that listening to the recordings of multiple speakers in the production training gave learners in the iCPR group the additional exposure that helped improve their perception skills. This could explain why the treatment group outperformed the control group on the ADT while the comparison group did not, and it is in line with the descriptively larger gains in the treatment group on the SDT. Furthermore, it could explain why the in-class group did not significantly outperform the control group on the ADT task even though—as opposed to the iCPR group—they completed all perception exercises unless they were absent that day in class: apparently, one day a week of HVPT-based listening training alone was not sufficient to lead to significant gains on the perception task.

Interestingly, the differences in patterns of results between the iCPR and the in-class group on the perception task did not translate to differences in results on the production task, where patterns of significance as well as effect sizes were consistent between the iCPR and in-class group. In part, this challenges previous findings about the association between speech production and perception whereby L2 learners’ perception accuracy serves as a predictor of
their production accuracy, because not only the treatment but also the comparison group—which showed smaller gains on perception accuracy—significantly outperformed the control group on measures of production skills. Yet, it has to be considered that previous studies on the perception-production link and studies that showed success of HVPT in helping learners improve their perception and production skills were usually not carried out in a classroom setting, but have been tested in laboratory settings (Lambacher et al., 2005; Thomson, 2011; Y. Wang et al., 2003). Conditions in classroom settings vary from conditions in laboratory settings because they usually entail a more diverse group of learners, various combinations of explicit instruction, more authentic engagement with learning tasks, and opportunities to reflect on learning outcomes and to compare these outcomes to peers. Therefore, additional classroom-based research on HVPT is necessary to explore the relationship between perception and production, and the developmental path to gains in perception and production accuracy among novice L2 learners.

3.3.4 Homework-based vs. In-class Pronunciation Training

Findings showed no significant difference between the posttest scores for any of the perception and productions tasks between the iCPR and the in-class group, and only negligible and one small effect size for the comparison of posttest scores between these two groups. This suggests that the homework-based, computer-delivered iCPR training was as effective in improving learners’ perception and production skills as teacher-delivered, in-class pronunciation training. This is an important empirical finding which reveals that the outcome of pronunciation instruction is similar between the two methods. Before concluding that they are equally valuable, however, it is important to also evaluate the learners’ feedback on both methods of pronunciation
instruction. Learners’ feedback and suggestions for each method was elicited through the exit questionnaires administered in both groups at the end of the training. None of the learners experienced both types of pronunciation instruction since they were either in the iCPR or in the in-class group, so a direct comparison is not possible. Yet, learners’ responses to questions targeting the perceived usefulness of each type of training can be compared. In the first part of the questionnaire, learners had to express the extent to which they agreed with various statements on a scale from 1 to 7, with 1 indicating “I strongly disagree” and 7 indicating “I strongly agree”. The two prompts stating that learners felt that they had learned something through their respective pronunciation training and that they thought their pronunciation had improved received very close scores in both groups (“learned something”: iCPR: 6.2, in-class: 6.3; “pronunciation improved”: iCPR: 6.4, in-class: 6.1). The fact that these scores were not only very close but also generally very high—all above 6 out of 7—shows that pronunciation training in general had a high degree of face validity for students in both groups. The statement assessing whether learners enjoyed working on the pronunciation training was rated at a 4.5 in the iCPR group and at a 5.3 in the in-class group, suggesting that the in-class group enjoyed the pronunciation training a little more than the iCPR group. When asked if each group would have preferred the other groups’ type of training (i.e., in-class training for the iCPR learners and homework-based training for the in-class learners), learners disagreed and showed a rating of 2.8 in the iCPR group and 2.1 in the in-class group. The low scores that the statements in this question received indicate that neither group was dissatisfied with their method of pronunciation instruction and did not clearly wish for another type of pronunciation training, again suggesting that both types of pronunciation training were of similar face validity to the learners.
In the second part of the questionnaire, which assessed habits in completing the pronunciation training, a majority of learners in both groups reported spending 10-15 minutes on each pronunciation intervention. This corroborates the fact that learners in both groups received the same amount of time on task for pronunciation instruction (as a reminder, the control group received additional grammar and vocabulary exercises to ensure that all groups received the same time on task for German instruction throughout the semester). The remaining questions, however, do show differences between the two groups. First of all, 12 of the 18 learners in the iCPR group report that they completed less than half of the perception units. This was not assessed as a question in the questionnaire for the in-class learners, but unless learners missed a day of class, they participated in all perception training units (see Section 3.3.3 for possible implications of this finding). Interestingly, this seems to have carried over to the learners’ perspective of the perception exercises: while 15 of 18 learners in the iCPR group reported that they liked the speaking exercises better than the listening exercises, in the in-class pronunciation group, the preferences were split more evenly (listening: 6 learners, speaking: 5 learners, liked both equally: 4 learners). This is somewhat surprising seeing that the listening training was almost identical in both groups: in-class learners received the same recordings in a Powerpoint format as the iCPR learners. The only difference between the groups was that, in the in-class training, the recordings were played in front of the entire class by the instructor and learners had to raise either a green or a blue card to indicate their answers whereas the iCPR learners simply had to click to the next slide to see the answer. It is possible that the learners enjoyed the colored-card-based answer method so much that their perception of the listening task was more positive. A more likely explanation, however, might be found in the fact that, for the listening exercises in the iCPR group, there was no control mechanism of whether the learners completed
the exercise or not. In their open-ended responses in the third part of the questionnaire, several learners remarked that they did not complete the listening exercises because there was no control mechanism of whether the learners completed them or not. In fact, the learners even suggested adding a quiz at the end of the listening exercises to force students to take them more seriously. This suggests that learners might have perceived the listening exercises as superfluous and chose not to spend their time on them since these exercises did not affect their grade. In the in-class condition, on the other hand, the listening exercises were part of the standard lesson for which learners had to be present anyways. It is then possible that the learners in the in-class group approached the listening exercises with a more open mind and did not perceive them as superfluous.

Overall, with the exception of commitment to the perception exercises, it is apparent that both types of pronunciation training were not only equally successful in improving learners’ L2 pronunciation, but also similarly well received by the learners. Taken together with the finding that the learners in the iCPR group seemingly benefited from the multiple voices in the native speaker sample recordings throughout the production exercises as illustrated above (see Section 3.3.3), this suggests that both types of pronunciation instruction can be used to improve classroom learners’ pronunciation instruction. Seeing that several studies have shown that a majority of teachers did not receive adequate training to teach pronunciation in the classroom, and considering the tight time-constraints on L2 curricula, being able to assign pronunciation as homework rather than having to teach it in class likely comes as a relief to foreign language teachers (Derwing & Munro, 2005; Grim & Sturm, 2016; Henderson et al., 2012; Foote et al., 2011). (For a detailed discussion of pedagogical implications see Section 5.3).
3.3.5 iCPRs and CAPT

As outlined in Section 2.2.5 in the review of the literature, Computer Assisted Pronunciation Training (CAPT) has many advantages, but also some disadvantages. The method of iCPRs falls under the general category of CAPT since all pronunciation training is delivered through the use of Microsoft Powerpoint presentations. One major advantage that is often named in favor of CAPT is the absence of time restrictions, which also showed to be a major advantage in the use of iCPRs. Not only did it save valuable in-class time to be able to assign pronunciation training as homework, but learners were also more flexible in the amount of time they dedicated to improving their pronunciation of individual sounds or suprasegmental features. If they struggled with any particular sound or felt that they needed more training on that feature, they were free to keep practicing with the iCPR materials. In the classroom-based pronunciation instruction, on the other hand, all pronunciation exercises were paced for the whole group of learners and it was not possible for an individual learner to decide to spend more time on one sound, but less on another. The fact that learners valued the flexible approach in time management and access to training is illustrated by comments in the exit questionnaire such as “Leave the exercises on Angel so we can go back and review words” (Participant 408).

One major problem of CAPT generally is the unreliability of feedback, which often includes false automatic feedback, for instance in automatic speech recognition programs. This problem is not an issue in the method of iCPRs: in the perception units, the feedback provided simply stated which of the two choices constituted the correct answer. In the production units, no feedback was provided at all. Generally, the method of iCPRs would absolutely allow individual feedback if the instructor chose to provide feedback on the recordings that each learner submitted after working through each production unit. However, even this feedback would not exhibit the
traditional problems of CAPT since the feedback would be provided by the instructor, a real person, rather than a computer.

Furthermore, learners agreeing with the statement “Any technological problems that I encountered were worth the overall benefit” on the exit questionnaire (rating of 5.2 of 7) shows that technological obstacles did not discourage learners from using this type of CAPT. In fact, learners’ feedback on the advantages of CAPT was very positive as comments such as “It helped me a lot… in classroom I tend to be nervous so helped build confidence” (Participant 404) and “Helped to hear what was supposed to sound like; there was less pressure than in class” (Participant 409) show.

Importantly, the method of iCPR revealed much larger effect sizes for the treatment than were previously found for the use of computers in pronunciation training. J. Lee et al.’s (2015) meta-analysis found an average Cohen’s $d = 0.75$ for CAPT for within-group comparisons and $d = 0.24$ for between-group comparisons. With an effect size of $d = 1.47$ for the within-group comparison in the iCPR group and an effect size of $d = 0.76$ for the comparison between the treatment and control group at the posttest, it is apparent that the iCPR training exhibited larger effects than are generally found in the field of CAPT. This is an encouraging finding which further highlights the overall benefit of the iCPR method.

The finding that learners benefited from iCPRs and enjoyed and valued this type of CAPT is in line with results from most recent CAPT studies. For instance, Bajorek’s (2017) comparison of the latest programs in commercially available CAPT training found that there are programs that have overcome the most common problems that CAPT exhibited in the past, such as Babbel. Moreover, Lima’s (2015), Supra Tutor—a four-week fully online pronunciation tutor designed to help international teaching assistants improve their comprehensibility through
targeted suprasegmental exercises—showed to be effective in significantly improving international teaching assistants’ overall comprehensibility. Furthermore, Okuno and Hardison’s (2016) computer-assisted auditory-visual training using waveform displays led to significant improvement in vowel duration among learners of L2 Japanese. Finally, Becker and Sturm’s (2017) computer-assisted online audio-visual materials helped L2 French learners’ listening comprehension development. Overall, this suggests a positive trend for CAPT and shows that iCPR training is in line with the direction of the field of pronunciation training.

3.3.6 (Lack of) Feedback

As discussed in Section 2.4.1.2, the pronunciation training in the iCPR and in-class group differed in terms of learners’ access to feedback. As a reminder, the following instances of feedback were provided in each type of treatment: in the iCPR group, learners received automated feedback in all perception units: the solution to each binary choice task was provided on the following slide, and learners were encouraged to listen to the correct form once more. Learners received no feedback on any of their submitted recordings for the production tasks. In the in-class group, learners received automated feedback in all perception units: after indicating their choice of correct answer by use of a color-coded card, the solution was displayed in front of the entire class by the instructor, and the instructor played the correct form once more. During the in-class production exercises, learners were encouraged to provide peer-feedback on their partner’s pronunciation and the instructor walked around the classroom, providing corrective feedback if she felt that it was appropriate. The quantity and quality of the peer- and instructor feedback during these exercises, however, was not assessed.
Previous findings on the role of corrective feedback in pronunciation instruction are not entirely clear. Saito (2015) showed that corrective feedback (here in form-focused instruction) was not necessary to improve Japanese L2 learners’ skills in the perception and production of the /l/-/ɹ/ contrast among lower proficiency learners. A. Lee and Lyster (2016a, 2016b), on the other hand, found corrective feedback to be beneficial in improving Korean learners’ perception skills, but their findings only applied to trained words and no significant benefit for corrective feedback was found on untrained words. In a follow-up study, A. Lee and Lyster (2017) found that some types of corrective feedback (i.e., repeating the targeted form, which is also the type of feedback iCPR and in-class learners received in the perception units in the present study) on perception can even lead to improvement in production. Yet, none of these studies were carried out in intact classrooms (Saito, 2015 and A. Lee and Lyster, 2016b employed simulated classrooms, but participants were recruited specifically for this task and class sizes were very small), therefore, the applicability to real-life classrooms still has to be investigated in future research.

It is possible that the learners in the iCPR group would have benefited even more if they had received feedback on their productions, which was reflected in learners’ comments on the exit questionnaire and on their 4.9/7 average agreement with the statement “I would have taken the pronunciation homework more seriously and would have benefited more from it if I had received teacher feedback throughout the semester”. Implementing such feedback could have taken various forms: phonetically trained teachers could have provided detailed feedback and given explicit metalinguistic instructions on how to improve the remaining pronunciation problems. But even teachers who are not trained to teach pronunciation could have provided feedback, for instance in the form of a comprehensibility rating throughout the semester, similarly to what the untrained raters in the present study did (even an untrained instructor would
be able to tell whether a learner improved from one recording to the next) . They could also have used materials from the listening units in class to administer a quick accentedness detection quiz. And even very busy instructors could have provided feedback on some of the submitted recordings since even a spot check or simply encouraging the learner to continue the good work might make a difference from the learners’ perspective, as it shows that the instructor is paying attention. Yet, the results of the present study also showed that feedback was not necessary to see improvement in the iCPR group: learners in the iCPR group significantly outperformed learners in the control group on perception and production skills, and there were no significant differences between iCPR and in-class learners’ pronunciation improvement. Moreover, the iCPR training showed consistently higher effect sizes than are being reported in J. Lee et al.’s (2015) meta-analysis for the use of pronunciation training without feedback: within-group comparison \( d_{\text{no feedback}} = 0.89 \), \( d_{\text{iCPR}} = 1.47 \) and between-group comparison \( d_{\text{no feedback}} = 0.62 \), \( d_{\text{iCPR}} = 0.73 \). Therefore, while additional feedback might have very well translated into even greater gains, the convenience for instructors to be able to assign pronunciation instruction as homework without having to do any additional grading or corrections has to be weighed against the additional benefits and work that providing feedback creates. The role of feedback in iCPRs should be addressed in future research (see discussion in Section 5.4), but in the end, it might just depend on teaching context and personal preference as well as skills of each individual instructor to determine how iCPRs should be included in the curriculum, with or without feedback. Importantly, instructors can be reassured that their learners benefit from iCPR pronunciation training even without corrective feedback.
CHAPTER 4: Experiment 2 – Distance Learning Environment

4.1 Methodology

4.1.1 Participants

A total of 67 learners from four sections of first-semester online German classes offered through Penn State World Campus were eligible to participate in the distance learning environment part of this study. Fifty of these learners gave consent to participate. Of these 50 learners, 16 had to be excluded for one of the following reasons\(^\text{15}\): a) they were not L1 speakers of English \((n = 2)\), b) they did not turn in 20% or more of the iCPR homeworks \((n = 2)\), c) they dropped the class before the final round of data collection \((n = 3)\), d) they did not submit any parts of the pre- or posttest data \((n = 6)\), or e) the quality of their spoken recordings was so poor that their data could not be rated by the native speaker raters \((n = 3)\). After these exclusions, there were 34 participants who met all of the criteria for inclusion in the study. However, only a subset of 29 of these participants had completed both pre- and posttest for perception and production. To avoid additional participant loss, learners were therefore included in the analyses if they had completed at least one set of the pre- and posttest: either the perception or the production tests. That is, learners were included in the perception analyses if they had completed the perception pre- and posttest, even if the production data collection for the learner was incomplete, and learners were included in the production analyses as long as they had completed the production pre- and posttest. This led to a total of 30 participants in the perception analyses and 33 participants in the production analyses. There was one learner that was included in the perception

\(^{15}\) Four learners in the distance learning environment reported that they had lived in Germany on military bases for several years. This situation is very common among online learners taking German classes through World Campus and this type of learners repeatedly reports having had none or very limited contact with Germans or the German speaking environment off base during their time in the country. Therefore, these learners were not excluded from the analyses.
analyses but not in the production analyses and there were four learners that were included in the
production analyses but not in the perception analyses.

Two sections each were assigned to the treatment group \((N = 20; n_{\text{perception}} = 17; n_{\text{production}} = 20)\) and the control group \((N = 14; n_{\text{perception}} = 13; n_{\text{production}} = 13)\). There was no comparison
group in the distance learning environment because the online learners did not have in-class
instruction. Table 13 shows learners’ average age and self-rated proficiency for German reading,
writing, speaking, and listening on a scale form 1 (very poor) to 10 (very proficient) by group.
Independent sample \(t\)-tests (equal variances not assumed) revealed no significant difference in
any of these measures between the two groups \((G_{\text{speaking}}: t(32) = 1.08, p = .288; G_{\text{listening}}: t(32) =
1.90, p = .067; \text{all other } t < 1)\).

Table 13: Experiment 2 – Learners’ Age and German Proficiency Ratings

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 20)</td>
<td>(n = 14)</td>
</tr>
<tr>
<td></td>
<td>((f = 6; m = 14))</td>
<td>((f = 4; m = 10))</td>
</tr>
<tr>
<td><strong>M (SD)</strong></td>
<td><strong>Range</strong></td>
<td><strong>M (SD)</strong></td>
</tr>
<tr>
<td>Age (years)</td>
<td>28.9 (7.4)</td>
<td>18-44</td>
</tr>
<tr>
<td>Self-ratings (max. 10)</td>
<td>1.8 (1.0)</td>
<td>1-5</td>
</tr>
<tr>
<td>Reading</td>
<td>1.6 (.89)</td>
<td>1-4</td>
</tr>
<tr>
<td>Writing</td>
<td>1.5 (.67)</td>
<td>1-3</td>
</tr>
<tr>
<td>Speaking</td>
<td>1.7 (.80)</td>
<td>1-3</td>
</tr>
</tbody>
</table>

Furthermore, two German instructors participated in this part of the study. A native
speaker of German with two years of teaching experience taught one section of the treatment
group and one section of the control group. A native speaker of English with native-like German
proficiency, who had lived five years in Austria and had six years of German teaching
experience, taught the other section of the treatment and the control group. The author of this dissertation did not act as an instructor in this study.

Finally, four native speakers of German (2 female, 2 male) were recruited as raters for Experiment 2 by means of a mass email sent through the international office at Penn State. In line with previous L2 pronunciation research (Oh, Jun, Knightly, & Au, 2003; Saito, Webb, et al., 2016; Yeni-Komshian, Flege, & Liu, 2000), these native speakers ($M_{age} = 27$) had lived in the US for no more than five years ($M_{length\_of\_stay} = 15$ months), had never taught German, and had no hearing impairment. All raters were Master’s, PhD, or exchange students at Penn State majoring in architecture, rural sociology, demography, and sports sciences. All raters grew up in Germany with German as their first language. They self-reported their German reading, speaking, writing, and listening proficiency at either 9 or 10 out of 10. Additional information on inter-rater reliability will be provided in Section 4.1.4.2.1.

4.1.2 Materials

All instructional and testing materials were identical to the materials used in the treatment and control group in Experiment 1 (see Section 3.1.2).

4.1.3 Procedure

4.1.3.1 Overview of Experimental Procedure

A total of four sections of GER 001 participated in this part of the study. These sections were divided into two sections which served as a treatment group and received pronunciation instruction via iCPR units, and two sections which served as a control group that did not receive a particular pronunciation training, but followed the same GER 001 curriculum.
Both groups received the same amount of total German instruction: while the treatment group was assigned three iCPR units each week, the control group was assigned additional grammar or vocabulary practice.

Data collection took place in Fall 2016. Pretest data collection took place in Week 2 of the semester, pronunciation treatments were assigned in Weeks 3-13 with one week pause for midterm exams, and posttest data collection took place in Week 14 of the semester.

All learners who were registered for one of the four sections participating in this study had to partake in all parts of this study as relevant for their respective groups. However, their data were only used if they gave consent to participate.

4.1.3.2 Testing Procedure

The pre- and posttest procedure was identical for the treatment and control group in the distance learning environment. During Week 2 of the semester, the instructors sent out an email to their online students on behalf of the author of this dissertation. In this email, the study was introduced and the learners were asked to give or withhold consent to participate. Consent forms were electronically signed by the participants by typing their initials and were uploaded to a private folder on Canvas, so that the instructors were not able to see who chose to participate or not. All students were then asked to download the language background questionnaire, fill it out, and upload their responses to another private dropbox on Canvas. Moreover, as part of their assigned exercises for Week 2, all students had to download the perception and production pretest and complete it. For the perception pretest, participants downloaded the PowerPoint file containing the recordings as well as the perception handout as a Word Document that they filled out during the testing, and uploaded it to a dropbox on Canvas. While it was possible to control
the number of times a sound file was played during the perception test in the face-to-face environment, in the distance learning environment, it was not possible to control how often the participants played each sound file before giving an answer.

For the production pretest, participants followed the exact same steps as discussed in Section 3.1.4.2 for the face-to-face learning environment. The three blocks in the production test were again counter-balanced between instructors and sections so that one of two sections within one group received the order 123 and the other section received the order 312.

In Week 14 of the semester, the same procedure was followed to administer the perception and production posttests: they were assigned as part of the students’ exercises for that week. The perception posttest contained the same items as the pretest, but they appeared in a new randomized order in both tasks. The production posttest was identical to the pretest, but each section that received the blocks in the order 123 at the time of the pretest, now received them in the order 312 for the posttest. After all posttest data were collected, the author of this dissertation sent out an email to all learners to explain the goal and purpose of the study, ask for any remaining questions, and thank the students for their participation.

4.1.3.3 Pronunciation Training Procedure

Learners in the treatment group received pronunciation training in the form of iCPR units. These iCPR units were identical to the units used in the face-to-face environment and were assigned in the same order as outlined in Chapter 3 (see Section 3.1.4.3).
4.1.4 Scoring and Data Analysis

4.1.4.1 Perception

The scoring procedure for the perception task was identical to that outlined for Experiment 1 (see Section 3.1.5.1).

As not all of my data were normally distributed (according to Shapiro-Wilks tests as recommended by Larson-Hall, 2016: accentedness detection task treatment group dataset \( p = .001 \)) and variances were not equal, I present the results of nonparametric tests and provide standardized estimates of the effect (Cohen’s \( d \)), and an estimate of the error associated with the comparison (95% confidence intervals) whenever possible (Norris, Plonsky, Ross & Schoonen, 2015). All analyses were carried out separately for exercises 1 and 2. Since my principal interest was to see whether the type of treatment would lead to differences in perception skills between the two groups at the time of the posttest, my primary analyses target between-group differences. I first compared pretest scores between the treatment and control group using Mann-Whitney tests to confirm that there were no significant differences between the groups prior to the treatment. I then compared posttest scores in the same manner. Since I was also interested in gains over time for each group, I finally compared pretest and posttest scores for both groups using Wilcoxon signed-rank tests.

As recommended by Plonsky and Oswald (2014), Cohen’s \( d \) field-specific benchmarks were used for interpretation and effect sizes for between-group comparisons (e.g., treatment group vs. control group at posttest) were considered large when \( d = 1.00 \), medium when \( d = 0.70 \), and small when \( d = 0.40 \). Effect sizes for within-group comparisons (e.g. treatment group at pretest vs. at posttest) were considered large when \( d = 1.40 \), medium when \( d = 1.00 \), and small

---

16 For parity with other studies, I carried out parametric tests (independent sample \( t \)-tests and paired \( t \)-tests) and found the same patterns of findings as for nonparametric tests.
when \( d = 0.60 \). Since the effect sizes between the three groups at the pretest were very similar and therefore there were no baseline differences that had to be accounted for, no adjusted effect sizes will be reported here.

4.1.4.2 Production

4.1.4.2.1 Rating Procedure

Procedure

The materials to be rated (i.e., the individual sound files) were obtained in the same manner as outlined for Experiment 1 (see Section 3.1.5.2.1). The rating procedure was the same as outlined for Experiment 1 (see Section 3.1.5.2.1) with the following adjustments for participant numbers: four native-speaker raters were recruited to rate the learners’ productions (for details on the raters see Section 4.1.1). Raters were paid an hourly rate for their participation. Each rater rated data from all 33 learners plus all data from four native and four near-native speakers as a control measure. Thus, each rater rated a total of 1776 word productions and 222 paragraph productions for a total of approximately 420 minutes or seven hours of rating per rater. The 1776 individual words and 222 paragraph excerpts that were presented to each rater were randomly distributed to 16 blocks of word ratings and ten blocks of paragraph ratings.

Intraclass correlation coefficients were used to calculate the inter-rater reliability of the raters. As shown in Table 14, high inter-rater consistency was observed for all raters on both continua of ratings (Cronbach’s \( \alpha = .89-.92 \)), in all cases exceeding the recommended .70-.80 benchmark (Larson-Hall, 2016). Reliability values were thus deemed sufficiently high to calculate mean comprehensibility and accentedness scores for each speaker, averaging across all raters, separately for each speech continuum. The average values for each learner then
constituted the data points for the production analyses for comprehensibility and accentedness, presented in the following section.

Table 14: Experiment 2 – Rater Consistency (Cronbach’s α, [95% CI]) for Rated Continua by Task

<table>
<thead>
<tr>
<th></th>
<th>Rater Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cronbach’s α, [95% CI]</td>
</tr>
<tr>
<td>Words</td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.90 [.88, .92]</td>
</tr>
<tr>
<td>Accentedness</td>
<td>.91 [.89, .93]</td>
</tr>
<tr>
<td>Paragraphs</td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.92 [.91, .93]</td>
</tr>
<tr>
<td>Accentedness</td>
<td>.89 [.88, .90]</td>
</tr>
</tbody>
</table>

4.1.5.2.2 Statistical Analyses

Standard practice before analyzing rating data is to exclude learners who performed at ceiling at the time of the pretest. Ceiling performance for this task was set at 8 with 9 being the highest possible rating. No learner received an average comprehensibility or accentedness score higher than 7.42, therefore, no data was excluded. Moreover, it was confirmed that all raters recognized native speakers and near-native speakers as such, and used the entire rating scale.

The production data in the distance learning environment were normally distributed (according to Shapiro-Wilks tests as recommended by Larson-Hall, 2016: all $p > .295$), but since I present nonparametric tests for all other analyses in this dissertation, I present the results of
nonparametric tests here as well to be consistent. I further provide standardized estimates of the effect (Cohen’s $d$) and an estimate of the error associated with the comparison (95% confidence intervals) whenever possible (Norris, Plonsky, Ross & Schoonen, 2015). All analyses were carried out separately for measures of comprehensibility and accentedness, as well as for the word- and paragraph-level productions. Since my principal interest was to see whether the type of treatment would lead to differences in comprehensibility and in accentedness between the two groups at the time of the posttest, my primary analyses target between-group differences. I first compared pretest scores between the treatment and control group using Mann-Whitney tests to confirm that there were no significant differences between the groups prior to the treatment. I then compared posttest scores in the same manner. Since I was also interested in gains over time for each group, I finally compared pretest and posttest scores for both groups using Wilcoxon signed-rank tests.

As recommended by Plonsky and Oswald (2014), Cohen’s $d$ field-specific benchmarks were used for interpretation and effect sizes for between-group comparisons (e.g., treatment group vs. control group at posttest) were considered large when $d = 1.00$, medium when $d = 0.70$, and small when $d = 0.40$. Effect sizes for within-group comparisons (e.g. treatment group at pretest vs. at posttest) were considered large when $d = 1.40$, medium when $d = 1.00$, and small when $d = 0.60$. Since the effect sizes between the three groups at the pretest were very similar and therefore there were no baseline differences that had to be accounted for, no adjusted effect sizes will be reported here.

---

17 For parity with other studies, I carried out parametric tests (independent sample $t$-tests and paired $t$-tests) and found the same patterns of findings as for nonparametric tests.
4.2 Results

4.2.1 Perception

4.2.1.1 Accentedness Detection Task

For a visual representation of the results, see Figure 16 and Figure 17. For descriptive statistics beyond the visual representation, see Table 4. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest ($U = 86.00, z = -1.04, p = .320, d = 0.02$), but they revealed a significant difference at the time of the posttest ($U = 58.00, z = -2.24, p = .028, d = 0.94$). With a between-group effect size of $d = 0.94$, the difference in posttest scores between the treatment group and the control group approached the benchmark for a large effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -2.34; p = .019, d = 0.89$), but not for the control group ($z = -0.87, p = .385, d = 0.33$). With a within-group effect size of $d = 0.89$, the gains over time in the treatment group approached the benchmark for a medium-sized effect. The gains in the control group ($d = 0.33$) would be considered a negligible effect.
Figure 16. Experiment 2 – Accentedness Detection Task – Between-group Results

![Graph showing between-group results with Pre and Post data points for Treatment and Control groups.](image)

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.4); # = small effect size (0.4 ≤ d < 0.7); ## = medium effect size (0.7 ≤ d < 1.0); ### = large effect size (d ≥ 1.0)

Figure 17. Experiment 2 – Accentedness Detection Task – Within-group Results

![Graph showing within-group results with Pre and Post data points for Treatment and Control groups.](image)

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.6); # = small effect size (0.6 ≤ d < 1.0); ## = medium effect size (1.0 ≤ d < 1.4); ### = large effect size (d ≥ 1.4)
Table 15: Experiment 2 – Descriptive Statistics for Accentedness Detection Task

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD) 95% CI Mdn  IQR</td>
<td>M (SD) 95% CI Mdn  IQR</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 17)</td>
<td>80.39 (13.32) [73.54, 87.24] 83.33 8.3</td>
<td>89.95 (7.37) [86.16, 93.74] 87.50 10.4</td>
</tr>
<tr>
<td>Control</td>
<td>80.13 (7.45) [75.63, 84.63] 79.17 10.4</td>
<td>82.69 (8.13) [77.78, 83.33] 87.61 12.5</td>
</tr>
</tbody>
</table>

4.2.1.2 Sound Discrimination Task

For a visual representation of the results, see Figure 18 and Figure 19. For descriptive statistics beyond the visual representation, see Table 16. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest (U = 101.50, z = -0.38, p = .711, d = 0.07), but they revealed a significant difference at the time of the posttest (U = 58.00, z = -2.23, p = .028, d = 1.00). With a between-group effect size of d = 1.00, the difference in posttest scores between the treatment group and the control group would be considered a large effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group (z = -2.27; p = .023, d = 1.03), but not for the control group (z = -1.02, p = .309, d = 0.20). With a within-group effect size of d = 1.03, the gains over time in the treatment group would be considered a medium-sized effect. The gains in the control group (d = 0.20) would be considered a negligible effect.
**Figure 18.** Experiment 2 – Sound Discrimination Task – Between-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.4$); # = small effect size ($0.4 \leq d < 0.7$); ## = medium effect size ($0.7 \leq d < 1.0$); ### = large effect size ($d \geq 1.0$)

**Figure 19.** Experiment 2 – Sound Discrimination Task – Within-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.6$); # = small effect size ($0.6 \leq d < 1.0$); ## = medium effect size ($1.0 \leq d < 1.4$); ### = large effect size ($d \geq 1.4$)
Table 16: Experiment 2 – Descriptive Statistics for Sound Discrimination Task

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>Mdn</td>
</tr>
<tr>
<td>Treatment</td>
<td>70.88 (9.13)</td>
<td>70.00</td>
</tr>
<tr>
<td>(n = 17)</td>
<td>[62.90, 78.87]</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>70.00 (8.17)</td>
<td>70.00</td>
</tr>
<tr>
<td>(n = 13)</td>
<td>[65.07, 74.93]</td>
<td></td>
</tr>
</tbody>
</table>

4.2.2 Production

4.2.2.1 Comprehensibility

4.2.2.1.1 Word-level Productions

For a visual representation of the results, see Figure 20 and Figure 21. For descriptive statistics beyond the visual representation, see Table 17. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest ($U = 127.00, z = -0.11, p = .928, d = 0.04$), but they revealed a significant difference at the time of the posttest ($U = 73.50, z = -2.08, p = .036, d = 0.80$). With a between-group effect size of $d = 0.80$, the difference in posttest scores between the treatment group and the control group would be considered a medium-sized effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -3.70; p < .001, d = 1.02$), but not for the control group ($z = -1.43, p = .152, d = 0.23$). With a within-group effect size of $d = 1.02$, the gains over time in the treatment group would be considered a medium-sized effect. The gains in the control group ($d = 0.20$) would be considered a negligible effect.
Figure 20. Experiment 2 – Comprehensibility on Word-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.4$); # = small effect size ($0.4 \leq d < 0.7$); ## = medium effect size ($0.7 \leq d < 1.0$); ### = large effect size ($d \geq 1.0$)

Figure 21. Experiment 2 – Comprehensibility on Word-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.6$); # = small effect size ($0.6 \leq d < 1.0$); ## = medium effect size ($1.0 \leq d < 1.4$); ### = large effect size ($d \geq 1.4$)
Table 17: Experiment 2 – Descriptive Statistics for Comprehensibility on Word-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th></th>
<th></th>
<th>Posttest</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
</tr>
<tr>
<td>Treatment</td>
<td>5.90</td>
<td>[.84]</td>
<td>6.02</td>
<td>1.06</td>
<td>6.70</td>
<td>[6.36, 6.74]</td>
<td>1.10</td>
</tr>
<tr>
<td>(n = 20)</td>
<td></td>
<td>[5.50, 6.29]</td>
<td></td>
<td></td>
<td></td>
<td>[6.36, 6.74]</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>5.87</td>
<td>[.70]</td>
<td>5.83</td>
<td>1.15</td>
<td>6.06</td>
<td>[5.50, 6.23]</td>
<td>1.34</td>
</tr>
<tr>
<td>(n = 13)</td>
<td></td>
<td>[5.44, 6.29]</td>
<td></td>
<td></td>
<td></td>
<td>[5.50, 6.23]</td>
<td></td>
</tr>
</tbody>
</table>

4.2.2.1.2 Paragraph-level Productions

For a visual representation of the results, see Figure 22 and Figure 23. For descriptive statistics beyond the visual representation, see Table 18. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest ($U = 122.50, z = -0.28, p = .785, d = 0.01$), but they revealed a significant difference at the time of the posttest ($U = 69.00, z = -2.25, p = .024, d = .75$). With a between-group effect size of $d = 0.75$, the difference in posttest scores between the treatment group and the control group would be considered a medium-sized effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -3.92; p < .001, d = 1.09$), but not for the control group ($z = -1.29, p = .196, d = 0.49$). With a within-group effect size of $d = 1.09$, the gains over time in the treatment group would be considered a medium-sized effect. The gains in the control group ($d = 0.49$) would be considered a negligible effect.
Figure 22. Experiment 2 – Comprehensibility on Paragraph-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; -- = negligible effect size (d < 0.4); # = small effect size (0.4 ≤ d < 0.7); ## = medium effect size (0.7 ≤ d < 1.0); ### = large effect size (d ≥ 1.0)

Figure 23. Experiment 2 – Comprehensibility on Paragraph-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; -- = negligible effect size (d < 0.6); # = small effect size (0.6 ≤ d < 1.0); ## = medium effect size (1.0 ≤ d < 1.4); ### = large effect size (d ≥ 1.4)
Table 18: Experiment 2 – Descriptive Statistics for Comprehensibility on Paragraph-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Treatment</td>
<td>5.91</td>
<td>[5.43,</td>
</tr>
<tr>
<td>(n = 20)</td>
<td>(1.04)</td>
<td>6.40]</td>
</tr>
<tr>
<td>Control</td>
<td>5.92</td>
<td>[5.43,</td>
</tr>
<tr>
<td>(n = 13)</td>
<td>(.82)</td>
<td>6.42]</td>
</tr>
</tbody>
</table>

4.2.2.2 Accentedness

4.2.2.2.1 Word-level Productions

For a visual representation of the results, see Figure 24 and Figure 25. For descriptive statistics beyond the visual representation, see Table 19. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest ($U = 123.50$, $z = -0.24$, $p = .813$, $d = 0.05$), but they revealed a significant difference at the time of the posttest ($U = 62.50$, $z = -2.49$, $p = .011$, $d = 0.86$). With a between-group effect size of $d = 0.86$, the difference in posttest scores between the treatment group and the control group would be considered a medium-to-large effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -3.21; p = .001$, $d = 1.24$), and for the control group ($z = -2.34, p = .019, d = 0.26$). With a within-group effect size of $d = 1.24$, the gains over time in the treatment group would be considered a medium-to-large effect. The gains in the control group ($d = 0.26$) would be considered a negligible effect.
Figure 24. Experiment 2 – Accentedness on Word-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.4$); # = small effect size ($0.4 \leq d < 0.7$); ### = medium effect size ($0.7 \leq d < 1.0$); #### = large effect size ($d \geq 1.0$)

Figure 25. Experiment 2 – Accentedness on Word-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$; – = negligible effect size ($d < 0.6$); # = small effect size ($0.6 \leq d < 1.0$); ### = medium effect size ($1.0 \leq d < 1.4$); #### = large effect size ($d \geq 1.4$)
Table 19: Experiment 2 – Descriptive Statistics for Accentedness on Word-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 20)</td>
<td>4.50</td>
<td>[4.19, 4.81]</td>
<td>4.39</td>
<td>1.19</td>
</tr>
<tr>
<td>(n = 13)</td>
<td>(.66)</td>
<td>(.70)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.2.2.2 Paragraph-level Productions

For a visual representation of the results, see Figure 26 and Figure 27. For descriptive statistics beyond the visual representation, see Table 20. Mann-Whitney tests did not show a significant difference between the treatment and control groups at the time of the pretest ($U = 127.50, z = -0.09, p = .928, d = 0.08$), but they revealed a significant difference at the time of the posttest ($U = 74.50, z = -2.05, p = .040, d = 0.76$). With a between-group effect size of $d = 0.76$, the difference in posttest scores between the treatment group and the control group would be considered a medium-sized effect. Wilcoxon signed-rank tests revealed a significant difference between mean pretest and posttest scores for the treatment group ($z = -3.31; p = .001, d = 0.77$), but not for the control group ($z = -0.55, p = .582, d = 0.07$). With a within-group effect size of $d = 0.77$, the gains over time in the treatment group would be considered a small effect. The gains in the control group ($d = 0.07$) would be considered a negligible effect.
Figure 26. Experiment 2 – Accentedness on Paragraph-level Productions – Between-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.4); # = small effect size (0.4 ≤ d < 0.7); ## = medium effect size (0.7 ≤ d < 1.0); ### = large effect size (d ≥ 1.0)

Figure 27. Experiment 2 – Accentedness on Paragraph-level Productions – Within-group Results

Note: y-axis: percentage correct; n.s. = p > .05; * = p ≤ .05; ** = p ≤ .01; *** = p ≤ .001; – = negligible effect size (d < 0.6); # = small effect size (0.6 ≤ d < 1.0); ## = medium effect size (1.0 ≤ d < 1.4); ### = large effect size (d ≥ 1.4)
Table 20: Experiment 2 – Descriptive Statistics for Accentedness on Paragraph-level Productions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th></th>
<th></th>
<th>Posttest</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
<td>M (SD)</td>
<td>95% CI</td>
<td>Mdn</td>
<td>IQR</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 20)</td>
<td>4.27 (.98)</td>
<td>[3.81, 4.73]</td>
<td>4.46</td>
<td>1.56</td>
<td>5.03 (.97)</td>
<td>[4.57, 5.49]</td>
<td>5.08</td>
<td>1.71</td>
</tr>
<tr>
<td>Control</td>
<td>4.20 (.71)</td>
<td>[3.77, 4.63]</td>
<td>4.17</td>
<td>1.17</td>
<td>4.26 (1.05)</td>
<td>[3.63, 4.90]</td>
<td>4.17</td>
<td>1.46</td>
</tr>
</tbody>
</table>

4.2.3 Summary of Findings

Table 21 summarizes the findings of Experiment 2 (for the same table with exact p-values and d-scores, see Appendix K).

Table 21: Experiment 2 – Summary of Findings

<table>
<thead>
<tr>
<th></th>
<th>Between-group comparison (at posttest)</th>
<th>Within-group comparison (pretest to posttest)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment &gt; Control</td>
<td>Treatment</td>
</tr>
<tr>
<td>Perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness Detection Task</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Sound Discrimination Task</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>
The between-group comparisons—that is, the comparison of posttest scores between the treatment group and the control group—revealed the following: for all measures of both perception and production, learners in the treatment group significantly outperformed learners in the control group. All comparisons showed a medium-sized effect, except in the sound discrimination task, in which the difference in posttest scores exhibited a large effect.

The within-group comparisons—that is, the comparison from pretest to posttest within each of the two groups—revealed the following: learners in the treatment group improved significantly from pre- to posttest on all measures of perception and production tests, with mostly medium-sized effects for the gains over time. Learners in the control group did not improve significantly over time on any measures of the perception task, and only on one measure (i.e., in accentedness on the word-level) in the production task. All effects for the gains over time were negligible in the control group.

### 4.2.4 Exit Questionnaire Responses

Seventeen of the 20 learners in the treatment group completed the Exit Questionnaire. Table 22 provides an overview of the mean scores for the six statements on the perceived usefulness of the iCPR pronunciation training in the first part of the debrief. These statements had to be rated on a scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating

<table>
<thead>
<tr>
<th>Accentedness</th>
<th>Word-level</th>
<th>**</th>
<th>***</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>##</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paragraph-level</td>
<td></td>
<td></td>
<td>***</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = $p > .05$; * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$

- = negligible effect size; # = small effect size; ## = medium effect size; ### = large effect size
“strongly agree”. Results showed very high values (means above 6 out of 7) for question 4, indicating that the learners strongly agreed that the benefits of the pronunciation training outweighed any technological difficulties. In fact, two students responded with a write-in, stating that the question was not even applicable because they did not encounter any technological difficulties. Results showed high values (means above 5 out of 7) for questions 1, 2, 3, and 5: learners agreed that they enjoyed working on the pronunciation homework, that they felt like they learned something from the homework, that they felt like their pronunciation of German improved through the exercises, and that the exercises had the additional benefit of reviewing vocabulary. Results showed medium-high values (means above 4 out of 7) for question 6, indicating that learners might have benefited from receiving teacher feedback throughout the semester.

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Statement To Be Rated By Learners</th>
<th>Mean Response (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I enjoyed working on the pronunciation homework.</td>
<td>5.2 (3-7)</td>
</tr>
<tr>
<td>2.</td>
<td>I feel like I learned something from the pronunciation homework.</td>
<td>5.7 (2-7)</td>
</tr>
<tr>
<td>3.</td>
<td>I feel like my pronunciation of German improved through the homework exercises.</td>
<td>5.9 (3-7)</td>
</tr>
<tr>
<td>4.</td>
<td>Any technological problems that I encountered were worth the overall benefit.</td>
<td>6.5 (5-7)</td>
</tr>
<tr>
<td>5.</td>
<td>Not only did the pronunciation homework improve my pronunciation but it also helped review some vocabulary.</td>
<td>5.5 (3-7)</td>
</tr>
<tr>
<td>6.</td>
<td>I would have taken the pronunciation homework more seriously and would have benefited more from it if I had received teacher feedback throughout the semester.</td>
<td>4.5 (1-7)</td>
</tr>
</tbody>
</table>

Note: Statements were rated on scale from 1 to 7, with 1 indicating “strongly disagree” and 7 indicating “strongly agree”.

Table 22: Experiment 2 – Mean Scores on Exit Questionnaire – Online iCPR Training Group
The second part of the questionnaire comprised of seven questions targeting the learners’ behavior in working through the pronunciation units, with a limited number of possible responses per question. For Question #1 (How much time, on average, did you spend on each pronunciation homework?), the majority of the learners reported spending 15-20 minutes on each iCPR unit (5-10 min: 0 learners; 10-15 min: 4 learners; 15-20 min: 10 learners; 20-25 min: 3 learners). For Question #2 (How often, on average, did you listen to the native speaker recording of each word?), eight learners reported listening 2-3 times and the remaining nine learners were distributed over the remaining answer options (1-2 times: 2 learners; 2-3 times: 8 learners; 3-5 times: 4 learners; 5-7 times: 2 learners; more than 7 times: 1 learner). For Question #3 (How often, on average, did you practice saying each word before you recorded yourself on all the words?), the majority of the learners reported listening between 2-5 times (1-2 times: 1 learner; 2-3 times: 5 learners; 3-5 times: 8 learners; 5-7 times: 2 learners; more than 7 times: 1 learner). For Question #4 (Did you always practice saying each word until you were completely happy with your own pronunciation before you recorded yourself?), the majority of learners answered “Usually yes” (always: 1 learner; usually yes: 12 learners; about half the time: 3 learners; usually not: 1 learner; never: 0 learners). For Question #5 (There was no control of whether you completed the listening exercises. Honestly, how many of the 10 listening exercise homework units did you complete?), a majority of learners reported completing all ten units (all 10: 12 learners; 9 of 10: 2 learners; 8 of 10: 0 learners; 7/6 of 10: 0 learners; 5/4 of 10: 2 learners; 3/2 of 10: 0 learners; 1/0 of 10: 1 learner). For Question #6 (Did you enjoy the listening exercises or the speaking exercises more?), the learners reported either liking both equally or liking the speaking exercises most, but only two learners reported liking the listening exercises most (listening: 2 learners; speaking: 8 learners; I enjoyed both equally: 7 learners). For Question #7 (Do you think
that three days of homework practice on each pronunciation problem was enough to help you or would you have preferred more or less practice?), the majority of learners answered that three units were a good amount (more practice: 3 learners; it was a good amount: 14 learners; less practice: 0 learners).

The third page of the questionnaire comprised of five open-ended questions. In response to the question “What did you like about the pronunciation homework exercises?”, three learners mentioned the metalinguistic explanations being helpful:

**Excerpt 1**
The chance to practice my speaking. Very helpful when they went into specific techniques or ways of shaping the mouth/tongue with more difficult German sounds. (Participant 1309)

Three more learners reported that they simply liked receiving pronunciation practice:

**Excerpt 2**
It was good to get some practice. I felt like I learned the words better when I hear and say them rather than just read them in the book. (Participant 1104)

Two learners mentioned that they enjoyed the additional vocabulary review:

**Excerpt 3**
I enjoyed being able to work on certain pronunciations as well as being able to learn new vocabulary. (Participant 1305)

In response to the second question “What did you not like about the pronunciation exercises?”, the majority—that is eight learners—answered by saying that the question was not applicable or wrote explicitly that they could not think of anything they did not like about the iCPR units:

**Excerpt 4**
I found the exercises very helpful. I can’t think of anything I didn’t like. (Participant 1113)
Three learners, however, mentioned technical difficulties or said that the technological component could be improved:

**Excerpt 5**
*The PPT format was a bit of a pain, clicking on the button and having to have it in display mode. I think a web-based version with easily clickable buttons would be ideal. Optionally, have a record button next to the practice words on the site (maybe flash based w/ access to my microphone?). Would eliminate the need for audacity, PPTs, conversion and uploading. (Participant 1309)*

Two learners mentioned the lack of instructor feedback and the lack of a grade as things they did not like about the iCPR units:

**Excerpt 6**
*I was never actually sure if the stuff I recorded was correct or not. (Participant 1105)*

**Excerpt 7**
*I did not like that they weren’t for a grade, I spent more time than I would’ve liked every week on them. (Participant 1101)*

In response to the question “Do you have any suggestions for future improvements of the pronunciation exercises? If so, what are they?”, answers varied widely. Two learners mentioned the lack of feedback:

**Excerpt 8**
*I think giving more feedback on how correct our recordings are would be good. (Participant 1105)*

Two students provided technological suggestions:

**Excerpt 9**
*Maybe create a site or upload them to a google drive since having to access and record at the same time was a pain. I ended up downloading powerpoint on my ipad to look and listen to the slides while I recorded on my computer. (Participant 1101)*

Other responses targeted the implementation of the iCPR units in the overall course structure and the practice of vocabulary items.
In response to the question “What, if any, are the areas of pronunciation not treated in the pronunciation exercises that you would have liked to work on and improve”, ten learners responded that they did not miss anything:

**Excerpt 12**
* I felt that all areas were covered. (Participant 1113)

The other learners suggested including more reviews and words in sentences:

**Excerpt 13**
* I would have liked to see some repeats and refreshers of the most used sounds like the ch, the z, and the ei. I noticed in the live chat that everyone is still struggling with these pronunciations. Also sentence and paragraph exercises. (Participant 1101)

**Excerpt 14**
* I would have liked to have more sentences dealing with what we were working on at that time, especially the numbers as they got more complex. (Participant 1110)

One student reports that the iCPR units were their favorite part of the course:

**Excerpt 15**
* Generally, I enjoyed these exercises the most. I just wished there was more of a chance to do spoken or listening comprehension for the course. (Participant 1111)

In response to the last question “Any final comments?”, six learners did not respond or wrote thank you notes:

**Excerpt 16**
* I’ve enjoyed this course, I just wanted to convey my enjoyment to the professor and staff. Vielen Danke! I am happy to have the pronunciation exercises. (Participant 1111)
Two learners commented on the course in its entirety rather than on the pronunciation exercises:

**EXCERPT 17**

*It was interesting taking an online language class. (Participant 1301)*

### 4.3 Discussion

Experiment 2 investigated the effectiveness of computer-delivered iCPR pronunciation training in an online learning environment by comparing learners who received iCPR pronunciation training (i.e., the treatment group) with learners who did not receive targeted pronunciation training, but otherwise followed the same first semester German curriculum (i.e., the control group). Just as outlined for Experiment 1 in Chapter 3, learners’ perception and production skills were assessed. Perception skills were assessed via an accentedness detection task and a sound discrimination task, and production skills were assessed via comprehensibility and accentedness ratings on the word- and paragraph-level.

Comparisons between the treatment and control group at the time of the posttest revealed that the learners who received iCPR pronunciation training significantly outperformed learners who did not receive pronunciation training on all measures of perception and production. The difference in posttest scores showed medium to large effect sizes, thus, we can assume that these findings are robust. Comparisons within each group from pretest to posttest revealed that the iCPR group improved significantly over time on all conditions, which is also reflected in the mostly medium-sized effects for the iCPR treatment over time. The control group, however, did not improve significantly over time on any measures of perception and production, except for measures of accentedness on the word-level, but this improvement only showed a negligible effect size, thus it is probably not a robust finding.
Overall, these findings suggest that learners who received iCPR training significantly improved their pronunciation skills from the beginning to the end of the semester. Not only were learners in the iCPR group significantly better than their peers at detecting accented German and discriminating between similar sounds in nonsense words, but they were also significantly easier to understand and significantly less accented in producing speech at the word- and paragraph-level. The results further suggest that simply following the standard first semester online curriculum did not suffice to improve learners’ pronunciation skills from the beginning to the end of the semester. This is a discouraging and somewhat surprising finding, since it has to be kept in mind that most of these learners began the semester with no previous knowledge of German, and while surely their knowledge of German improved—otherwise they would not be able to pass the class—, apparently their pronunciation skills did not improve significantly. Unfortunately, there are no other studies on the development of pronunciation training in online language learning that would shed light on the question whether this finding is common or not. The scarce previous literature that exists only addresses oral proficiency development as a whole. As Lin’s (2014) meta-analysis of studies investigating the influence of computer-mediated communication on oral proficiency showed, most studies defined oral proficiency in terms of lexical and syntactic accuracy and overall fluency, but only four studies even considered pronunciation as a factor among others in oral proficiency. Thus, while previous research has shown that oral proficiency develops mostly on par in online and F2F classrooms, we have no data on the development of pronunciation skills alone (R. Blake et al., 2008; C. Blake, 2009; Isenberg, 2010). In an unpublished pilot study presented at the 9th Pronunciation in Second Language Learning and Teaching Conference, Inceoglu (2017) investigated the effect of pronunciation training on the development of learners’ pronunciation, fluency, and
comprehensibility in an asynchronous online French class at the lower intermediate level. Her pronunciation training was comprised of different computer-assisted exercises, such as discrimination exercises, IPA transcriptions, and oral recordings. She found significant improvement in segmental productions, and a significant decrease in unfilled pause frequency and use of forbidden French liaisons, suggesting that learners in the online French class benefited from the pronunciation intervention. Yet, Inceolgu’s study did not include a control group. Thus, again, we have no indication how learners’ pronunciation skills develop without pronunciation instruction. Nevertheless, Inceoglu’s findings are in line with the findings in the present study. Particularly in light of the absence of improvement in the control group, it is important to see that the iCPR training had such a robust positive effect on the pronunciation skills of online learners of German.

The empirical findings of benefits of iCPR pronunciation training in distance language learning were further reflected in the learners’ assessment of the method on the exit questionnaire. That is, learners agreed with the statements that they enjoyed working on the pronunciation exercises (5.2/7), that they felt like they learned something (5.7/7), and that they felt that their pronunciation had improved through the use of the iCPR units (5.9/7). Additionally, learners agreed that the iCPR units helped them review vocabulary as well (5.5/7). It is also worth noting that learners encountered almost no technological difficulties with the method of iCPR, which becomes apparent in the high rating of the statement “Any technological problems that I encountered were worth the overall benefit” (6.5/7) and the fact that several students wrote in that they did not encounter any technological problems—even though this was not a write-in question. Thus, it appears as if the method was indeed very accessible for learners. On the second part of the exit questionnaire, which targeted learners’ habits in completing the
homework exercises, learners report having spent 15-20 min on average on each unit, which is five minutes more than the results in the F2F environment showed. Moreover, as opposed to the face-to-face environment, 12 of 17 learners reported having completed all perception exercises and an additional two reported having completed 9 of 10 perception exercises. Thus, a large majority of learners completed the perception exercises, which is in line with the significant improvement on both perception and production measures in the treatment group. (For further discussion of the differences between the F2F and online group see Section 5.2). In the open-ended questions at the end of the exit questionnaire, many learners mentioned that they enjoyed receiving targeted pronunciation training as part of their online course. It is also important to note that even in response to the question “What did you not like about the pronunciation exercises?”, a majority of the learners responded that there was nothing they did not like. This strongly suggests that learners saw the pronunciation training as a very positive component of their learning experience. Some learners did mention, however, that the technology for iCPRs could have been improved and that iCPRs should have included practice on the sentence level. These are valid concerns that can be addressed in future research (see also Section 5.4).

Overall, both qualitative and quantitative results suggest that learners in an online learning environment can greatly benefit from including iCPR pronunciation training in the standard curriculum. Seeing that most online language classes currently do not include targeted pronunciation training, the method of iCPR is particularly appealing, since it can simply be added-on to any existing curriculum, and does not cause additional work for the instructors. In light of the lack of improvement in the control group, it seems even more important to find ways to offer pronunciation training to all beginning language learners in distance language environments.
CHAPTER 5: Conclusions

5.1 Overview of Findings

The present dissertation explored the effectiveness of homework-based, computer-delivered iCPR pronunciation training in face-to-face (F2F) and distance learning environments. In the F2F learning environment (Experiment 1), two levels of comparisons were employed: (1) learners who received iCPR pronunciation training (i.e., the treatment group) were compared to learners who did not receive targeted pronunciation training, but otherwise followed the same GER 001 curriculum (i.e., the control group), and (2) learners who received iCPR pronunciation training were compared to learners who received teacher-directed in-class pronunciation training (i.e., the comparison group). In the distance learning environment (Experiment 2), only one level of comparison was employed since there was no in-class instruction in this learning environment. Thus, learners who received iCPR pronunciation training (i.e., the treatment group) were compared to learners who did not receive targeted pronunciation training, but otherwise followed the same GER 001 curriculum (i.e., the control group). The training was administered over the course of one academic semester, and gains in learners’ L2 pronunciation skills were measured through a pretest in the second week of the semester and a posttest in the 14th week of the semester. Each test consisted of a perception and a production task. The perception task was a binary-forced-choice task, and learners’ pre- and posttest perception skills were measured based on the total score learners reached in this task. The production task consisted of word- and paragraph-readings. Learners’ pre- and posttest productions were rated for comprehensibility and accentedness by native speakers of German. The perception scores and the native speaker ratings of learners’ productions constituted the data for statistical analyses.
As stated in Section 2.4.2, the present dissertation investigated the following research questions:

For the face-to-face learning environment:

1. Is pronunciation training delivered through iCPR units in a face-to-face learning environment effective in significantly improving novice learners’ L2 perception skills when compared to a control group?

2. Is pronunciation training delivered through iCPR units in a face-to-face learning environment effective in significantly improving novice learners’ L2 production skills when compared to a control group?

3. Is pronunciation training delivered through iCPR units in a face-to-face learning environment as effective as in-class pronunciation training in improving novice learners’ L2 perception skills?

4. Is pronunciation training delivered through iCPR units in a face-to-face learning environment as effective as in-class pronunciation training in improving novice learners’ L2 production skills?

For the distance learning environment:

5. Is pronunciation training delivered through iCPR units in a distance learning environment effective in significantly improving novice learners’ L2 perception skills when compared to a control group?

6. Is pronunciation training delivered through iCPR units in a distance learning environment effective in significantly improving novice learners’ L2 production skills when compared to a control group?
The major findings addressing each research question are presented below.

*For the face-to-face learning environment (Experiment 1):*

1. Learners in the iCPR group significantly outperformed learners in the control group on some, but not on all, measures of perception skills. Considering descriptive trends and effect sizes, however, data overall suggest that iCPR pronunciation training helped learners improve their L2 perception skills compared to a control group.

2. Learners in the iCPR group significantly outperformed learners in the control group on measures of ease of understanding and strength of accent on word-level productions. Descriptively, learners in the iCPR group also outperformed learners in the control group on paragraph-level productions, although this difference did not reach significance. This suggests that iCPR pronunciation training helped learners improve their L2 production skills on the word-level, and could potentially help them improve their L2 production skills on the paragraph-level as well.

3. There were no significant differences between the learners’ gains over time and posttest scores between the iCPR and the in-class group on any of the perception tests. This suggests the iCPR pronunciation training was at least as effective as in-class pronunciation training in improving novice learners’ L2 perception skills.

4. There were no significant differences between the learners’ gains over time and posttest scores between the iCPR and the in-class group on any of the production tests. This suggests the iCPR pronunciation training was at least as effective as in-class pronunciation training in improving novice learners’ L2 production skills.
For the distance learning environment (Experiment 2):

5. Learners in the iCPR group significantly outperformed learners in the control group on all measures of perception skills. This suggests that iCPR pronunciation training helped learners improve their L2 perception skills compared to a control group.

6. Learners in the iCPR group significantly outperformed learners in the control group on all measures of production skills. This suggests that iCPR pronunciation training helped learners improve their L2 production skills compared to a control group.

5.2 Differences in Experiment 1 and Experiment 2

The differences in learner backgrounds\(^\text{18}\) did not allow for a direct comparison of the quantitative results between the learners in the F2F and distance learning environment. Nevertheless, we can still draw conclusions about each learning environment based on a general comparison of the findings in Experiment 1 and Experiment 2. Learners in both learning environments were similar in that they benefited from the iCPR pronunciation training and significantly outperformed the respective control groups at the time of the posttest. Yet, learners in both learning environments differed in the fact that, in the F2F learning environment, both the treatment and the control group improved significantly over time—the gains in the treatment group were just so much larger that these learners still outperformed the control group at the end of the semester, even though the control group had improved significantly over time as well—whereas in the distance learning environment, only the treatment group improved significantly over time. As discussed in Section 3.3, we would expect all learners to show significant

---

\(^{18}\) Learners in the F2F and distance learning environment differed in age, language background, experience abroad, and most importantly, learning circumstances. That is, while most F2F learners were fulltime students under 21, most distance learners were adult learners who completed an online degree while working part- or fulltime.
improvement of their pronunciation skills over time, since the learners in the present study were first semester learners who started without prior knowledge of German and should improve significantly after 14 weeks of German instruction. The fact that the control group learners in the distance learning environment did not show significant improvement of their pronunciation skills from the second to the 14th week of the semester then points to a deficiency in the current GER 001 online curriculum, and stresses the importance of a pronunciation intervention as an obligatory component in distance language learning.

It has to be pointed out that previous research has shown that distance language learners’ oral proficiency generally develops on par with their peers in F2F environments (R. Blake et al., 2008). Moreover, it can be assumed that learners’ knowledge of vocabulary and grammar as well as their writing, reading, and listening skills developed on par with F2F learners. That is, learners in the online environment used the same textbook as their F2F peers, were taught by instructors who had taught the same class in a F2F environment at Penn State before teaching it online for World Campus, and had to pass similar midterm and final exams as their peers in the F2F learning environment. Thus, we should not draw the conclusion that learners cannot learn a foreign language in distance language instruction, but rather, that there is a difference between distance and F2F learning environments that has an effect on the development of learners’ pronunciation skills. The most obvious difference between the two learning environments of course is the time F2F learners spend in the classroom. With a format of four 50-minutes meetings a week in the F2F environment, this time equals 60 hours of in-class time over 15 weeks, in which the learners not only listen to spoken input from their instructor and their peers, but also have many opportunities to speak themselves. Looking at the differences in results in Experiment 1 and Experiment 2 then, it is likely that the time spent in the classroom in the F2F

175
environment allowed learners in the control group to improve significantly over time, while the lack of this classroom-environment hindered learners in the distance learning environment to improve their pronunciation skills significantly. Again, this stresses the importance to make a targeted pronunciation training, such as iCPR training, available to learners in the distance environment. Further, it raises the question whether online learners would benefit from more interaction with their instructor and peers—e.g., in additional live video chats—in order to improve their pronunciation skills even in the absence of targeted pronunciation training.

In addition to comparing the quantitative findings for the two learning environments, it is worth looking at the qualitative findings taken from the exit questionnaires. The responses in both learning environments yielded very similar results: learners in both environments were similar in their ratings of whether they enjoyed the iCPR units, whether they felt like their pronunciation improved through the practice, etc. The two groups differed, however, in that the online learners reported having spent more time on each iCPR unit (15-20 minutes instead of 10-15 minutes), and that a majority of online learners completed 90% or more of the perception exercises, while in the F2F environment, a majority of learners did not even complete 50% of the perception exercises. These results suggest that the online learners took the iCPR units more seriously—possibly because they had less opportunities to practice their German speaking due to the lack of classroom meetings, and thus valued opportunities for additional practice.

Finally, when asked whether the overall amount of pronunciation training over the course of the semester had been too little, just the right amount, or too much, no learner in either learning environment answered that they received too much pronunciation training (in the F2F learning environment, all 18 learners responded that it was a good amount, in the online learning environment three learners responded that they would have liked more training and the
remaining 14 learners responded that it was a good amount). This strengthens previous findings that show that learners themselves believe in the value of pronunciation training and want to be offered opportunities to improve their pronunciation (Couper, 2003; Derwing & Munro, 2015). Apparently, it does not matter what learning environment learners are taking foreign language classes in; they generally value pronunciation training opportunities.

5.3 Pedagogical Implications

As outlined in Section 2.1.2 of the Literature Review, recent studies have found that pronunciation instruction is still neglected in the foreign language classroom (Foote et al., 2016), despite the growing number of research studies on L2 pronunciation that were published in the past two decades. Reasons for this gap between research and practice seem to be manifold—for instance, teachers show very limited interest in learning about findings published in research journals (Wahid & Sulong, 2013), and research studies do not always address the right questions for practical implementation of pronunciation instruction (Levis, 2016b). Yet, the most important reason why pronunciation instruction is still neglected in L2 classrooms seems to be the fact that many teachers do not feel comfortable or adequately prepared to teach pronunciation (Breitkreutz et al., 2011; Derwing & Munro, 2005; Foote et al., 2011; Grim & Sturm, 2016; Henderson et al., 2012). Considering this reason while keeping in mind findings that have shown that learners themselves see pronunciation training as very important (Couper, 2003; Derwing & Munro, 2015), and that learners’ pronunciation skills are likely to improve through training (Thomson & Derwing, 2015), it becomes apparent that a solution had to be found for how to include pronunciation instruction in the standard L2 curriculum without burdening instructors.
that lack the preparation to teach pronunciation. The homework-based method of iCPRs investigated in the present dissertation can offer such a solution.

J. Lee et al. (2015) warn that in the design of pronunciation interventions (and teaching materials in general), “instructional costs (time and energy) must be weighed against the potential benefits for L2 learners” (p. 349). Just as with the creation of any effective teaching material, it certainly takes some time to create iCPR units at first, but their design allows for them to be easily reused and shared, which makes up for the time invested in their assembly. Practical tips on how to get started with the assembly of iCPR units—e.g., how to choose practice items, how best to record tokens, etc.— can be found in Martin (2015) and Martin (2017). It is important to note that once the units are created, it is extremely easy to share them with colleagues at the same school or at other institutions, which allows for a large number of learners to benefit from the pronunciation instruction with limited instructional cost. Moreover, the method of iCPRs causes little extra work for the instructor, since the results of the present study have shown that corrective feedback on the part of the instructor is not necessary to improve learners’ pronunciation. While the instructor might still choose to provide regular or occasional feedback on the learners’ recordings, the improvement in the absence of feedback is a noteworthy finding that might motivate instructors to adopt iCPR units in their teaching. Finally, the computer-based nature of iCPR units allows for them to be assigned as homework, which saves valuable in-class time.

The design of the iCPR units further allows for them to be added on to any existing curriculum. Each unit is designed to take learners 10-15 min to work through, so they can be assigned in addition to other homework. Moreover, instructors can be very flexible in deciding which units to assign as homework. That is, it would be entirely possible to assign only those
units that instructors deem beneficial for their learners, or to assign different units to different learners based on their individual pronunciation problems. Thus, iCPR units are versatile in their implementation and offer multiple options to help intact classes or individual learners improve their pronunciation skills.

The present study suggested that iCPR pronunciation training is particularly beneficial for learners in online learning environments, who receive less spoken input and have fewer chances to practice their speaking skills than learners in F2F environments. The fact that, in the distance learning environment, the learners in the control group did not significantly improve their pronunciation skills over the course of the semester suggests that an instructional method that allows for pronunciation training to be included in any L2 online course is direly needed. Since iCPR units can serve as an easy add-on to any L2 online curriculum, they serve as a promising method to allow for pronunciation training to be included in online language classes. On a similar note, there is no previous research that suggests that this type of pronunciation instruction should only be beneficial to learners in post-secondary education. In fact, the versatile use of iCPR units as add-ons to existing curricula, and the easily accessible technology used in their design make them a viable option for middle- and high school classrooms as well.

Finally, when considering pedagogical implications of the present study, it has to be noted that the learners’ feedback on the method of iCPRs was largely very positive. Learners in both the face-to-face and distance learning environments reported that they enjoyed working with the iCPR units, and that they felt that their pronunciation improved through the training. Thus, the face validity of this instructional method was high, which is a desirable educational goal in foreign language instruction and can lead to improved learner motivation (Brown & Abeywickrama, 2010).
5.4 Limitations and Directions for Future Research

The findings presented in this dissertation suggest that iCPR pronunciation training is an effective method to provide learners in face-to-face and distance learning environments with targeted pronunciation training. Nevertheless, the present study had some limitations and leaves several questions unaddressed and open for future research.

One of those questions is whether the findings—which reveal the effectiveness of iCPR-based pronunciation training—are only relevant for L2 German or whether they would transfer to other languages. Similarly, it remains open for future research whether iCPRs are only beneficial to novice learners or can also be helpful to learners of more advanced proficiency levels. Nothing in the theoretical approaches that informed the design of iCPR units (see Section 2.4.1.2) suggests that the method would only be successful for use in L2 German. In fact, Tanner and Landon’s (2009) original CPR study was conducted with ESL learners. Moreover, the effectiveness of iCPRs for L2 French is currently being investigated at Purdue University by Jessica Sturm. Sturm implemented iCPRs in the first four semesters of French language instruction and her findings will thus not only inform the question of whether iCPRs can be transferred to French, but also what proficiency levels they are (most) beneficial for.

Another question that remains open for future research is what would have changed if the learners in the iCPR group had received feedback on their recordings throughout the semester. As discussed in Section 3.3.6, a variety of ways to provide feedback would have been possible, for example, a spot check on only some of the recordings, or encouraging comments to support the training’s face validity and reassure learners that the instructor is aware of their work and progress. It would have been interesting to see if such feedback could have led to even larger gains in the iCPR group. That is, particularly in the face-to-face environment where significant
differences in production skills were found only on the word-level, future research could address whether feedback can lead to a significant difference in production skills between the iCPR and the control group on the paragraph-level as well.

Two more methodological limitations of the present study were the lack of measures of free speech, and of a delayed posttest. J. Lee et al. (2015) and Thomson and Derwing (2015) stress the importance of assessing pronunciation in a more free and communicative context because these types of tests have greater authenticity and ecological validity than the commonly used word- and sentence-reading tests. While this is certainly true, the present study was still designed with measures of word- and sentence-reading tests because the pretest data was collected from novice learners in the second week of the semester, when learners are simply not yet able to communicate freely or answer a prompt spontaneously in the L2. Nevertheless, additional free speech data (responses to six short questions) were elicited from each learner at the posttest. I plan to analyze these data in the near future to investigate whether there are correlations between the ratings of the free speech productions and learners’ results on the word- and paragraph-level. This would allow me to gauge whether learners’ gains in controlled speech environments are related to their performance of free speech tasks at the end of the semester. Further, it would be desirable for future studies to include a measure to assess if the gains from the iCPR pronunciation training are retained over time. A delayed posttest several weeks after the end of the pronunciation training would provide insight into how robust the improvement in perception and production skills are. Yet, since the present study was designed with a posttest in the last week of the semester, no delayed posttest measures were collected. While it is rather difficult to reach learners after the class disperses at the end of a semester, it would certainly be valuable to collect additional data at a later point—even if only a subset of the learners could be
reached. Such a design might be possible, for instance, if learners continue in the institution’s language program the following semester and would agree to a delayed assessment of their pronunciation skills.

Another limitation of the present study was that the iCPR training did not include a control measure for completion of the perception units, but only for production units. The exit questionnaire responses showed that, in the face-to-face environment, this led to a majority of learners completing less than half of the perception units. This raises the question whether the results would have looked different if the learners had completed more of the perception training. It also suggests that the design of future iCPR interventions should include a control measure for completion of perception units, such as a short quiz at the end of each unit, or even a cumulative quiz in class that would increase face validity of the exercises and encourage learners to take them more seriously.

Finally, in the present study, learners with L1s other than English were excluded from the data analysis. This was a methodologically motivated choice since the iCPR units were designed specifically with pronunciation problems of L1 American English speakers learning L2 German in mind. Learners from other language backgrounds often exhibit different pronunciation problems than American learners of German. To make the present study feasible, however, pronunciation targets had to be limited in number and, thus, the most prominent features impeding American learners’ intelligibility were selected. Nevertheless, given the increasingly multicultural population in classrooms at US institutions, it is important to find teaching materials that are equally beneficial for learners of all L1 backgrounds. Therefore, the design of future iCPR materials should account for more diverse language backgrounds and find a way to make pronunciation training valuable for all learners in the classroom or at least investigate
whether learners with L1s other than English benefit as much from the training as L1 English learners.

5.5 Final Remarks and Conclusion

The goal of the present dissertation was to present a solution for the discrepancy between the importance of pronunciation in L2 communication and the neglect of pronunciation training in the classroom. To this end, I investigated the effectiveness of a homework-based method of pronunciation instruction, called iCPRs, among novice learners of L2 German in F2F and distance learning environments. All in all, results from both learning environments showed that learners’ perception and production skills improved significantly through the use of iCPR pronunciation training. A qualitative exit questionnaire further revealed that learners in both learning environments enjoyed working with the iCPR pronunciation units, and that the method had a large face validity among the learners. Taken together, these findings suggest that pronunciation training cannot only be taught in the classroom, but can also be assigned as homework. This is an important finding for instructors teaching in both F2F and online environments. In F2F settings, assigning pronunciation training as homework can relieve in-class time pressure and minimize the burden on instructors that do not feel adequately prepared to teach pronunciation. In online environments, assigning pronunciation instruction through iCPRs offers a new possibility to add a pronunciation component to curricula that often do not focus on pronunciation at all.

In conclusion, the impact of this study can be wide-reaching. The described research is directly applicable to any foreign language classroom at the K-12 and post-secondary education levels. The method of iCPRs is designed for direct implementation by language learners and can
supplement existing world language curricula with minimal effort by the instructors. The findings of this study directly link linguistic research to L2 learning and can positively impact learners’ overall L2 proficiency and oral communication skills.
References


Murphy, J. (2014). Teacher training programs provide adequate preparation in how to teach pronunciation. In L. Grant (Ed.), Pronunciation myths: Applying second language research to classroom teaching (pp. 188-244). Ann Arbor, MI: University of Michigan Press.


### Appendix A: List of Sound Discrimination Training Items

<table>
<thead>
<tr>
<th>target</th>
<th>distractor</th>
<th>target</th>
<th>distractor</th>
<th>target</th>
<th>distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>consonantal [r]</strong></td>
<td></td>
<td><strong>[e:]</strong></td>
<td></td>
<td><strong>[ø:]</strong></td>
<td></td>
</tr>
<tr>
<td>ara</td>
<td>acha</td>
<td>pet</td>
<td>pät</td>
<td>fös</td>
<td>füs</td>
</tr>
<tr>
<td>üri</td>
<td>üRI R=English r</td>
<td>gek</td>
<td>geyk</td>
<td>sök</td>
<td>suk</td>
</tr>
<tr>
<td>ora</td>
<td>oha</td>
<td>bet</td>
<td>böt</td>
<td>büt</td>
<td>but</td>
</tr>
<tr>
<td>ara</td>
<td>aRa</td>
<td>def</td>
<td>deif</td>
<td>föt</td>
<td>fot</td>
</tr>
<tr>
<td>uri</td>
<td>uchi</td>
<td>kek</td>
<td>keik</td>
<td>döp</td>
<td>dup</td>
</tr>
<tr>
<td>ora</td>
<td>ocha</td>
<td>pet</td>
<td>peyt</td>
<td>döf</td>
<td>dof</td>
</tr>
<tr>
<td>üro</td>
<td>üho</td>
<td>tef</td>
<td>täf</td>
<td>tōf</td>
<td>tof</td>
</tr>
<tr>
<td>iro</td>
<td>iRo</td>
<td>fet</td>
<td>föt</td>
<td>gök</td>
<td>guk</td>
</tr>
<tr>
<td>ari</td>
<td>achi</td>
<td>fes</td>
<td>fäs</td>
<td>pöt</td>
<td>püt</td>
</tr>
<tr>
<td>erü</td>
<td>ehü</td>
<td>tep</td>
<td>teyp</td>
<td>kōk</td>
<td>kek</td>
</tr>
<tr>
<td><strong>&lt;z&gt;</strong></td>
<td></td>
<td><strong>[y:]</strong></td>
<td></td>
<td><strong>[o:]</strong></td>
<td></td>
</tr>
<tr>
<td>azi</td>
<td>aßi</td>
<td>füs</td>
<td>fies</td>
<td>pot</td>
<td>put</td>
</tr>
<tr>
<td>ize</td>
<td>ise</td>
<td>püt</td>
<td>pöt</td>
<td>tōf</td>
<td>toaf</td>
</tr>
<tr>
<td>üzi</td>
<td>ütschi</td>
<td>büt</td>
<td>but</td>
<td>dok</td>
<td>dök</td>
</tr>
<tr>
<td>oza</td>
<td>oßa</td>
<td>füt</td>
<td>föt</td>
<td>fot</td>
<td>føt</td>
</tr>
<tr>
<td>üzo</td>
<td>üso</td>
<td>düp</td>
<td>düp</td>
<td>dot</td>
<td>dut</td>
</tr>
<tr>
<td>iza</td>
<td>itha (Engl Th)</td>
<td>düf</td>
<td>dief</td>
<td>sot</td>
<td>sout</td>
</tr>
<tr>
<td>üzo</td>
<td>üßo</td>
<td>tüf</td>
<td>tuf</td>
<td>bof</td>
<td>buf</td>
</tr>
<tr>
<td>uzi</td>
<td>usi</td>
<td>gük</td>
<td>giek</td>
<td>fos</td>
<td>fös</td>
</tr>
<tr>
<td>ezo</td>
<td>eto</td>
<td>sük</td>
<td>sök</td>
<td>kof</td>
<td>kouf</td>
</tr>
<tr>
<td>azi</td>
<td>asi</td>
<td>kük</td>
<td>kük</td>
<td>gok</td>
<td>gök</td>
</tr>
<tr>
<td><strong>[xl, [ç]</strong></td>
<td></td>
<td><strong>&lt;ie&gt;, &lt;ei&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acho</td>
<td>ako</td>
<td>piet</td>
<td>peit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>icha</td>
<td>ischa</td>
<td>giek</td>
<td>geik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uchi</td>
<td>uri</td>
<td>kiek</td>
<td>keik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>üchi</td>
<td>üschi</td>
<td>fięp</td>
<td>feip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>achü</td>
<td>atschü</td>
<td>dief</td>
<td>dief</td>
<td></td>
<td></td>
</tr>
<tr>
<td>echu</td>
<td>eschu</td>
<td>feit</td>
<td>fiet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ocha</td>
<td>oka</td>
<td>teis</td>
<td>ties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auche</td>
<td>aure</td>
<td>beit</td>
<td>biet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eicho</td>
<td>eischo</td>
<td>seik</td>
<td>siek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iche</td>
<td>ische</td>
<td>deip</td>
<td>diep</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Sample iCPR Production Unit

Pronunciation of German <ö>

- Just as it is the case for ö, you have to know that the sound Germans make when they say ö has no equivalent in English.
- That means, it is a new sound that you have to learn from scratch!
- For your mouth this means that it is not used to producing this sound and you need to actively control what your lips and tongue are doing to arrive at producing the right sound.

Now that you had a refresher on the German /e/, say the sound but pucker your lips as if to whistle.

Long German <ö> is just like German /e/, except that you round your lips!

This should sound something like:

- Make sure you really round your lips a lot. You can even whistle a little first to make sure you really round your lips as much as you can.
- No, you do not look silly! You are trying hard to utter better at a foreign language. Two thumbs up for this from MPH!

Okay, let’s practice your best ö: just say ö together with the recording.

Remember to pay attention to all your cues: tongue touches the front teeth and upper molars, really rounded lips, think of saying German /e/.

A quick review of German /e/

- A few weeks ago, you practiced saying the German /e/ without gliding into a second vowel. This /e/ sound will be the basis of producing ö.
- Here is a reminder on how to pronounce German /e/ without sounding very accented: the most important thing is to produce only one vowel, when you say /e/.
  - You achieve this by concentrating on not moving your mouth after beginning to produce the vowel.
  - Begin saying /e/ the English way as in gate but FREEZE at the beginning of the vowel.
  - That means do not move your mouth or tongue. Your lips should be spread and your cheeks tense.
  - Keep this position and then directly produce /ö/ without gliding into another vowel first.
  - This is what you should sound like:

Here is what is happening in your mouth when you produce ö:

- The tip of your tongue touches the lower front teeth
  - It can touch your teeth very lightly, but that’s the region it should be in
- The sides of the front of the tongue are in contact with the inner sides of your upper molars
  - Imagine your tongue as being rather flat and spread out than long and thin
- Your lips are really rounded.
  - When you say German /e/, your lips are spread.
  - Remember: the only difference between saying German /e/ and ö is rounding your lips!

Let’s practice ö in words...

- First, we’ll practice saying the ö-sound in the beginning of German words.
  - The advantage of this is that you have time to bring your lips in the correct position before you have to start saying the words.
- Here we go! Listen to the recording and repeat after the native speaker!
• Österreich
Austria

• Okay, if that worked out all right, you are ready for the next step:

• Now try saying these words containing an ‘ö’:

• (der) Hörsaal
lecture hall

• (das) Wörterbuch
dictionary

• (die=Pl, der=Sing) Söhne
sons

• schwören
to swear
* (die=Pl, der=Sg) Vögel
  birds

* tödlich
  deadly

* schön
  beautiful

* (die) Höhle
  cave

* lösen
  to solve

* Well done!

* As always, you’ll get a chance to practice all sounds again on the next slide. Practice them until you feel confident saying them for the recording.

**Practice Review**

- ÖI
- Öfen
- Österreich
- Hörsaal
- Wörterbuch
- Söhne

- schwören
- Vögel
- tödlich
- schön
- Höhle
- lösen

- Ok, now open your audacity program and record yourself saying the words from the list.

- The entire list will be on the next slide.
Record yourself saying these words:

- Öl
- Öfen
- Österreich
- Hörsaal
- Wörterbuch
- Söhne
- schwören
- Vögel
- tödlich
- schön
- Höhle
- lösen

- Now, please stop your recording and 'export' it in order to save it as a .wav file.

- Then, please label your file with the following:
  - your PSU Username (xyz123)
  - the number „1“
  - the number „1“
- Example: xyz123_11_1.wav

- Finally, upload your recording to this week's dropbox on ANGEL.

- Das war super! Sehr gut!!

- You are done for today. ☺
Appendix C: Handout Perception Pretest

PSU User ID (xyz123): ____________

Listening Practice

Exercise 1: Recognizing German vs. English pronunciation

Please listen to both versions of the target word and indicate in the table below if the first (#1) or the second (#2) word you heard was spoken with Standard German Pronunciation.

<table>
<thead>
<tr>
<th>Target word</th>
<th>Choice #1</th>
<th>Choice #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice Item: Ball</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Löwe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bühne</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wiegen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ebene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheibe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Möbel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sauer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Braut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Batterie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lachen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zahn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>schlicht</td>
<td></td>
<td></td>
</tr>
<tr>
<td>schneien</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hauch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beten</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lokal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wohl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tüte</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exercise 2: Recognizing German sounds

You will hear words that don’t actually exist in German. (They are just strings of letters that could be words in German and can be pronounced, but that have no meaning). Listen for the targeted sound written on top of each slide and write down whether word 1 or word 2 contains this sound.

<table>
<thead>
<tr>
<th>Item #</th>
<th>Target sound</th>
<th>Choice #1</th>
<th>Choice #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>ä</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>German r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ü</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ei</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ö</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ei</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>ö</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>German r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>ch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>ü</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>ch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>ie</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix D: Production Task Word Reading Items

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ofen</td>
<td>1. übel</td>
<td>1. Biene</td>
</tr>
<tr>
<td>2. wohnen</td>
<td>2. aktiv</td>
<td>2. Graf</td>
</tr>
<tr>
<td>5. Vampir</td>
<td>5. oben</td>
<td>5. Bogen</td>
</tr>
<tr>
<td>11. fegen</td>
<td>11. Hof</td>
<td>11. liberal</td>
</tr>
<tr>
<td>15. Flöhe</td>
<td>15. Süden</td>
<td>15. Million</td>
</tr>
<tr>
<td>22. mieten</td>
<td>22. Ring</td>
<td>22. Biber</td>
</tr>
<tr>
<td>23. glühen</td>
<td>23. flach</td>
<td>23. Fluch</td>
</tr>
<tr>
<td>25. Los</td>
<td>25. Löhne</td>
<td>25. bügeln</td>
</tr>
</tbody>
</table>
Appendix E: Production Task Reading Paragraphs

Key:
- Novel words highlighted in yellow. (= words that are not in Chapters 1-5 and not included in the iCPR trainings)
- Names are only highlighted if they contain a target sound and if that target sound is the only token in the reading paragraph.
- All names are standard German names that raters should recognize as such.
- All paragraphs contain items for at least 10 out of 13 of the targeted sounds.

<table>
<thead>
<tr>
<th></th>
<th>[a]</th>
<th>[e]</th>
<th>&lt;z&gt;</th>
<th>[x]</th>
<th>[ɛ:]</th>
<th>[y], [y]</th>
<th>&lt;ie&gt;</th>
<th>&lt;ei&gt;</th>
<th>[o], [œ]</th>
<th>[ɔ:]</th>
<th>&lt;v&gt;</th>
<th>stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Delfin</td>
</tr>
<tr>
<td>#2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>aktiv</td>
</tr>
<tr>
<td>#3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Muslim</td>
</tr>
<tr>
<td>#4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Person</td>
</tr>
<tr>
<td>#5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Natur</td>
</tr>
<tr>
<td>#6</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Athlet</td>
</tr>
</tbody>
</table>

Reading: Paragraph #1:

Reading: Paragraph #2:

Reading: Paragraph #3:

Reading: Paragraph #4:

Reading: Paragraph #5:

Reading: Paragraph #6:
Appendix F: Sample of Production Test PowerPoint Slides

Introduction
- Today, we’ll start to assess your current level of German pronunciation.
- You’ll record several words and sentences today and then again at the end of the semester. That way, we can tell you how much you improved or what still causes difficulties for you.
- Just do your best for the recordings today, but don’t despair if you have problems with some of the words! (Most words will be unknown to you at this point, but you can find a picture illustrating what they mean on each slide).

Instructions
- Please read the German words and sentences out loud and record yourself using your Audacity Software.
- During the recording:
  - Make sure the room is quiet
  - No televisions, loud roommates, etc.
  - Avoid typing on the keyboard
  - If you still want to look up a word, please don’t do it at the same time that you are speaking.
  - Try to speak as clearly as possible
  - Minimize background noise & static from your computer’s microphone

Instructions for reading
- You’ll have to insert each individual word you’re supposed to say in a sentence. The sentence is always the same and will appear on top of each slide:
  - It is: „Ich habe das Wort XX gesagt“.
    - XX of course stands for the word you’ll see on the slide.
    - It is always the same sentence, only the targeted word will change!
    - The sentence simply translates to „I said the word XX”.
- Here are two practice items:

Practice Item 1
- You would see:

- You would say out loud:
  - „Ich habe das Wort Stift gesagt“.

- Are you recording yourself?
- Then let’s begin with the first 25 words:
Ich habe das Wort XX gesagt:

- Ofen

Awesome! Now here comes the paragraph reading task:

Instructions for paragraph readings

- On the next page, you will see a short paragraph in English. Just read through it, so that you know what it is about.
  - You can leave the audacity recording program running while you read, even if you don’t talk...
- Afterwards, you will see a German translation of the paragraph.
- Please read the German translation out loud!
  - You do not need to read the English paragraph out loud!
- The pictures are just there to help you remember what the paragraph is about because there is some new vocabulary in the German paragraphs.

This is my friend Valko. He is from Norway and he is a sincere person. Valko likes Switzerland and sailing.
He has a girlfriend. She works as a hairdresser and she goes to veterinary school. She wants to be a vet. She has a dolphin and a mouse.

Pictures omitted
for copyright reasons

- On the next slide, you will see a German translation of the paragraph.

Please read the German text out loud.

Das ist mein Freund Valko. Er kommt aus Norwegen und er ist aufrichtig. Valko mag die Schweiz und Segeln.
Er hat eine Freundin. Sie arbeitet als Frisörin und studiert Medizin. Sie will Tierarzt werden. Sie hat einen Delfin und eine Wühlmaus.

Pictures omitted
for copyright reasons

- And here is a second paragraph. Again, first in English, so that you know what it is about, then in German to read out loud.

This is my aunt Nora and my uncle Dietrich. They live in Bremen. Dietrich has a big belly, but he is an active person. Nora plays the guitar in a band. Her sister lives in Europe too. She has many animals: she has a goat, a parrot, a seagull and a howler monkey.

Pictures omitted
for copyright reasons
• On the next slide, you will see a German translation of the paragraph.

• Please read the German text out loud.


Pictures omitted for copyright reasons

• Great job!

• Please stop your recording now and save it. Remember: in order to save a file in Audacity, you have to go to „File“ → „Export“ (not “save” but “export”)

• Please label your file with your PSU Username and the number „1“
  * Example: xyz123_1.wav

• Now, we’re going back to saying individual words in a sentence.

• Please start recording again with Audacity in a new Audacity window:
  * If you just click „record“ again in the old window, it just adds to the old recording. But we want a new recording to cut file size down...

• Here are 25 more words:

Ich habe das Wort XX gesagt:

• übel

You’re done! Super gemacht!!!

• Please stop your last recording now and export it as a .wav file.

• Please label your file with your PSU Username and the number „3“
  * Example: xyz123_3.wav

• Then, please upload all three files to the dropbox on Angel called „Dropbox Week 2 - Assessment“.
  * Don’t worry if the upload takes a while, your files are probably rather big. Be patient ☺️.
Appendix G: Language Background Questionnaire

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 visual and/or hearing problems (either corrected or uncorrected)?
   □ No
   □ Yes [Please explain: ____________________________]

4. Native Country/Countries (Please check all that apply.)
   □ United States
   □ Other [Please specify: ____________________________]

5. Native Language(s) (Please check all that apply.)
   □ English
   □ Other [Please specify: ____________________________]

6. Language(s) spoken at home. (Please check all that apply.)
   □ English
   □ Spanish
   □ German
   □ Chinese
   □ Other [Please specify: ____________________________]

Part II

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

7. Prior to taking this GER 001 class, have you studied any second language(s)?
   □ No
   □ Yes → If yes, which language(s)?

8. If you studied any second language(s) (including German) before college, please check all of the following that apply and indicate the starting age and length of study for any second language(s) learned before college.
   □ 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? ________
9. Have you studied any second language(s) (including German prior to this GER 001 class) in college?
   □ No
   □ Yes → If yes, which language(s)? __________________________________________

   For how long? Language 1:  Language 2:
   □ Less than one semester  □ Less than one semester
   □ 1-2 semesters  □ 1-2 semesters
   □ 3-4 semesters  □ 3-4 semesters
   □ 5-6 semesters  □ 5-6 semesters
   □ 7-8 semesters  □ 7-8 semesters
   □ 8+ semesters  □ 8+ semesters

10. Are you currently taking any language courses besides GER 001?
    □ No
    □ Yes → If yes, which course(s)? __________________________________________

11. Have you studied and/or lived/served abroad? If you were not born in the United States, please include your stay(s) in the United States in the table below.
    □ Yes
    □ No

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

    | Country | Approx. dates or your age at the time | Length of stay | Language(s) spoken (mostly English or local language?) |
    |---------|--------------------------------------|----------------|------------------------------------------------------|
    |         |                                      |                |                                                      |
    |         |                                      |                |                                                      |
    |         |                                      |                |                                                      |
    |         |                                      |                |                                                      |

12. 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: __________________________________________]
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.

16. Your reading proficiency in German. (1 = not literate and 10 = very literate)
   1 2 3 4 5 6 7 8 9 10

17. Your writing proficiency in German. (1 = not literate and 10 = very literate)
   1 2 3 4 5 6 7 8 9 10

18. Your speaking ability in German. (1 = not fluent and 10 = very fluent)
   1 2 3 4 5 6 7 8 9 10

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

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.

20. Your reading proficiency in English. (1 = not literate and 10 = very literate)
   1 2 3 4 5 6 7 8 9 10

21. Your writing proficiency in English. (1 = not literate and 10 = very literate)
   1 2 3 4 5 6 7 8 9 10

22. Your speaking ability in English. (1 = not fluent and 10 = very fluent)
   1 2 3 4 5 6 7 8 9 10

23. Your speech comprehension ability in English. (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! ©
Appendix H: Participant Exit Questionnaire – iCPR Group

PSU User ID (xyz123): ___________

Debrief on Pronunciation Homework

Scale: 1 = strongly disagree ☹ to 7 = strongly agree ☺

1. I enjoyed working on the pronunciation homework.
   1  2  3  4  5  6  7

2. I feel like I learned something from the pronunciation homework.
   1  2  3  4  5  6  7

3. I feel like my pronunciation of German improved through the homework exercises.
   1  2  3  4  5  6  7

4. I would have felt more comfortable to practice pronunciation in-class rather than as homework (even if that meant the same amount of homework because I would have received more grammar exercises as homework then).
   1  2  3  4  5  6  7

5. Any technological problems that I encountered (e.g. downloading the PPT slides, recording myself etc.) were worth the overall benefit.
   1  2  3  4  5  6  7

6. Not only did the pronunciation homework improve my pronunciation but it also helped review some vocabulary.
   1  2  3  4  5  6  7

7. I would have taken the pronunciation homework more seriously and would have benefited more from it if I had received teacher feedback throughout the semester.
   1  2  3  4  5  6  7
1. How much time, on average, did you spend on each pronunciation homework?

<table>
<thead>
<tr>
<th></th>
<th>5-10 Min</th>
<th>10-15 Min</th>
<th>15-20 Min</th>
<th>20-25 Min</th>
</tr>
</thead>
</table>

2. How often, on average, did you listen to the native speaker recording of each word?

<table>
<thead>
<tr>
<th></th>
<th>1-2 times</th>
<th>2-3 times</th>
<th>3-5 times</th>
<th>5-7 times</th>
<th>more than 7 times</th>
</tr>
</thead>
</table>

3. How often, on average, did you practice saying each word before you recorded yourself on all the words?

<table>
<thead>
<tr>
<th></th>
<th>1-2 times</th>
<th>2-3 times</th>
<th>3-5 times</th>
<th>5-7 times</th>
<th>more than 7 times</th>
</tr>
</thead>
</table>

4. Did you always practice saying each word until you were completely happy with your own pronunciation before you recorded yourself?

<table>
<thead>
<tr>
<th></th>
<th>Always</th>
<th>Usually yes</th>
<th>About ( \frac{1}{2} ) the time</th>
<th>Usually not</th>
<th>Never</th>
</tr>
</thead>
</table>

5. There was no control of whether you completed the listening exercises (= the first pronunciation homework of the week). Honestly, how many of the 10 listening exercise homework units did you complete?

<table>
<thead>
<tr>
<th></th>
<th>All 10</th>
<th>9 of 10</th>
<th>8 of 10</th>
<th>7/6 of 10</th>
<th>5/4 of 10</th>
<th>3/2 of 10</th>
<th>1/0 of 10</th>
</tr>
</thead>
</table>

6. Did you enjoy the listening exercises or the speaking exercises more?

<table>
<thead>
<tr>
<th></th>
<th>Listening</th>
<th>Speaking</th>
<th>I enjoyed both equally</th>
</tr>
</thead>
</table>

7. Do you think that 3 days of homework practice on each pronunciation problem was enough to help you or would you have preferred more or less practice?

<table>
<thead>
<tr>
<th></th>
<th>More practice</th>
<th>It was a good amount</th>
<th>Less practice</th>
</tr>
</thead>
</table>
1. What did you like about the pronunciation homework exercises?


2. What did you not like about the pronunciation homework exercises?


3. Do you have any suggestions for future improvements of the pronunciation exercises? If so, what are they?


4. What, if any, are the areas of pronunciation not treated in the homework exercises that you would have liked to work on and improve?


5. Any final comments?


Thank you for your participation 😊
Appendix I: Participant Exit Questionnaire – In-Class Group

Debrief on Pronunciation Exercises

Scale: 1 = strongly disagree ☐ to 7 = strongly agree ☐

1. I enjoyed working on the pronunciation exercises in class.
   1  2  3  4  5  6  7

2. I feel like I learned something from the pronunciation exercises in class.
   1  2  3  4  5  6  7

3. I feel like my pronunciation of German improved through the exercises.
   1  2  3  4  5  6  7

4. I would have felt more comfortable practicing pronunciation as homework assignments.
   1  2  3  4  5  6  7

   If yes: Why?
   ☐ I would have felt less singled out in front of the class.
   ☐ I could have taken more time to focus on it.
   ☐ Other: ____________________________

5. Not only did the pronunciation exercises improve my pronunciation but it also helped review some vocabulary.
   1  2  3  4  5  6  7

1. How much time, on average, did you spend on the pronunciation exercises in class per day (on days that you practiced pronunciation in class)?

   5-10 Min  10-15 Min  15-20 Min  20-25 Min
2. Did you enjoy the listening exercises or the speaking exercises more?

<table>
<thead>
<tr>
<th>Listening</th>
<th>Speaking</th>
<th>I enjoyed both equally</th>
</tr>
</thead>
</table>

3. Do you think that 3 days of in-class practice on each pronunciation problem was enough to help you or would you have preferred more or less practice?

<table>
<thead>
<tr>
<th>More practice</th>
<th>It was a good amount</th>
<th>Less practice</th>
</tr>
</thead>
</table>

1. What did you like about the pronunciation exercises?

2. What did you not like about the pronunciation exercises?

3. Do you have any suggestions for future improvements of the pronunciation exercises?

4. What, if any, are the areas of pronunciation *not* treated in the pronunciation exercises that you would have liked to work on and improve?

5. Any final comments?

Thank you for your participation 😊
### Appendix J: Summary of Findings – Exp. 1 – p-values, effect sizes

<table>
<thead>
<tr>
<th></th>
<th>Between-group comparison (at posttest)</th>
<th>Within-group comparison (pretest to posttest)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Comparison</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Control</td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness</td>
<td>.007</td>
<td>.142</td>
</tr>
<tr>
<td>Detection Task</td>
<td>(1.10)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>Sound Discrimination Task</td>
<td>.102</td>
<td>.240</td>
</tr>
<tr>
<td></td>
<td>(0.67)</td>
<td>(0.47)</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>.017</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>(0.88)</td>
<td>(0.93)</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>.133</td>
<td>.101</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>Accentedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(0.88)</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>.151</td>
<td>.166</td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
<td>(0.59)</td>
</tr>
</tbody>
</table>

Note: p-value (effect size Cohen’s d)
### Appendix K: Summary of Findings – Exp. 2 – $p$-values, effect sizes

<table>
<thead>
<tr>
<th></th>
<th>Between-group comparison (at posttest)</th>
<th>Within-group comparison (pretest to posttest)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment &gt;</td>
<td>Treatment</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td><strong>Perception</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accentedness</td>
<td>.028</td>
<td>.019</td>
</tr>
<tr>
<td>Detection Task</td>
<td>(0.94)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Sound Discrimination Task</td>
<td>.028</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.03)</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>.036</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>(0.80)</td>
<td>(1.02)</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>.024</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(1.09)</td>
</tr>
<tr>
<td>Accentedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word-level</td>
<td>.011</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(1.24)</td>
</tr>
<tr>
<td>Paragraph-level</td>
<td>.040</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>(0.76)</td>
<td>(0.77)</td>
</tr>
</tbody>
</table>

Note: $p$-value (effect size Cohen’s $d$)
VITA
Ines A. Martin

EDUCATION
2018  Ph.D. in German Applied Linguistics and Language Science, Pennsylvania State University
2012  Erweiterungsprüfung Deutsch als Fremdsprache, Philipps-Universität Marburg, Germany
2011  Erstes Staatsexamen für das Lehramt an Gymnasien, Philipps-Universität Marburg, Germany
2009  Study Abroad, Institut Linguistique Adenet, Montpellier, France
2008  Study Abroad, Juniata College, Huntingdon, PA

SELECTED PUBLICATIONS


SELECTED GRANTS AND AWARDS
2017  Dissertation Grant (RGSO) Semester Release, PSU, College of the Liberal Arts
2016-2017  Watz Dissertation Fellow Award, Center for Language Acquisition, Penn State
2016  TOEFL Small Grants for Doctoral Research in L2 Assessment, ETS
2016  Language Learning Dissertation Grant, Language Learning Journal
2014-2015  Superior Teaching and Research Award, PSU, College of the Liberal Arts
2012-2013  Max Kade Fellowship – 1 Year Teaching Release, Max Kade Foundation

SELECTED CONFERENCE PRESENTATIONS
2015  “L2 Pronunciation Training Boosts the Acquisition of L2 Grammar Among Classroom Learners.” Poster Presentation. The 10th International Symposium on Bilingualism.