The production and perception of L2 phonological vowel contrasts by L1 Italian and L1 Spanish learners of English

Claudia Concas

Note that some graphs/tables/images have been removed in order to comply with copyright restrictions.

MSc in Applied Linguistics for Language Teaching, 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION



1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].

3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.

4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.

5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Claudia Concas
Date:	26/10/2025

Abstract

This dissertation explores how first language (L1) Italian and Spanish adult EFL learners navigate the perception and production of challenging English vowel contrasts. It focuses on how L1 phonological systems, second language (L2) proficiency, and modality-specific factors shape learners' performance with specific vowels. A quasi-experimental study was conducted with 101 participants (71 Italian, 30 Spanish), who completed a picture-selection perception task targeting seven English vowels (/iː/, /ɪ/, /æ/, /ʌ/, /ɑː/, /ɒ/, /əʊ/), and a read-aloud production task on four vowels (/iː/, /ɪ/, /æ/, /ɑː/). Drawing on contrastive and error analyses, the study assessed whether predicted areas of difficulty aligned with actual learner performance.

The findings revealed that /ɑː/ and /ɒ/ were the least accurately perceived vowels, whilst /ɪ/ and /ɑː/ posed the greatest challenges in production. Perception and production scores were not significantly different overall, but item-level patterns highlighted important modality-specific effects. Learners with higher receptive vocabulary (LexTALE) scores tended to perform better in both tasks, although the relationship between proficiency and accuracy varied depending on the vowel. Whereas the production of /ɪ/ improved with increased proficiency, /ɑ/ remained problematic across different proficiency levels. No statistically significant differences were found between Italian and Spanish groups in either modality; however, this may be partly due to unequal sample sizes (71 vs 30).

By combining contrastive and error analysis, this study validates the predictive power of phonological comparison while underscoring the need for empirical validation. It also highlights the importance of examining perception and production jointly through contrast-specific analyses. Pedagogical implications include the need for explicit instruction on problematic contrasts and the integration of both perceptual and articulatory training in EFL curricula for Romance L1 learners, in particular, Italian and Spanish.

Acknowledgements

Among all the paths I had imagined for myself in life, I would never have thought to be blessed with the opportunity of studying at the University of Oxford. I feel extremely fortunate and grateful for this wonderful opportunity, which has allowed me to meet truly impressive people and to be taught by world-leading academics who are also remarkable human beings.

I would like to express my deepest gratitude to my supervisor, Dr Faidra Faitaki, without whose guidance and enthusiasm this dissertation would not have been possible. Thank you, Faidra, for never letting me be disheartened: your support, encouragement, and belief in my work have been a constant source of motivation. I aspire one day to have even half of your confidence and expertise.

I am also especially grateful to Dr Catherine Hamilton for her invaluable help in navigating the statistical journey required for this project. Thank you for welcoming me as an extra student on your course, and taking the time to guide me through the intricacies of R. You are, without a doubt, my "stats hero"!

I wish to thank all the brilliant academics on the ALLT course, Faidra, Hamish, Heath, Liz, and Victoria, whose teaching and enthusiasm sparked my passion for Applied Linguistics and inspired me to adopt a quantitative data collection approach and embark on experimental research. To my ALLT cohort, it was such a pleasure to meet and learn from such a diverse group of people from all over the world, each bringing something unique to our shared journey. Your perspectives, ideas, and support throughout this adventure made the ALLT environment enriching and intellectually stimulating!

Turning to my study, I am profoundly grateful to all my participants for dedicating their time and energy to my experiment. I am equally appreciative of those who, despite technical difficulties, made repeated attempts to complete the tasks. Your commitment is what enables research to move forward.

Finally, my heartfelt thanks go to those closest to me. To Cristiana, my number one fan, thank you for always being there to encourage me to give my best. To Andrea, for cheering me on even from afar, as if the miles between us did not exist. To Camilo, thank you for your unwavering support and for encouraging me through every high and low. And to my family, the foundation of everything I am, without whom none of this would have been possible: my mum, whose strength and determination inspire me daily, my sister Enrica, my aunts Marcella and Dorotea, my uncles Gigi and Massi, and my cousins Federico, Ludovica and Eleonora. These past two years have been an emotional rollercoaster, and your presence has been my anchor.

Table of Contents

List of tables 14

List of figures 15

List of abbreviations 16

1. Introduction 17

1.1 Italian 18

1.2 Spanish 19

1.3 Overview of the Dissertation 21

2. Literature Review 22

2.1 The Acquisition of the L2 Phonology: The Role of the L1 22

2.2 Perception vs Production: The Current Debate 23

2.3 Contrastive & Error Analyses: Strengths and Weaknesses 25

2.4 Contrastive Analysis 26

2.4.1 Vowels: Italian & Spanish vs English 26

2.4.2 Inventory size 27

2.4.3 Tense - Lax & Long – Short Distinctions 28

2.4.4 Diphthongs ____ 29

2.4.5 Consonants: Italian vs English ____ 30

2.4.6 Consonants: Spanish vs English ____ 33

2.4.7 Predicted Problematic Features ____ 34

2.5 Empirical Evidence from Previous Error Analysis Studies ____ 36

2.5.1 EA Studies with Italian Participants ____ 36

2.5.2 EA Studies with Spanish Participants ____ 37

2.5.3 Integrating CA Predictions with EA Findings ____ 39

2.6 Research Questions and Study Aims ____ 40

3. Methodology ____ 41

3.1 Methodological Approach ____ 41

3.2 Participants ____ 41

3.3 Data Collection Instruments ____ 42

3.3.1 Background Questionnaire ____ 42

3.3.2 LexTALE: Receptive Vocabulary Test ____ 42

3.3.3 Stimuli Selection ____ 43

3.3.4 Production Task: Read-aloud ____ 44

3.3.5 Perception Task ____ 44

3.4 Procedure ____ 45

3.5 Piloting and Technical Issues ____ 45

3.5.1 Technical Pilot ____ 45

3.5.2 Official Pilot ____ 46

3.6 Coding and Interrater Reliability ____ 47

3.7 Ethical Considerations ____ 47

3.8 Methodological Limitations ____ 48

4. Results ____ 50

4.1 Descriptive Statistics ____ 50

4.1.1 Overview ____ 50

4.1.2 LexTALE ____ 50

4.1.3 Read-Aloud Task ____ 51

4.1.4 Perception Task ____ 53

4.2 RQ1 Which L2 vowels pose the greatest difficulty for L1 Italian and L1 Spanish learners of English in (a) perception and (b) production? ____ 55

4.2.1 Perception Data – Repeated Measures ANOVA ____ 55

4.2.2 Perception Data – 2x7 Repeated Measures ANOVA ____ 56

4.2.3 Production Data – Repeated Measures ANOVA ____ 57

4.2.4 Production Data – 2x4 Repeated Measures ANOVA ____ 58

4.3 RQ2 Does proficiency affect which vowels learners find problematic in (a) perception and (b) production? ____ 60

4.3.1 Relationship Between Proficiency and Perception Accuracy ____ 60

4.3.2 Relationship Between Proficiency and Production Accuracy ____ 62

4.4 RQ3 Is the learners' perception of L2 vowels more accurate than their production of the same vowels? ____ 65

5. Discussion ____ 67

5.1 RQ1: Which L2 vowels pose the greatest difficulty for L1 Italian and L1 Spanish learners of English in (a) perception and (b) production? ____ 67

5.1.1 (a) Perception ____ 68

5.1.2 (b) Production ____ 69

5.2 RQ2: Does proficiency affect which vowels learners find problematic in (a) perception and (b) production? ____ 70

5.2.1 (a) Perception ____ *71*

5.2.2 (b) Production ____ *72*

5.3 RQ3. Is the learners' perception of L2 vowels more accurate than their production of the same vowels? ____ *72*

6. Conclusions ____ *74*

6.1 Implications for EFL Teaching ____ *74*

6.2 Limitations of This Study and Recommendations for Future Research ____ 75

References ____ *78*

Appendices ____ *93*

Appendix A: Dialects spoken by Italian participants ____ *93*

Appendix B: Participant regional and national origins ____ *95*

Appendix C: Demographics and Language Background Questionnaire ____ *96*

Appendix D: Sentences Used in the Production Task ____ 101

Appendix E: Perception Task: Vowel Contrasts, Stimuli, and Conditions ____ 103

Appendix F: Rating Instructions, Coding Template, and Participant Labels for the Production Task ____ 104

F1. Rating Instructions (as shared with raters) ____ 104

F2. Coding Template Structure ____ 105

F3. Participant Relabels 105

Appendix G: Observed Agreement Calculations for Interrater Reliability 108

Appendix H: Information Sheet and Consent Form 109

Appendix I: Ethics Approval Letter 115

Appendix J: Distribution of LexTALE scores (whole sample and by group) 117

Appendix K: Distribution of Read-Aloud Task accuracy scores per vowel (whole sample) 119

Appendix L: Distribution of Perception Task accuracy scores per vowel (whole sample) 122

Appendix M: Distribution of Read-Aloud Task accuracy scores per vowel in the Italian Group 126

Appendix N: Distribution of Read-Aloud Task accuracy scores per vowel in the Spanish Group 128

Appendix O: Distribution of Perception Task accuracy scores per vowel in the Italian Group 130

Appendix P: Distribution of Perception Task accuracy scores per vowel in the Spanish Group 134

Appendix Q: Boxplots of Read-Aloud Task Accuracy by Vowel and Group 138

Appendix R: Boxplots of Perception Task Accuracy by Vowel and Group 141

Appendix S: Point-and-Line Plot of Mean Perception Accuracy by Vowel (Whole Sample) 145

Appendix T: Point-and-Line Plot of Mean Perception Accuracy by Vowel and Group 146

Appendix U: Point-and-Line Plot of Mean Production Accuracy by Vowel (Whole Sample) 147

Appendix V: Point-and-Line Plot of Mean Production Accuracy by Vowel and Group 148

Appendix W: Normality checks for LexTALE and vowel accuracy scores in the Perception and Production Tasks 149

W1. Visual Checks: Histograms and Q-Q Plots 149

W2. Shapiro-Wilk Test: LexTALE Scores 157

W3. Shapiro-Wilk Test: Overall Accuracy Scores 157

W4. Shapiro-Wilk Test: Perception Task – Individual Vowels 157

W5 Shapiro-Wilk Test: Production Task – Individual Vowels 158

Appendix X: Correlation Analyses Scatterplots for Perception and Production Tasks 159

Appendix Y: Regression Scatterplots for Perception and Production Tasks 166

Y1. Perception Task 166

Y2. Production Task _____ **168**

Appendix Z: Q-Q Plot Assessing the Normality of the Difference Scores Between Perception and Production Accuracy _____ **170**

List of tables

Table 1. Italian consonant phonemes

Table 2. English consonant phonemes

Table 3. Spanish consonant phonemes

Table 4. Predicted problematic phonological segments for L1 Spanish and L1 Italian learners of English

Table 5. The six most common vowel and consonant misformations and their average error rate

Table 6. Summary of Participant Demographics

Table 7. Descriptive Statistics for the LexTALE Scores by Group

Table 8. Descriptive Statistics per Vowel for the whole participant sample in the Read-Aloud Task

Table 9. Descriptive Statistics per Vowel for the whole participant sample in the Perception Task

Table 10. Summary of Simple Linear Regression Results Predicting Vowel Perception Accuracy from LexTALE Scores

Table 11. Summary of Simple Linear Regression Results Predicting Vowel Production Accuracy from LexTALE Scores

List of figures

Figure 1. Geographical distribution of Italian regional varieties

Figure 2. Number and proportion of Spanish speakers around the world between 2000 and 2005

Figure 3. Italian vowel system

Figure 4. Spanish vowel system

Figure 5. English pure vowel system

Figure 6. Boxplot of LexTALE Scores for the whole participant sample

Figure 7. Violin Plot of Perception Accuracy Across Vowels for the Whole Participant Sample

Figure 8. Violin Plot of Perception Accuracy Across Vowels for the Italian and Spanish Learner Groups

Figure 9. Violin Plot of Production Accuracy Across Vowels for the Whole Participant Sample

Figure 10. Violin Plot of Production Accuracy Across Vowels for the Italian and Spanish Learner Groups

Figure 11. Regression Scatterplot for LexTALE and Overall Perception Accuracy Scores

Figure 12. Regression Scatterplot for LexTALE and Overall Production Accuracy Scores

Figure 13. Mean Perception and Production Accuracy Scores Across All Vowels with 95% Confidence Intervals

List of abbreviations

ANOVA	Analysis of Variance
CA	Contrastive Analysis
CEFR	Common European Framework of Reference
CPH	Critical Period Hypothesis
EA	Error Analysis
EF	Education First
EFL	English as a Foreign Language
EPI	English Proficiency Index
ESL	English as a Second Language
GA	General American
HVPT	High Variability Perception Training
IPA	International Phonetic Alphabet
L1	First Language
L2	Second Language
L2LP	L2 Linguistic Perception Model
NAE	North American English
PAM	Perceptual Assimilation Model
PT	Perception Task
RAT	Read-Aloud Task
SAA	Speech Accent Archives
SLA	Second Language Acquisition
SLM	Speech Learning Model
SSBE	Southern Standard British English

1. Introduction

Pronunciation and listening comprehension are consistently reported as two of the most challenging skills to master for second language (L2) English learners (Derwing & Munro, 2015; Field, 2019). Yet, pronunciation and perceptual training continue to receive limited attention in many EFL and ESL contexts. This is often due to institutional constraints, such as limited classroom time for English instruction, which tends to prioritise grammar and vocabulary over speaking and listening, as well as a lack of teacher training (Munro, 2018).

This study focuses on first language (L1) Italian and Spanish learners of English, a choice motivated by both pedagogical relevance and linguistic interest. These two populations account for a large proportion of English language learners across Europe. According to the European Commission (2023), over 90% of students in both Italy and Spain study English from primary through secondary school. Yet, the English proficiency levels of these groups remain relatively low. The Education First (EF) English Proficiency Index (EPI, 2024) ranks Italy 46th and Spain 36th out of 116 countries globally, placing both countries in the moderate proficiency band, behind most European nations. Despite extensive exposure to English through formal education, persistent difficulties in L2 communication point to the need for a more focused investigation of L2 speech perception and production in these learner groups (Stan, 2024; Vargas Jiménez, 2015).

It is widely acknowledged that all L2 learners face difficulties in acquiring a new phonological system due to the influence of their L1 (Flege, 1991). For Spanish and Italian speakers, the English vowel system poses particular challenges, with its higher number of vowel items and diphthongs. This can lead to systematic perception and production errors, which, though common, may reduce intelligibility and increase the risk of communication breakdowns (Jarosz, 2023; Munro, 2008). While the importance of pronunciation is recognised, its instruction remains underprioritised. This study investigates both perception and production, two key modalities of phonological competence that are often assumed to develop in parallel but frequently do not. Early models (Best & Tyler, 2007; Escudero, 2009; Flege, 1995) emphasised the primacy of perception, while more recent studies (De Leeuw et al., 2021; Kartushina & Frauenfelder, 2014) suggest modality mismatches, where learners may correctly identify sounds but fail to produce them. Addressing both dimensions is crucial for capturing the full scope of learners' phonological development.

The present study aims to investigate how L1 Italian and Spanish speakers perceive and produce challenging L2 English vowels. From a linguistic perspective, Italian and Spanish are typologically similar Romance languages; yet even closely related L1s can result in distinct L2 phonological

acquisition outcomes (Serratrice et al., 2012). Comparing learners with closely related L1s offers a unique lens to explore cross-linguistic influence. It also helps identify which English contrasts are especially susceptible to L1-transfer effects. To achieve this aim, the present dissertation draws on Contrastive Analysis (CA) and Error Analysis (EA). In CA, the systems of the L1 and the L2 are compared to identify differences which may cause learning difficulties (Rustipa, 2011), while EA examines the errors produced by learners (Lennon, 2008). Therefore, the study contributes to a more granular understanding of L1 influence and learners' needs and offers insights to enhance pronunciation teaching and curriculum development. The remainder of this chapter provides a brief overview of the two languages development and characteristics, followed by an overview of the dissertation.

1.1 Italian

To date, Italian is spoken by approximately 67 million native speakers and 18 million L2 users (World Population Review, 2024). Bertinetto and Loporcaro (2005) define Italian as one of the most *conservative* Romance languages; that is, the Romance that remains closest to its source language, Latin, especially in terms of phonetic features. It is the official language of Italy, San Marino and the Vatican City, as well as one of the official languages of Switzerland, where Italian is spoken mainly in the Ticino and Graubünden cantons (Bertinetto & Loporcaro, 2005). The standard variety of Italian is based on the dialect spoken in Florence (Tuscany), where the Italian language was born, and is mainly used by a restricted elite of speakers, including journalists and stage actors.

Italian phonology is characterised by great variation both at the segmental and prosodic levels. Particularly relevant is the heterogeneity with which phonological segments are realised across different geographical regions. One reason for such diversity is the relatively recent birth of Italian as a spoken language in the peninsula. After Italy’s unification in 1861, Italian remained mostly a written standard while local dialects were predominant in everyday communication. Only in the 1950s/60s did Standard Italian begin to spread more broadly due to the influence of mass media and education (Berruto et al., 2018).

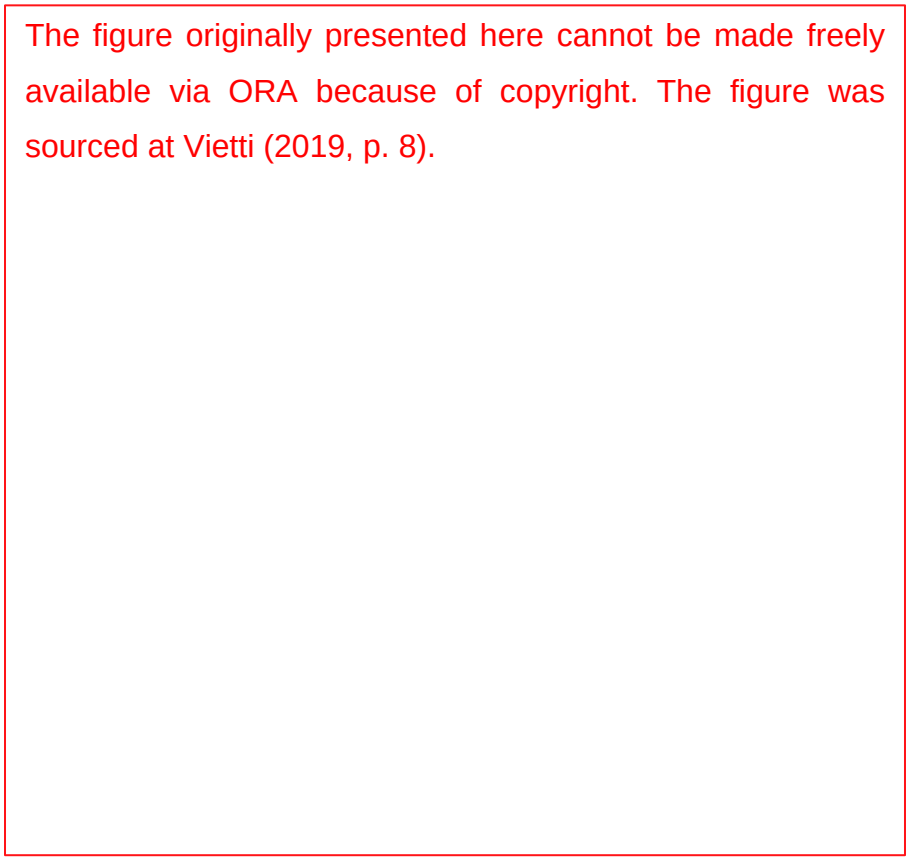
Yet, the main cause underlying such phonological variation in Italian is the long-standing contact of the standard variety with the primary dialect specific to each geographical region (De Mauro, 2014). This contact has given rise to regional Italian varieties, forms of Italian shaped by the phonological features of each area’s native dialect (Vietti, 2019). The result of this language contact led to the formation of well-established regional varieties, which can be divided into four areas: central, southern,

northern, and Sardinia. Sociolinguistic variation adds another layer of complexity, producing a continuum from the standard to more localised forms.

For practical reasons, namely, the lack of a unified phonological system across regional varieties and the need for consistency in the CA, this dissertation adopts Standard Italian as the reference variety. However, given the sensitivity of phonology as a language area, reference to regional varieties will be made throughout the analysis when necessary.

Figure 1

Geographical distribution of Italian regional varieties



Note. Source: Vietti (2019, p. 8).

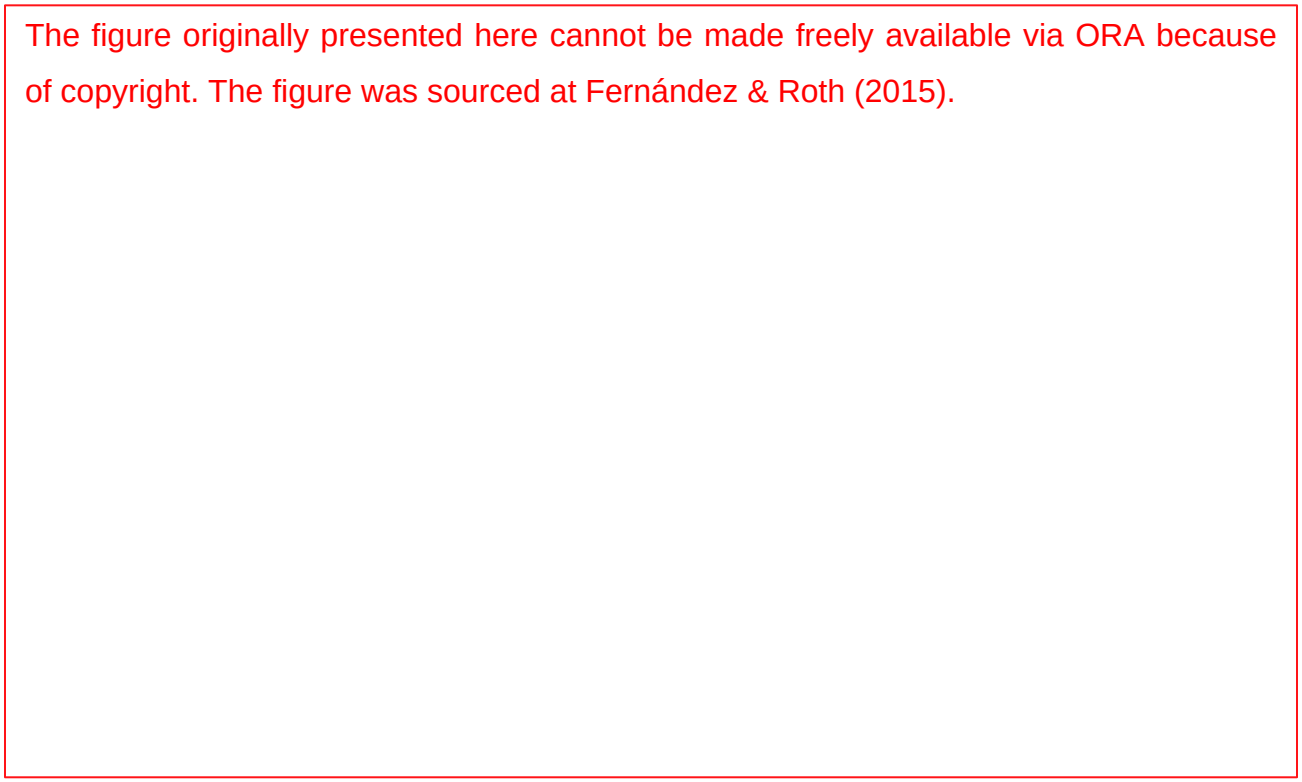
1.2 Spanish

Spanish is the second most spoken language in the world by the number of native speakers. According to the Instituto Cervantes (2023), almost 500 million people speak Spanish as their mother tongue, and when including L2 speakers, the total number corresponds to approximately 599 million. Spanish is the official language of 21 countries in total. Additionally, Spanish is the second most spoken language

in the US and various other regions worldwide due to the spread of Hispanic communities (World Population Review, 2024).

Figure 2

Number and proportion of Spanish speakers around the world between 2000 and 2005



Note. Source: Fernández & Roth (2015).

Considering the extensive geographical distribution of Spanish, as displayed in Figure 2, the existence of several dialects and varieties of this Romance language does not come as a surprise. Variation in the Spanish-speaking world particularly affects the sphere of phonology, leading to the formation of different dialectal groups. One of the main factors that helps distinguish between different Spanish varieties is whether they maintain the phonological distinction between the voiceless interdental and alveolar fricatives: /s/ and /θ/, respectively (Lipski, 2013). If they do, the words *casa*, 'house', and *caza*, 'hunting', would constitute a minimal pair, being pronounced as [ka.sa] and [ka.θa], respectively. Such phonological opposition is maintained in all of the varieties of Peninsular Spanish, except for those spoken in Andalusia and in the Canary Islands (Lipski, 2013). The latter varieties, along with the dialects of Spanish spoken in the Americas, lose the /θ/ - /s/ distinction in favour of the latter phoneme; a phenomenon known as *seseo* (Núñez-Méndez, 2022). Therefore, in this dialectal

area where /θ/ is a mere allophone, the words *casa* and *caza* are homophones, as they share the same pronunciation: [ˈka.sa].

Another phonological opposition which is either lost or maintained depending on the variety is that between the palatal lateral approximant /ʎ/ and the voiced palatal fricative /ʝ/, exemplified by the minimal pair *calló* [kaˈʎo], 'he/she became silent', and *cayó* [kaˈʝo], 'he/she fell'. The loss of this opposition in favour of /ʝ/ is known as *yeísmo* and is becoming increasingly more common, especially among younger generations (Rost Bagudanch, 2017). Variation also affects the pronunciation of /s/ in coda position, being aspirated or even omitted completely in some varieties (Campos-Astorkiza, 2013). The phoneme /x/ also varies; it is realised as uvular [χ] (e.g., *junta* [ˈχunta], 'meeting') in Peninsular Spanish and as [h] in Andalusia, the Canary Islands, Central America and the Caribbean; e.g. *caja* [ˈkaha], 'box' (Campos-Astorkiza, 2013).

1.3 Overview of the Dissertation

Having presented the phonological characteristics of the focal languages, the remainder of this dissertation is structured as follows. Chapter 2 will review relevant literature in the field of L2 phonology acquisition. It will also present a CA comparing the learners' L1s (Italian and Spanish) and English, discuss evidence from previous error analysis studies on Italian and Spanish learners and outline the study's research questions. Chapter 3 will outline the study's methodology, including participant selection, task design, and analysis procedures, and Chapter 4 will then present the results of the descriptive and inferential statistical analyses conducted on the data. Chapter 5 will discuss the findings in relation to theoretical frameworks and pedagogical concerns. Finally, conclusions will be drawn in Chapter 6, which will acknowledge the study's limitations and offer recommendations for future research and pedagogy.

2. Literature Review

2.1 The Acquisition of the L2 Phonology: The Role of the L1

It is now well established in the field of Second Language Acquisition (SLA) that a learner's L1 can have a considerable impact on the acquisition of an L2 system (Flege, 1995; Jarosz, 2023; Munro, 2008). This influence, often referred to as L1 transfer, can manifest either positively or negatively, depending on the degree of similarity between the two systems. L1 transfer can lead to either facilitation or interference; when L1 and L2 structures are similar, it may facilitate learning, but when they differ, it may result in negative transfer or interference (Bardovi-Harlig & Sprouse, 2024). The latter type of transfer is likely to contribute to persistent perceptual and pronunciation errors (Flege, 1991).

L1 transfer is particularly salient in the domain of pronunciation, where L2 learners often display segmental and suprasegmental features characteristic of their L1. Foreign accent, for instance, is a common outcome of such influence, which has been widely documented and is frequently attributed to persistent L1-based phonological patterns (Derwing & Munro, 2015). These errors are not random but are systematically shaped by the learner's phonological repertoire and the phonological properties of the L2 input. One relevant factor in this regard is the functional load of a phoneme; that is, the extent to which a given sound is used to distinguish between words (Meyerstein, 2019). For example, the phonemes /ɪ/ and /iː/ carry a high functional load in English because they differentiate between several minimal pairs, such as *bit* and *beat*. L2 phonemes that are absent from the learner's L1 but carry a high functional load in the target language are more likely to pose challenges. These predictions can be formalised to some extent through systematically comparing the L1 and L2 phonological systems.

This relationship between the L1 and the L2 has also been addressed within the framework of interlanguage theory, first introduced by Selinker (2015). According to this view, learners construct a dynamic linguistic system that is neither entirely L1 nor fully target-like, but rather an evolving grammar. Within this system, errors are not simply signs of failure but evidence of active hypothesis testing about the L2. As Corder (1967) argues, errors provide insights into the learner's developing linguistic competence and should be viewed as an integral part of language learning, rather than as mere signs of deficiency. In the phonological domain, interlanguage manifests in consistent substitution patterns, simplification strategies, or misapplication of L1 phonotactic rules, patterns that can be empirically studied through error analysis (James, 2013).

One important factor that interacts with L1 transfer in the phonological domain is age (Singleton & Lengyel, 1995). The Critical Period Hypothesis (CPH) posits that there is a biologically determined window during which language acquisition occurs most naturally and effectively (Lenneberg, 1967). According to Lenneberg (1967), this period ends with the onset of puberty, after which the brain loses the plasticity necessary for successful language learning, making acquisition virtually impossible. Although the CPH was initially formulated for L1 acquisition, it was later extended to L2 learning. This hypothesis has since been widely investigated in the L2 domain, with particular regard to phonological development (Flege, 1991, 2019; Flege et al., 1995). One piece of evidence often cited in support of this view is that most adults who learn an L2 after childhood tend to retain a foreign accent (Abrahamsson & Hyltenstam, 2009; Scovel, 2000).

The age of 12 is often regarded as a threshold for achieving native-like ultimate attainment in phonology (Asher & García, 1969; Balari & Lorenzo, 2015; Oyama, 1976). However, this threshold should not be interpreted as a clear-cut dividing line, as age effects tend to be gradual. Flege et al. (1995) found a continuous decline in perceived foreign accent as age of arrival increased among native Italian speakers in Canada. In general, the older the learner, the lower the likelihood of fully acquiring the L2 phonological system. This has implications not only for how L2 phonology is acquired, but also for how it is taught. The focus of the present dissertation on adult learners is, therefore, motivated by the observation that adults are known to struggle more with the acquisition of accurate L2 phonology (Flege & MacKay, 2004; Moyer, 2007; Piske et al, 2001).

2.2 Perception vs Production: The Current Debate

In the context of L2 phonological acquisition, the relationship between speech perception and production plays a pivotal role. These two skills are central to the development of intelligible and accurate spoken language, yet they are often treated separately in both research and pedagogy. Thomson (2022) operationalised speech perception as the listener's ability to decode phonetic input and recognise individual speech segments as intended by the speaker. By contrast, speech production concerns how speakers generate those segments during spoken utterances (Thomson, 2022). While these are distinct cognitive and physiological processes, the extent to which they are interconnected remains a subject of ongoing debate in SLA (Arjmandi & Behroozmand, 2024; Song & Eckman, 2024).

The well-established view that accurate perception is a necessary condition for accurate production is central to many models of L2 speech learning. One such model is Flege's (1995) Speech Learning Model (SLM), which posits that the degree of perceived similarity between L1 and L2 sounds affects the formation of new phonetic categories, with greater perceived similarity making accurate

production more difficult. Unfamiliar L2 sounds (i.e., those not assimilable to an L1 category) are more likely to result in the formation of new phonological representations. This perception-production link is also apparent in other models (Best, 1994; Best & Tyler, 2007; Escudero, 2009). Empirical evidence further supports this connection: numerous studies have found that auditory training targeting difficult L2 contrasts not only improved learners' perception but also led to gains in production (Sakai & Moorman, 2018; Thomson, 2011; Thomson & Derwing, 2015).

However, recent empirical research has challenged the view that perception necessarily precedes production. A growing number of studies have shown that the two skills do not always develop in tandem and may be only weakly correlated (De Leeuw et al., 2021; Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011). For example, Kartushina and Frauenfelder (2014) found that French learners of English could perceive certain vowel contrasts accurately (e.g., /iː/-/ɪ/) but failed to produce them correctly. Similarly, De Leeuw et al. (2021) observed that Spanish-English bilinguals were capable of perceiving English sounds yet continued to produce them inaccurately, suggesting a dissociation between phonological discrimination and articulation.

Other studies have suggested the opposite development sequence; namely, that production can precede perception (Hattori & Iverson, 2010; Sakai, 2016). Hattori and Iverson (2010) examined Japanese learners' acquisition of the English /r/-/l/ contrast and found that while identification and production were moderately correlated, other perceptual abilities were weakly related to production. Likewise, Sakai (2016) demonstrated that Spanish-speaking learners of English who received production-only training showed significant improvements in the perception of the /i/-/ɪ/ contrast.

This variability across studies illustrates the mixed and sometimes contradictory nature of the empirical evidence. One important limitation of this body of research is its methodological and contextual heterogeneity; studies differ considerably in terms of participant populations (beginner vs advanced), language combinations, target phonological contrasts, and the types of training or testing methods used. While this diversity offers a broad perspective, it also complicates attempts to draw generalisable conclusions and weakens the overall strength of the evidence base.

Such findings indicate that perception and production may be governed by partly independent cognitive mechanisms, and that it is not always the case that L2 phonetic representations are translated into articulatory output (De Leeuw et al., 2021). This may be particularly true for some specific contrasts, which learners may perceive successfully but fail to produce correctly. To better understand this relationship, it is necessary to conduct item-level analyses. Although a few studies have begun to adopt item-level approaches, such analyses remain relatively underused in this field (Melnik-Leroy et

al., 2021; Song & Eckman, 2024). The present study addresses this gap by examining whether and to what extent L1 Italian and Spanish learners' perception of English vowels aligns with their production.

2.3 Contrastive & Error Analyses: Strengths and Weaknesses

Contrastive Analysis (CA) is a linguistic approach that compares a learner's L1 with the target language to predict learning difficulties (Rustipa, 2011). Though developed in the 1960s under behaviourist assumptions, CA remains a useful diagnostic tool in L2 teaching (Munro, 2018). CA posits that similarities between the L1 and the L2 facilitate acquisition, while differences can lead to errors. Initially, this assumption led educators to view errors as bad habits to be eliminated (Lennon, 2008). However, this view has since evolved; contemporary SLA research recognises errors as a natural and informative part of the learning process (Manirakiza et al., 2021; Nzerem & Bob, 2021).

CA began to face criticism from the late 1960s onwards for its limited predictive power (Corder, 1994; Wardhaugh, 1970). Many errors observed in learner data were not anticipated by L1–L2 contrasts, and some predicted difficulties did not lead to errors (James, 2013). Furthermore, the CA's overemphasis on language interference often led to overlooking other sources of difficulty, such as *developmental errors*, which occur regardless of the learners' L1 (Lennon, 2008). As a result, its explanatory power was deemed insufficient to account for the complexity of L2 acquisition.

In response, Error Analysis (EA) emerged as an alternative (Corder, 1967; Wardhaugh, 1970). Rather than predicting errors, EA focuses on identifying and categorising actual learner errors, thus offering insights into learners' interlanguage development. Errors are viewed as evidence of learners' active hypothesis testing about the L2 system. However, like CA, EA has its limitations: it lacks a comprehensive theory of acquisition, fails to account for individual learner differences, and does not distinguish between perception and production processes (Derwing & Munro, 2015). This last point is particularly relevant to phonology. As mentioned in Section 2.2, learners may accurately perceive a contrast without being able to produce it (De Leeuw et al., 2021). Failing to consider this dissociation risks misinterpreting the nature of phonological errors.

Combining CA and EA offers a more comprehensive picture of L2 phonological difficulties. CA helps anticipate areas of likely L1 influence, while EA validates or challenges these predictions empirically. This study adopts such a combined approach to evaluate the relevance of CA for identifying problematic vowel contrasts and to assess whether observed learner errors align with theoretical predictions. That is, the CA will be used as the starting point to evaluate its effectiveness as a method. Crucially, it will allow the researcher to tailor the experimental materials by focusing on problematic

contrasts based on systematic phonological comparison, thus validating the study's material development. This investigation also reflects on the continued relevance of CA by examining the connection between theory and practice.

2.4 Contrastive Analysis

Given the overlap between Italian and Spanish vowel systems, a joint CA will be conducted for vowels, while consonant systems will be analysed separately, given their greater divergence. This analysis aims to highlight predicted phonological challenges for each group of learners, which will then be compared against this study's error analysis findings.

2.4.1 Vowels: Italian & Spanish vs English

The vowel inventories of Italian and Spanish are almost equal: Spanish has only 5, /i, e, a, o, u/. Standard Italian has 7 pure vowels, /i, e, ɛ, a, ɔ, o, u/, as it distinguishes between mid-high and mid-low vowels. However, the opposition between tense and lax mid vowels in Italian varies depending on the region (Krämer, 2009). On the other hand, phonological variation across varieties of Spanish mostly concerns consonant production, rather than vowels (Anderson & Centeno, 2018). The two figures below display the vowel inventories of Standard Italian and Spanish, respectively.

Figure 3

Italian vowel system

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Vietti (2019, p. 23).

Note. Adapted from Vietti (2019, p. 23).

Figure 4

Spanish vowel system

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Campos-Astorkiza (2013, p. 90).

Note. Adapted from Campos-Astorkiza (2013, p. 90).

In contrast with the two Romance vowel systems, the English inventory consists of 12 pure vowels: 5 long, /iː, uː, ɑː, ɔː, ɜː/, and 7 short, /ɪ, e, ə, æ, o, ʌ, ʊ/ (Rogerson-Revell, 2018).

Figure 5

English pure vowel system

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Rogerson-Revell (2018, pp. 112-114).

Note. Source: Rogerson-Revell (2018, pp. 112-114).

2.4.2 Inventory size

Examining and comparing the vowel systems displayed above, the first difference that stands out is the size discrepancy between the 7- and 5-vowel inventories of Italian and Spanish, respectively, against the 12-vowel English system. Therefore, the first hurdle that Italian and Spanish learners of

English have to face is the larger size of the L2 phonological inventory. According to Iverson and Evans (2007), when the L1 inventory is smaller than the target language system, several L2 segments will be associated with a single L1 category. A larger size corresponds to a higher number of vowel phonemes that are absent in the learner's L1, making the acquisition process slower and more complex, compared to EFL learners with a larger-sized vowel inventory, e.g. German speakers learning English (Iverson & Evans, 2009).

When the L2 phoneme does not exist in the L1 inventory, the learner can either create a new category or assimilate it to the closest L1 category (Escudero et al., 2014). The latter assimilation process occurs, for instance, with the English long-short tense-lax contrast /iː/ - /ɪ/, which is absent in the phonologies of both Italian and Spanish, but usually assimilated with the closest L1 element /i/ (Escudero et al., 2014). Neither Spanish nor Italian has central vowels; therefore, these may be particularly complex to acquire and master for these two groups of learners. The central English vowels /ə, ʌ, ɜː/, for which no corresponding similar element exists in the two Romances, entail a greater learning challenge for these learners who will need to create new phonological categories to perceive and produce these phonemes correctly.

2.4.3 Tense - Lax & Long – Short Distinctions

The tense-lax and long-short oppositions do not carry any phonemic value in Spanish (G ómez González & Sánchez Roura, 2021) nor in Italian, where the tense/lax distinction between mid-vowels holds a mild functional load (Vietti, 2019). As far as vowel length is concerned, this is not contrastive in either of the two Romance languages or in English, although it is an integral part of many vowel contrasts in the latter language (e.g., *beat* /biːt/– *bit* /bɪt/, *caught* /kɔːt/ – *cot* /kɒt/, *pool* /puːl/ – *pull* /pʊl/) (Rogerson-Revell, 2018). In Italian, vowel length can be predicted based on the phonetic environment in which the vowel occurs. Vowels are usually long when they occur in an open and stressed syllable, and a non-final position; a case in point is the word *cioccolato* [tʃok.ko.laːto] (Wheelock, 2016). In Spanish, there is also some degree of variation in vowel production depending on whether they are in a stressed or unstressed syllable and on the phonetic environment in which they occur (Gallardo del Puerto & Gómez Lacabex, 2008).

Based on the contrastive and error analysis literature available, the English vowel contrasts which appear or are predicted to be most problematic for L1 Italian and Spanish learners are listed below, along with example minimal pairs (Duguid, 2001; Gómez González & Sánchez Roura, 2021; Rogerson-Revell, 2018; Wheelock, 2016).

/iː/ **- /ɪ/, /ʃ**i**ː**p**/ – /ʃɪp/** (sheep – ship).

/uː/ **- /ʊ/, /p**u**ː**l**/ – /p**ʊ**l/** (pool – pull).

/ɔː/ **- /ɒ/, /k**ɔ**ː**t**/ – /k**ɒ**t/** (caught – cot).

/æ ʌ ɑː/

/æ/ - /ʌ/, /kæ**t/ – /k**ʌ**t/** (cat – cut).

/ɑː/ **- /æ/, /p**ɑ**ː**t**/ – /p**æ**t/** (part – pat).

/ɑː/ **- /ʌ/, /st**ɑ**ː**f**/ – /st**ʌ**f/** (staff – stuff).

/ə ɜː/

/ɜː/ **- /ɑ**ː/, /ˈ**k**ɜ**ː.tən/ – /ˈk**ɑ**ː.tən/** (curtain – carton).

The schwa /ə/, the most frequently occurring vowel in English, can replace the pronunciation of any vowel or glide in unstressed syllables. Additionally, it can appear in stressed syllables as part of a centring glide or a triphthong.

2.4.4 Diphthongs

On top of the 12 pure vowels displayed above, English also stands out for its extensive use of diphthongs. A diphthong is usually defined as a sequence of two vowels with a different quality, belonging to the same syllable (Krämer, 2009). One of the two sounds in the sequence is a vowel sound called the nucleus, while the other is an "off-glide sound"; in English, the nucleus always comes before the glide segment (e.g., <here> /hɪə/ and <how> /haʊ/) (Rogerson-Revell, 2018). According to Rogerson-Revell (2018), there are 7 diphthongs in total in British English: /eɪ/ /aɪ/ /ɔɪ/ /ɪe/ /ʊə/ /əʊ/ /aʊ/. However, /aʊ/, /aɪ/ and /ɔɪ/ are the only common ones in both British and North American varieties of English due to the great dialectal variation which diphthongs undergo across different varieties (Rogerson-Revell, 2018).

The classification of diphthongs in other languages may vary depending on the definition that one chooses for describing diphthongs. In Spanish, Silva Valencia (2022) identifies 14 diphthongs and defines them as combinations of two vowel sounds. This definition may sound too simplistic, given that a hiatus is also a sequence of two adjacent vowels which belong to two different syllables. By contrast, only 5 Spanish diphthongs are listed in Uribe Enciso et al. (2019). Although the authors do not refer to any definition explicitly, one may assume that they are applying stricter criteria. With regards to Italian, diphthongs are also widely used and can be either falling, changing from mid/low

to high or from back to front (e.g., aj, aw, uj, ej, etc.) or rising, where the first part is always a glide (e.g., wi, we, wa, wo, jo, je, ja, etc.) (Krämer, 2009).

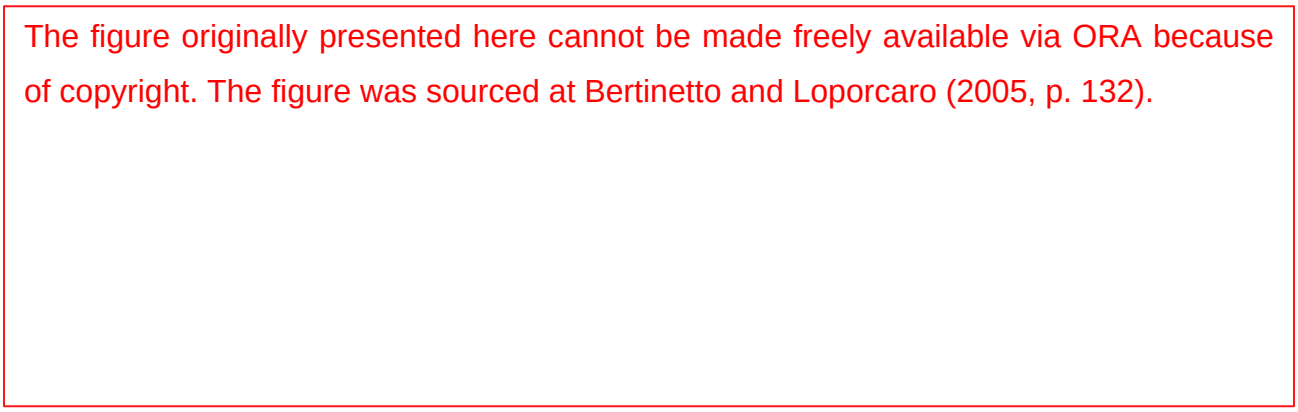
Usually, Italian and Spanish learners of English do not have major issues perceiving and producing English diphthongs. One exception may be the English /əʊ/ diphthong in words like ‘go’ or ‘goat’, which these learners tend to pronounce as a monophthong /o/ instead (Coe, 2001; Duguid, 2001). This is a case where the diphthong is not perceived, giving rise to confusion between /o/ and /əʊ/: for example, in the words got and goat, where both vowels are realised as the Italian or Spanish /o/. A similar confusion may arise between /e/ and /ei/: get and gate, both being realised as /e/ (Coe, 2001; Duguid, 2001).

2.4.5 Consonants: Italian vs English

The consonant inventory of Italian does not differ from that of English to a great extent. As can be observed in the two tables below, Italian has 23 consonant phonemes, whereas English has 24.

Table 1

Italian consonant phonemes



Note. Source: Bertinetto and Loporcaro (2005, p. 132).

Table 2

English consonant phonemes

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Rogerson-Revell (2018, pp. 100).

Note. Source: Rogerson-Revell (2018, pp. 100).

Both languages feature the same six plosives (/p, b, t, d, k, g/), in three pairs that differ in voicing. That is, the voiceless bilabial plosive /p/ and its voiced counterpart /b/ form a minimal pair in words such as *pari* [ˈpari] 'even', and *Bari* [ˈbari] 'place name' in Italian and *pie* [pʰaɪ] and *buy* [baɪ] in English. Other minimal pairs with plosives in Italian include: *tino* [ˈtiːno] 'vat' and *Dino* [ˈdiːno] 'proper name', *cola* [ˈkoːla] 'percolate.3sg' and *gola* [ˈgoːla] 'throat'; and in English: tin [tʰɪn], din [dɪn], Kate [kʰeɪt], gate [geɪt] (Bertinetto & Loporcaro, 2005; Rogerson-Revell, 2018).

Voiceless plosives are aspirated in English in word-initial position (e.g., *pat* [pʰæt]) or at the beginning of a stressed syllable (e.g., *atomic* [ə'tʰɒm.ɪk]), whereas voiced stops are not (e.g., *bat* [bæt]; *agglomerate* [ə'glɒm.ər.eɪt]). Italian lacks this phenomenon, so omission of aspiration may cause intelligibility problems given that aspiration, on top of voicing, represents a significant difference between /p, t, k/ and /b, d, g/ (Rogerson-Revell, 2018). Thus, failing to aspirate a syllable-onset voiceless stop may lead the native listener to perceive its voiced counterpart (e.g., the unaspirated pronunciation of the word *pin* *[pɪn] perceived as *bin*). /t/ and /d/ are considered dental phonemes in Italian, whereas in English, their place of articulation is usually alveolar (Duguid, 2001). This may cause intelligibility problems as dental realisations of English /d/ and /t/ are common and can be replaced or confused with interdental fricatives, which, as explained below, are often articulated as dental /t/ and /d/ by these learners (e.g., *they – day, thick – tick, thought – taught, though – dough* may sound the same) (Berti, 2015).

Moving on to consider fricatives, English shares all its fricative phonemes with Italian except for the interdental voiceless/voiced pair /θ, ð/ and the glottal /h/. Italian learners usually articulate /θ/ as a voiceless stop [t] or as a labio-dental fricative [f] (e.g., *[tɪŋk] or *[fɪŋk] for ***think*** [θɪŋk]), and the voiced /ð/ as [d] (e.g., *[dɪs] for ***this*** [ðɪs]) (Wheelock, 2016). The omission of the glottal fricative in Italians' L2 speech can also negatively affect their intelligibility (Berti, 2015). Although the <h> grapheme is used in writing, it is always silent in Italian; therefore, the L2 word ***house*** /haʊs/ may be pronounced as *[aʊs] due to the absence of /h/ in the Italian phonemic inventory (Rogerson-Revell, 2018). Moreover, /ʒ/ only appears in Italian loanwords such as ***abat-jour*** [abaˈʒur] 'table lamp', ***garage*** [gaˈraʒ] 'garage', and is therefore sometimes excluded from Italian phonological inventories (see Vincent, 1988). /z/ has a limited distribution, since it contrasts with /s/ only in intervocalic position within or at the right edge of a lexical morpheme (e.g., [ˈfuːso] 'spindle' vs. [ˈfuːzo] 'fused', both spelled ***fuso***) (Bertinetto & Loporcaro, 2005). Both /ʒ/ and /z/ are contrastive consonants in English.

There is only one lateral phoneme in English, /l/, whereas Italian has two: /l/ and /ʎ/; this leads to very few difficulties for Italian learners of English in producing this phoneme, except for, perhaps, word-final contexts in which this becomes a dark l, namely, a different allophone of /l/ (e.g., milk, fill, etc.) (Rogerson-Revell, 2018). As for the rhotic consonant, it can only occur either as an onset or as a coda in Italian, but not as a nucleus (Berti, 2015). According to Krämer (2009), it can be pronounced as a tap when short [ɾ] and as a trill when long [r]; the latter realisation usually occurs in intervocalic position when the consonant becomes a geminate (e.g., carro [ˈkarro] 'cart') and occasionally word-initially (e.g., ***Roma*** [ˈroma] 'Rome'). By contrast, this phoneme is articulated as an approximant and retroflex phone in English, as in the word ***red*** [ɻed]. However, the realisation of this phoneme varies largely within different accents of English, being, for instance, articulated as a trill in some Scottish accents (Rogerson-Revell, 2018).

As for nasals, the alveolar /n/ and the velar /ŋ/ are both part of the phonological inventory of English; however, whereas the former is a phoneme of Italian, the latter is an allophonic realisation of /n/ occurring before velar stops (e.g., ***ancora*** [ˈaŋkɔra] 'anchor'). Therefore, this difference may negatively affect intelligibility when transferred to the L2 speech; for instance, the word ***thing*** [θŋ] may be pronounced as *[θɪŋg], potentially leading the listener to interpret it as ***think*** [θɪŋk] (Wheelock, 2016). Furthermore, English phonotactics allows for /s/ + /l, m, n/ consonant clusters (e.g., ***sleep*** [sliːp]; ***small*** [smɔːl]), whereas Italian does not. That is, /s/ in Italian undergoes assimilation to its adjacent consonant and is, therefore, voiced in words like ***slitta*** [ʒlitta] 'sleigh'. Transferring this L1 phonotactic rule to the L2 phonology can result in mispronunciations like the following: *[zliːp] for ***sleep*** and *[zmɔːl] for ***small*** (Wheelock, 2016).

Finally, consonant length is contrastive in Italian with long consonants (geminates) occurring only word-internally, typically intervocalically (e.g., *fatto* [ˈfatːo] ‘deed’ vs *fato* [ˈfato] ‘fate’), or before glides (e.g., *occhiali* [oˈkːjaːli] ‘spectacles’). There is also a set of geminates consisting of oral stops and /f/ occurring before laterals or trills (e.g., *afflitto* [aˈfːlitːo] ‘sorrowful’). Consonant length in English is not distinctive and, although this rarely affects intelligibility, it can influence how Italian learners interact with English orthography (Bassetti et al., 2021). They tend to pronounce a double consonant grapheme in English as a geminate as they would do in their L1. For instance, <Finnish> and <finish>, which are homophones in English, would be pronounced with a long and a short /n/ sound, respectively (Bassetti et al., 2022). Regarding grapheme-to-phoneme correspondence, the orthographic system of Italian is highly transparent, as is that of Spanish, as opposed to English. This hurdle often leads to phonological errors in the L2 (Wheelock, 2016).

2.4.6 Consonants: Spanish vs English

The consonant inventory of Spanish contains 19 consonants as opposed to the 24 of English.

Table 3

Spanish consonant phonemes

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Campos-Astorkiza (2013, p. 93).

Note. Source: Campos-Astorkiza (2013, p. 93).

Given the close typological relation between Spanish and Italian, there are several areas where both languages contrast with English. One key difference lies in the realisation of voiceless plosives. In Spanish, as in Italian, /p, t, k/ are never aspirated, as opposed to English. This may lead to intelligibility problems, given that voiceless unaspirated stops in word-initial/syllable-onset position may be perceived as their voiced counterparts /b, d, g/ (Coe, 2001). For instance, the unaspirated pronunciation of *pear* *[peər] may be perceived as *bear*. Another area of divergence concerns the place of articulation of the stops /t, d/, which are usually dental in Spanish (e.g., *tomar* [to'maɾ] ‘take’, *dolor* [do'loɾ] ‘pain’), but alveolar in English (e.g., tear [teər], dear [dɪər]). This mismatch can lead to confusion with other

dental or interdental sounds (e.g., *they-day*). Spanish rhotics also differ considerably from their English counterparts; they are pronounced as a tap /ɾ/ (e.g., *cara* ['kaɾa] 'face') or trill /r/ (e.g., *tierra* ['tjera] 'earth'), whereas in English they are usually retroflex (e.g., *road* [ɻʊd]). Spanish learners often transfer their native rhotic sounds to English, which may be perceived as different sounds and cause confusion. Finally, the voiced velar nasal [ŋ] occurs in Spanish only word-internally as an allophone (e.g., *blanco* ['blaŋko] 'white'), whereas in English it is a phoneme typically found word-finally. Spanish speakers tend to pronounce a /k/ after it (e.g. 'sing' /sɪŋ/ pronounced as *[sɪŋk]) (Coe, 2001).

Moving on to consider the phonological features that are unique to Spanish, when voiced plosives /b, d, g/ are in intervocalic position, they undergo spirantisation, a phonological process under which they are weakened and articulated as voiced fricatives [β, ð, ɣ] (Anderson & Centeno, 2018). This may be an issue when Spanish speakers give English plosives their L1 values; for instance, the word 'adapt' /əˈdæpt/ would be produced as [əˈðæpt] (Uribe-Enciso et al., 2019). Voiced stops do not occur word-finally in Spanish, so L1 speakers often devoice them, leading to mispronunciations that may compromise intelligibility (e.g. pronouncing 'cub' as 'cup' [kʌp]) (Gómez González & Sánchez Roura, 2021). Moreover, the English contrast between /b/ and /v/ may be problematic for speakers of this language, considering that the graphemes <b> and <v> are pronounced as /b/ in Spanish, where /v/ is not a phoneme. Although /s/ is a phoneme of Spanish, [z] only exists as an allophone and, therefore, it can be hard for L1 Spanish learners to distinguish between these two English phonemes; they may, for example, pronounce the words 'eyes' /aɪz/ and 'ice' /aɪs/ as homophones (Veiga-Pérez, 2017).

Finally, another difference between Spanish and English concerns syllable structure and phonotactic constraints. Onsets and codas can have up to four consonants in English (e.g., ***spr**ay* /sreɪ/, ***sixths*** /sɪksθs/), whereas Spanish phonotactics allows a maximum of two (e.g., **tr**a**ns**-cribir 'transcribe') (Gómez González & Sánchez Roura, 2021, p. 157). Syllable-onset /sC/ clusters in English (e.g., **Sp**ain, **sl**ogan, **st**andard) are particularly challenging to produce for Spanish-speaking learners, given that this combination is not allowed by their L1 phonotactics. Therefore, a prothetic vowel [e] often precedes the /sC/ cluster in the L2 speech of these learners, resulting in the following realisations *[espeɪn, eˈsləʊ.gən, eˈstæn.dəd] (De Leeuw et al., 2021).

2.4.7 Predicted Problematic Features

Overall, the phonological features that are likely to be challenging to perceive and produce for Italian and Spanish learners of English are outlined in Table 4. These include both segmental (consonantal and vocalic) difficulties and issues related to phonotactic constraints. Prosodic differences exist but will not be discussed here, as they fall outside the scope of the present dissertation.

Table 4[1]

Predicted problematic phonological segments for L1 Spanish and L1 Italian learners of English

Italian	
Consonants:	/θ/ /ð/ /h/ /z/ - /s/ /ɽ/ /ŋ/
Vowels:	/iː/ - /ɪ/ ; /uː/ - /ʊ/ ; /ɔː/ - /ɒ/ ; /æ/ - /ʌ/ ; /ɑː/ - /æ/ ;/ɑː/ - /ʌ/ ; /ɜː/ - /ɑː/ ; /ə/ ; /əʊ/ - /ɒ/; /ei/
Features causing issues due to phonotactics:	word-initial [pʰ] [tʰ] [kʰ] word-initial /s/ + /l/ /s/ + /l, m, n/ clusters
Spanish	
Consonants:	/θ/ /ð/ /h/ /z/ - /s/ /ɽ/ /ŋ/ /b/ - /v/
Vowels:	/iː/ - /ɪ/ ; /uː/ - /ʊ/ ; /ɔː/ - /ɒ/ ; /æ/ - /ʌ/ ; /ɑː/ - /æ/ ; /ɑː/ - /ʌ/ ; /ɜː/ - /ɑː/ ; /ə/ ; /əʊ/- /ɒ/ ; /ei/
Features causing issues due to phonotactics:	intervocalic /b/ /d/ /g/ word-final /b/ /d/ /g/ syllable onset /sC/ clusters

Although consonants also pose difficulties for these learners, this dissertation focuses on vowels, for both theoretical and methodological reasons. From a theoretical perspective, English vowels are especially challenging for learners whose L1 has a smaller vowel inventory, making them more prone to L1-based transfer errors (Flege, 1995; Escudero, 2009). Methodologically, a selection had to be made among the segmental features to ensure focus and statistical power. Given that vowels have been shown to cause greater challenges than consonants for these learner groups, they were prioritised as the object of investigation, and only a limited number of vowel contrasts were selected for systematic analysis.

[1]In Table 4, a dash appears between two segments to signify that learners usually confuse them or use them interchangeably in English.

2.5 Empirical Evidence from Previous Error Analysis Studies

To bridge the gap between the theoretical framework provided by CA and the actual phonological issues learners face, it is crucial to conduct an error analysis (EA). Whereas the CA can predict potential challenges learners might encounter, empirical evidence is needed to confirm whether these predictions match learners' difficulties. EA provides valuable insight into which phonological contrasts are problematic for learners, as it is based on real-life performance rather than theoretical expectations. This section reviews past EA studies, which provide empirical data on the phonological errors produced by Italian and Spanish-speaking learners of English. This will help establish a foundation for understanding how L1 transfer affects L2 phonology.

2.5.1 EA Studies with Italian Participants

The literature on EA specific to Italian learners is rather limited. Wheelock's (2016) study, an important contribution to the field, explores the phonological challenges faced by Italian learners of English. First, a CA was conducted to compare the phonological systems of Italian and English. Then, an EA approach was adopted, which involved examining spoken learner data extracted from the online Speech Accent Archives (SAA). This database contained speech samples of 27 EFL Italian learners, all of whom had lived in the US for at least three months. The participants read an elicitation passage aloud, and their productions were phonetically transcribed, analysed and compared against a model North American English (NAE) transcription of the same passage. This allowed for the identification of the most common pronunciation errors, termed 'misformations', produced by the participants. The six most common vowel and consonant misformations are reported below:

Table 5

The six most common vowel and consonant misformations and their average error rate

Vowel Misformations	**Consonant Misformations**
/ɑ/ as [V][2] - (80.6 %)	/z/ as [s] - (51.5 %)
/oʊ/ as [o] - (60.5 %)	/v/ as [f] - (44.4%)
/æ/ as [V] - (32.6 %)	/ð/ as [d] - (42.9 %)
/ɜ/ as [ɛ] - (28.7 %)	/θ/ as [t] - (20.4%)
/i/ as [ɪ] - (12.0 %)	/g/ as [k] - (15.7 %)
/ɪ/ as [i] - (10.9 %)	/d/ as [t] - (8.5%)

[2] V stands for any other vowel.

Note. Adapted from Wheelock (2016, p. 49).

As shown in Table 5, some of these challenging sounds are not part of the Italian phonological inventory, such as /θ/ and /ð/. As for vowel phonemes, the misformation of /ɑ/ as any other vowel had the highest error rate, followed by /oʊ/ pronounced as [o], and /æ/ produced as different vowels (Wheelock, 2016).

The author was also interested in determining which problematic segments were retained in the speech of higher-performing learners. Therefore, a sub-cohort of 7 higher-performing participants, who had spent longer periods in the US, was selected. These participants kept mispronouncing both /ɑ/ and /oʊ/ at high rates, 75.0 % and 61.9 %, respectively. This suggests that even learners who had been living in the US for longer periods still had issues in mastering the pronunciation of these two English phonemes. However, lower error rates were observed for /æ/ (14.3 %), /ɜ/ (10.7 %), /i/ (9.9 %) and /ɪ/ (5.0 %), which may potentially reflect higher degrees of exposure to the target language while living in the USA. As for consonants, the error rate was lower for all the phonemes considered; however, the misformation of /z/ as [s] was still recurrent in the speech of higher-performing learners.

Despite its value, the study has some limitations. First, the sample size, especially in the subgroup analysis, was small, which limits generalisability. Furthermore, the author did not provide information about the participants' regional L1 background, which may influence their pronunciation patterns. Finally, the use of a read-aloud task raises questions about orthographic interference, as participants may have relied on Italian grapheme-phoneme correspondences during production. These issues highlight the need for further research involving more diverse samples and additional task types.

Wheelock's (2016) investigation appears to be the only EA study focusing specifically on Italian learners of English, indicating a gap in the literature. Although some other studies refer to English phonological features that seem to be problematic for Italian speakers, these cannot be regarded as EA studies and, therefore, cannot be included in this section (see Flege et al., 1999; Munro et al., 1966). This lack of comprehensive data underscores the need for further research regarding how Italian speakers navigate phonological contrasts in English.

2.5.2 EA Studies with Spanish Participants

A more substantial body of EA research has focused on vowel-related difficulties experienced by Spanish-speaking learners of English. This subsection reviews key studies that investigate both the

perception and production of English vowel sounds by L1 Spanish-speaking participants, identifying recurring problematic areas and relevant pedagogical implications.

Veiga-Pérez (2017) explored the L2 English perception and production abilities of 20 Spanish high school students, who had studied English for approximately 12 years. The participants completed listening and pronunciation tasks targeting vowel pairs /æ/-/ʌ/ and /ɪ/-/iː/, among other contrasts. First, they were administered two listening tasks (one for vowels and one for consonants) where they had to listen to three minimal pairs per contrast and select the word they heard. They were then tested individually on their pronunciation of the same words using 24 stimulus cards. The findings revealed a dissociation between perception and production abilities: while the learners showed relatively high perception accuracy for these vowel contrasts, their pronunciation performance was considerably weaker. This discrepancy was particularly evident in the production of /ae/, /ʌ/, and /ɪ/, suggesting that accurate perception does not automatically lead to accurate production.

Vera Diettes (2014) investigated the perception and production of the English schwa by Colombian Spanish speakers and tested the assumptions of Flege's (2003) SLM. Since the schwa is absent from the Spanish vowel inventory, it was expected to be acquired more easily than similar but not identical sounds. 10 L1 Colombian Spanish learners and 5 British English speakers participated in the study, which consisted of pre-tests, training, and post-tests. Learners completed word identification and read-aloud tasks, and their speech data was analysed using Praat. While initial perception accuracy for schwa reached 70%, production was poor, with participants replacing schwa with L1-equivalent vowels. Orthographic interference led learners to perceive and produce the schwa inaccurately. After targeted pronunciation training, the experimental group showed considerable improvement in producing schwa-like vowels, confirming the SLM prediction that new phonetic categories can be acquired with focused input.

Calvo Benzies (2017) investigated the production of English vowels by 10 advanced Spanish EFL learners enrolled in the third and fifth years of a BA in English Studies in Galicia, Spain. Drawing on prior CAs, the study focused on vowels widely regarded as problematic for Spanish learners, including /ʌ/, /ɑː/, /æ/, /ɔː/, /ɜː/, /əʊ/, and schwa. Unlike prior work, this study adopted an empirical approach, collecting oral data through personal interviews, a picture description task, and a read-aloud task. Participants' productions were transcribed using the IPA, and a total of 209 pronunciation errors were identified. Fifth-year students produced slightly fewer errors (99) than third-year students (110), suggesting a modest effect of instruction and experience. The most frequent mispronunciations occurred with schwa (62 errors), /ɔː/ (43 errors), /ɑː/ (37 errors), and /əʊ/ (15 errors). As noted by the

author, the read-aloud task may have contributed to orthographic interference, influencing pronunciation accuracy. Despite participants' advanced proficiency, the persistence of these errors indicates that certain L2 vowel contrasts remain challenging even at higher levels.

More recent studies reinforce these findings. Cebrià̀n et al. (2021) examined Spanish learners' perception and production of the /iː/-/ɪ/ contrast. Their results confirmed poor discrimination and production, especially for the shorter vowel /ɪ/, which tends to be assimilated into the Spanish /i/. Similarly, Conklin et al. (2019) reported difficulty among Spanish speakers in distinguishing between and producing /ae/, /ʌ/, and /ɑː/, a triad often subject to overlap or neutralisation due to the lack of corresponding vowel categories in Spanish. Taken together, these studies highlight consistent patterns of difficulty in the acquisition of English vowel sounds by L1 Spanish learners, with schwa, long-short vowel contrasts, and unfamiliar mid/low central vowels being particularly problematic. The evidence emphasises the importance of integrating focused pronunciation work into L2 instruction and of addressing not only L1 transfer but also orthographic interference and phonological awareness.

2.5.3 Integrating CA Predictions with EA Findings

CA, on the one hand, was used to identify potential areas of difficulty by comparing the English vowel system with those of Italian and Spanish (James, 2013). On the other hand, EA has the purpose of testing CA predictions empirically by analysing learners' actual performance in perception and production tasks (Corder, 1982; Rustipa, 2011). The integration of CA and EA may lead to a more comprehensive understanding of phonological challenges faced by L1 Italian and L1 Spanish learners of English.

The EA studies reviewed above highlight consistent vowel difficulties that correspond to the CA predictions in Section 2.4.7. For Italian learners, Wheelock (2016) reported issues with vowels like /ɑː/, /əʊ/, /æ/ and, to a lesser extent, /ɜ/, /iː/ and /ɪ/; in particular, /ɑː/ and /əʊ/ had the highest error rate even in higher performing learners. Along the same lines, studies on Spanish learners point to problems with central vowels such as /ʌ/, /ɜː/, and schwa, as well as the /iː/-/ɪ/ contrast, and recurring difficulties with /ae/, /ʌ/, and /ɑː/ (Calvo Benzies, 2017; Cebrià̀n et al., 2021; Conklin et al., 2019; Veiga-Pérez, 2017; Vera Diettes, 2014). All these vowels were identified as potentially problematic for both Spanish and Italian learners due to the limited size of their L1 vowel inventories.

Given the scope of the present study, it was necessary to narrow down the focus to a subset of vowels. 7 English vowel phonemes (/iː/, /ɪ/, /ɑː/, /æ/, /ʌ/, /ɒ/, /əʊ/) were selected for the perception task, and 4 (/iː/, /ɪ/, /ɑː/, /æ/) for the production task.

2.6 Research Questions and Study Aims

The present study aims to investigate the perceptual and productive difficulties that L1 Italian and L1 Spanish learners face when acquiring English vowel contrasts. While previous research has identified various L2 vowel categories as challenging for Romance language speakers, fewer studies have directly compared perception and production performance across different L1 groups, nor examined how learner proficiency shapes these difficulties. Proficiency has often been categorised into broad levels (e.g., low vs. high), which may overlook important individual differences. A continuous measure of proficiency may offer a clearer picture of how it interacts with learners' phonological performance. Moreover, while some difficulties appear to decrease with increasing proficiency, others persist regardless of level, raising further questions about the nature of L2 phonological acquisition. Furthermore, the extent to which perception and production of L2 vowel contrasts are aligned remains a topic of ongoing debate. Addressing these gaps, the study is guided by the following research questions:

1. Which L2 vowels pose the greatest difficulty for L1 Italian and L1 Spanish learners of English in (a) perception and (b) production?
2. Does proficiency affect which vowels learners find problematic in (a) perception and (b) production?
3. Is the learners' perception of L2 vowels more accurate than their production of the same vowels?

3. Methodology

This chapter outlines the methodological design of the study, including the rationale behind the quasi-experimental approach, participant recruitment, data collection instruments, procedural details, coding methods, ethical considerations and methodological limitations.

3.1 Methodological Approach

This study adopted a quasi-experimental design to investigate L2 vowel perception and production in L1 Italian and Spanish learners. The participants were recruited based on their linguistic background to target English vowel phonemes that were predicted to pose difficulties for these learner groups. The quasi-experimental nature of the research allowed for the collection of empirical data under controlled conditions while maintaining ecological validity by working with naturally occurring learner populations (Mackey & Gass, 2015).

To ensure broad accessibility, the study was conducted entirely online using the Gorilla Experiment Builder platform (Anwyl-Irvine et al., 2020). This allowed for controlled presentation of stimuli and automatic data collection, while facilitating participation from learners in different geographical locations. The online format also enabled the use of native-speaker audio recordings, image-based perception stimuli, and integrated voice recording for the production task, all within a single experimental session.

3.2 Participants

A total of 124 participants (87 Italian and 37 Spanish) were recruited[3] using a snowball sampling technique, whereby participants were referred through existing contacts. This approach was particularly appropriate given the researcher's lack of access to language classrooms or formal educational settings (Cohen et al., 2018). Participants could take part in the study if they were adult L1 speakers of Italian or Spanish with a history of learning English as an L2. However, following the export of the Gorilla dataset, an additional 23 participants were excluded during the data cleaning and wrangling phase due to the absence of their production and/or perception data. Specifically, for the production task, a threshold of 50% task completion was established: participants needed to have submitted at least 8 out of 16 experimental recordings to be retained in the final dataset.

[3] Before data analysis, 316 live participant entries were manually excluded from the Gorilla platform. These participants either experienced technical issues, failed to engage with the experimental tasks, or exited the study too early. The criterion used for inclusion at this stage was that participants must have reached the perception task, which followed the production task. This ensured that usable production data, and potentially perception data, were available for analysis.

Thus, the resulting dataset used for statistical analysis included 101 participants: 71 Italian and 30 Spanish learners of English. Their age and gender distributions are presented in Table 6. Among the Italian participants, a large proportion reported Sardinia as their region of origin, and Sardinian Campidanese (n = 38) was the most frequently reported dialectal variety spoken. Full regional and dialectal data are available in Appendices A and B.

Table 6

Summary of Participant Demographics

	Age	**Gender**			**Total**
	Mean (SD)	M	F	Other	
Spanish	27.6 (7.6)	14	16	0	30
Italian	37.0 (15.0)	42	28	1	71

3.3 Data Collection Instruments

The present study followed a quantitative approach for data collection. The focus was on collecting production and perception accuracy scores for specific English vowel phonemes, as well as receptive vocabulary scores. The data collection was structured, task-based, and automatically recorded using the Gorilla Experiment Builder platform.

3.3.1 Background Questionnaire

Participants first completed an online questionnaire designed to elicit background information on their language learning history. The questionnaire collected demographic details as well as linguistic information such as native language(s), current language use, and formal English learning experience. These data were used to describe the sample and provide contextual insight into participants' exposure to English. The full questionnaire is included as Appendix C.

3.3.2 LexTALE: Receptive Vocabulary Test

To assess participants' proficiency in English, the study used the LexTALE test, a validated lexical decision task developed by Lemhöfer and Broersma (2012). The test consists of English word and nonword stimuli, and participants have to select whether a stimulus is a word or not. LexTALE scores have been shown to correlate strongly with more comprehensive tests of L2 proficiency (e.g., the

Quick Placement Test) and are widely accepted as a proxy for general English proficiency in SLA and psycholinguistic research, especially with European language speakers (Chaouch-Orozco et al., 2023; Cilibrasi & Marková, 2022; Nakata et al., 2020). The version of LexTALE used in the current study was adapted for the Gorilla platform by Declerck and Kirk (2023) and is publicly available via Open Materials. Unlike the original version, it includes 63 items instead of 60. However, this small discrepancy is negligible as it does not affect the interpretation of continuous scores or their use in regression models.

Despite its strengths, recent research has raised cautionary notes about LexTALE's universality. Puig-Mayenco et al. (2023) found that correlations between LexTALE scores and standardised proficiency measures were moderate or low across L1 Spanish and L1 Chinese learners, particularly at lower proficiency levels. These findings suggest that while LexTALE remains a robust measure of vocabulary knowledge, its validity as a standalone indicator of global proficiency may vary across learner groups. Therefore, LexTALE was used in the current study not as a categorical proficiency classifier but as a continuous vocabulary-based estimate to examine the predictive relationship between receptive proficiency and phonological accuracy.

3.3.3 Stimuli Selection

The stimuli used in the production and perception tasks were selected through a two-stage process. First, a CA was conducted to compare the vowel inventories of English, Italian, and Spanish (see Section 2.4). Based on this analysis and previous EA studies (Cebrian et al., 2021; Veiga-Pérez, 2017; Wheelock, 2016), 4 English vowels were selected for the production task: /iː/, /ɪ/, /ɑ/, and /æ/. To reduce the likelihood that participants would consciously focus on these experimental vowels, /e/ and /ɒ/ were used as control vowels and selected based on their low predicted difficulty for the target learner groups.

For the perception task, the range of stimuli was expanded to include 7 vowels: the 4 used in the production task (/iː/, /ɪ/, /ɑː/, /æ/) plus /ʌ/, /ɒ/, and the diphthong /əʊ/. The vowel /ɒ/, although used as a control item in production, was included here as a target vowel due to literature suggesting that learners often struggle to perceptually differentiate it from /əʊ/, particularly in minimal pairs (Coe, 2001; Duguid, 2001). This decision was informed by research highlighting perceptual difficulty arising not from the individual vowel but from specific contrastive pairings, in this case /ɒ/ and /əʊ/, which learners may struggle to distinguish due to L1 phonological influence and overlap in acoustic features (Flege, 2003; see Section 2.5).

3.3.4 Production Task: Read-aloud

Participants were presented with short sentences containing target words designed to elicit specific L2 vowel contrasts; there were 24 sentences in total, with 4 sentences per vowel. The stimuli included 4 experimental vowels predicted to be problematic for the target population (/iː/, /ɪ/, /ɑː/, /æ/), and 2 control vowels (/e/, /ɒ/). Each vowel was represented in 4 separate sentences, resulting in 16 experimental and 8 control items. All sentences were 7 words in length and were manually matched for syntactic complexity to accommodate learners of varying proficiency levels. Control sentences helped reduce the salience of experimental manipulation and prevented participants from becoming overly focused on specific vowel categories. For this reason, only the experimental items were analysed in the final dataset, as controls only served as distractors. The full list of sentences used in the production task is provided in Appendix D.

Participants were instructed to read each sentence aloud and record their voice using the Gorilla platform. They were only allowed one trial per item; thus, they could not re-record their response. All recordings were later transcribed and coded to assess production accuracy; only the target vowel in each experimental item was transcribed following the International Phonetic Alphabet (IPA).

3.3.5 Perception Task

The perception task was presented after the production task to minimise the risk of priming effects influencing vowel production. Participants listened to pre-recorded auditory stimuli produced by a native speaker of Southern Standard British English (SSBE) and were instructed to identify the word they heard by selecting the corresponding image from a pair of pictures. The goal of the task was to determine whether the same English vowel contrasts that posed production difficulties for learners also caused perceptual confusion.

The stimuli were organised into minimal pairs or closely matched lexical items, targeting 7 English vowels: /iː/, /ɪ/, /ɑː/, /æ/, /ʌ/, /ɒ/, and the diphthong /əʊ/. Each trial featured two pictures on the screen, one corresponding to the target word and the other to a distractor. Participants heard a single word and had to click on the image that matched what they had heard. All stimuli/images were presented twice: once as the distractor and once as the target. For a selection of trials (n = 8), the same stimulus was presented again later to assess consistency (repetition). For other trials (n = 8), the target and distractor switched roles (reverse). The reverse items were used to maintain attention and reduce response strategies, such as consistently selecting the image on one side of the screen. A total of 8 items were repeated, and another 8 items were presented in reverse, summing to 16 out of the 32 total trials. The assignment of items to repetition and reverse conditions was determined randomly. All stimuli were

embedded in monosyllabic, high-frequency lexical items (e.g., *sheep*, *bin*, *cup*), and the images were selected to represent the referents without introducing visual ambiguity. A total of 32 trials were presented, distributed across 16 target words involving 7 vowel contrasts: 4 pairs targeting /iː/ - /ɪ/, 2 for /ɑː/ - /æ/, 2 for /ɑː/ - /ʌ/, 4 for /æ/ - /ʌ/, and 4 for /əʊ/ - /ɒ/. The full list of stimuli, including images and word pairs, presence of repetition or reversed conditions, is provided in Appendix E.

3.4 Procedure

The experiment was conducted entirely online using the Gorilla Experiment Builder platform. The experimental session followed a fixed sequence of tasks designed to guide participants smoothly from informed consent through to the final perception task, with a typical completion time of approximately 20 minutes. Participants were first presented with an informed consent form, which they were required to read and approve before proceeding. Then, they completed a background questionnaire (see Section 3.3.1), followed by the LexTALE test (see Section 3.3.2), followed by the production and perception tasks, respectively.

At the end of the session, participants were given the option to enter a lottery draw for a £50 voucher, with one winner drawn per language group (Italian and Spanish). Email addresses for the draw were collected separately via a secure link hosted on the University of Oxford's servers, ensuring that participants' experimental data remained fully anonymous.

3.5 Piloting and Technical Issues

Before launching the official experiment, two rounds of piloting were conducted to evaluate the validity and feasibility of the study design and minimise technical disruptions. The first round served as a technical pilot, while the second focused on final adjustments.

3.5.1 Technical Pilot

The first complete version of the experiment was tested by 5 participants to identify potential functionality issues. This group included 2 native English speakers (one from the US and one from Australia), as well as one Swiss German, one Italian, and one Spanish L1 speaker. At this point, the experiment had not yet incorporated the lottery draw link.

The technical pilot confirmed that the average completion time for non-native participants was approximately 20 minutes. Participants were not anonymous during this phase, as they were invited to provide feedback upon completing the tasks. This feedback highlighted several areas for improvement:

the term “L3” in the background questionnaire was unclear; CEFR proficiency benchmarks were missing from the self-assessment scale; and participants encountered slow page loading, occasional microphone malfunctions, and issues with smartphone compatibility. Based on this input, modifications were made to clarify questionnaire wording, include proficiency benchmarks, and improve the experiment’s interface functionality.

The production task was initially designed as a picture-naming activity, where participants would name objects depicted in images containing key English vowel contrasts (e.g., ship /ʃɪp/ vs. sheep /ʃiːp/). However, results from a preliminary pilot study revealed that the images frequently failed to elicit the intended lexical items due to abstractness, ambiguity, or lexical unfamiliarity. In the pilot study, participants were asked to write down the first word that came to mind after viewing each image. A total of 20 participants evaluated the control pictures (13 native speakers of English and 7 non-native speakers), while 22 participants evaluated the experimental pictures (15 native English speakers and 7 non-native speakers). The non-native group included speakers of various L1s, with only one L1 Italian participant and no L1 Spanish participants. The results revealed low convergence on the intended lexical items. For instance, the image for “chock” was frequently interpreted as “wooden block,” prompting its replacement with “sock”. In response, the task was redesigned as a read-aloud activity, which ensured that all participants produced the target words embedded in controlled sentence contexts.

3.5.2 Official Pilot

A second pilot was conducted with 5 anonymous L1 Italian participants. This version included the lottery draw link, which redirected participants who opted in to a secure University of Oxford server for email submission. Those who declined were able to complete the study without providing personal data.

During this phase, a misconfigured branch in the Gorilla experiment flow caused an error in the redirection to the lottery draw page. This issue was identified and resolved before the official launch. Additional technical problems were also reported: one participant was unable to complete the experiment on their laptop and had to switch to an iPad; another had to switch from laptop to phone mid-task; and one participant encountered a blank screen during the perception task and could not proceed. These issues were discussed with both the Gorilla support team and the project supervisor. Although further fixes were not possible at the platform level, the information sheet was updated to warn participants of potential technical difficulties. A summary of this information was also provided in Spanish and Italian to ensure accessibility for lower-proficiency participants.

3.6 Coding and Interrater Reliability

To assess vowel accuracy, the researcher manually coded the responses from the production task. 2 native English-speaking teachers with knowledge of the IPA were recruited to rate the intelligibility of the recorded speech samples as part of the interrater reliability procedure. Each rater was compensated £65, funded by a research grant awarded for the project. Each target word was evaluated as either correct or incorrect, based on whether the vowel produced matched the intended English phoneme in terms of phonological quality and, where applicable, quantity (e.g., vowel length). Coding decisions were informed by a detailed rubric specifying the target vowel contrasts, corresponding lexical items and included clear instructions and criteria for rating each token (see Appendix F).

To assess the reliability and consistency of the coding procedure, each external rater independently analysed the production data from 24 participants, representing approximately 20% of the full dataset. This subset consisted of 384 tokens per rater. Both raters used the same rubric and were provided with an Excel-based coding template in which they recorded binary accuracy judgements, optional error notes, and general comments.

Simple percentage agreement was used to calculate interrater reliability. According to Gass and Mackey (2005), anything above 75% represents an acceptable level of agreement among raters; however, there is no clear framework in SLA research, and, therefore, this can be taken as a rough guideline. The observed agreement between the researcher and Rater 1 was 82%, while agreement with Rater 2 was 70% (see Appendix G for the observed agreement calculations). These values were deemed acceptable based on existing standards in inter-rater reliability for categorical data and were adopted as the primary metric for assessing coding consistency. Following this exercise, the researcher reviewed the points of divergence, particularly with Rater 2, and found no systematic inconsistencies or evidence of conceptual misunderstanding. Given that the disagreements were limited in scope and did not indicate broader concerns about rubric interpretation, the researcher proceeded to code the remaining data independently, adhering strictly to the same rubric to maintain consistency.

In contrast to the production task, the perception task was automatically scored by the Gorilla platform, based on participants' image selections. No manual coding was required for this component of the data.

3.7 Ethical Considerations

All participants were over 18 and provided consent by confirming that they had read and understood the information sheet outlining the study's purpose, experimental tasks, and data use (see Appendix

H). They were informed that they could withdraw at any point before submitting their responses, without providing a reason and without any negative consequences.

No personally identifiable data were collected as part of the main experiment. The only exception was the optional provision of an email address at the end of the study, which was used exclusively to enter a lottery draw. Participants who opted to enter the draw were automatically redirected to a secure University of Oxford server, separate from the Gorilla platform, to ensure data confidentiality. Participants were explicitly informed that their IP addresses were not stored and that their experimental responses would remain anonymous throughout. A £50 lottery draw (£25 per language group) was planned as a participation incentive to ease the predicted high dropout rate, which is characteristic of online experiments (Cohen et al., 2018). A total of 61 participants submitted their email addresses for the draw; this data was permanently deleted immediately following the completion of data collection.

All data generated by the experiment were stored securely using Nexus365, the University of Oxford's password-protected cloud storage system. This study received ethics approval from the University of Oxford and was conducted in accordance with its ethical standards for research involving human participants (see Appendix I).

3.8 Methodological Limitations

Methodological limitations must be acknowledged when interpreting the findings of this research. First, while efforts were made to maximise sample size, some participants were excluded from all analyses due to incomplete data. Although a per-analysis exclusion strategy could have maximised sample size for individual tests, a uniform dataset was prioritised to ensure consistency across analyses and reduce potential bias from unequal groupings (Mackey & Gass, 2015). This decision, while necessary, may have resulted in the loss of potentially usable data, particularly from participants who completed only one of the two main tasks. A further limitation concerns the influence of orthography on production. While the use of a read-aloud task ensured consistent elicitation of target words, it also introduced the possibility that participants' pronunciations were shaped by English spelling conventions, particularly if they had limited phonological awareness of the target vowel contrasts (Bassetti et al., 2020; Colenbrander et al., 2019). This influence is difficult to quantify but is an acknowledged trade-off in production research involving written prompts.

Finally, the online nature of the experiment, while allowing for broad participation and automatic data collection, also presented technical challenges (Cohen et al., 2018). Numerous initial participants were lost due to device incompatibility, audio failures, or dropout, and several reported delayed stimulus

loading, particularly during the perception task, which may have affected response accuracy. In response, several optimisation steps were implemented, including setting a minimum internet speed requirement of 2 Mbps and reducing image file sizes to improve loading times to minimise strain on participant devices. However, given the constraints of running audio-based experiments online, it was not possible to fully eliminate these issues. Additionally, the online format limited the researcher's ability to monitor task engagement and environmental distractions, which could have affected task performance. To mitigate the impact of these limitations, the study employed robust design choices, including piloting, clear exclusion criteria, and inter-rater reliability procedures.

4. Results

This chapter presents the results of the statistical analyses conducted to address the three research questions. Descriptive statistics are first reported in Section 4.1, and the findings from the inferential statistical tests used to tackle each research question are discussed in Sections 4.2, 4.3, and 4.4. Inferential analyses were conducted using RStudio (R Core Team, 2023).

4.1 Descriptive Statistics

4.1.1 Overview

After completing the data collection and filtering process, the resulting participant sample (n = 101) was analysed descriptively first to check whether the data was normally distributed. To do so, data distribution was visualised by creating histograms in R of the data collected for each task: the LexTALE, the Read-Aloud Task, and the Picture-Selection Task. Means, standard deviations, skewness and kurtosis were also calculated for the data in each task, as described below.

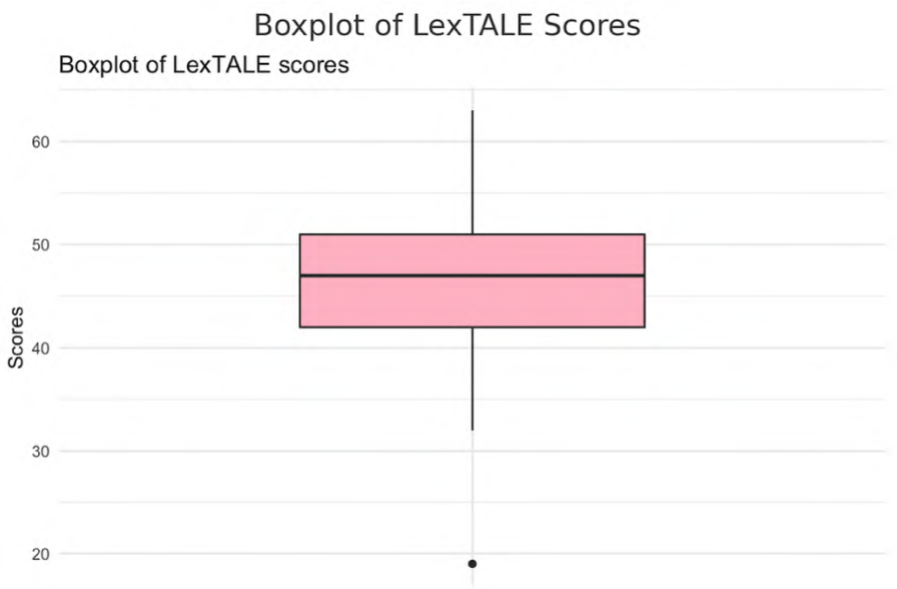
To explore group-level patterns, the data collected from the Italian (n = 71) and the Spanish (n = 30) participants were also analysed separately following the same procedures used for the whole sample. Before calculating the descriptive statistics for each group, the data were checked for normality with histograms and boxplots; particularly, histograms were created for each vowel and each group in all tasks (see Appendices J-P).

4.1.2 LexTALE

The LexTALE scores for the whole participant sample were first assessed for normality through visual inspection and descriptive statistics. As shown in Appendix J and Figure 6, the distribution appeared approximately normal, which was further supported by skewness (-0.22) and kurtosis (0.81), both within the acceptable range of -2 and 2. The mean LexTALE score was 46.93 (SD = 7.46), with a range from 19 to 63.

Figure 6

Boxplot of LexTALE Scores for the whole participant sample



When broken down by group, as illustrated in Table 7, both the Italian and Spanish participants showed similar distributions, with slightly higher average scores and lower variability in the Spanish group. Although individual variability occurs in both groups, the descriptive data indicate that the Spanish participants had a slightly higher proficiency level than their Italian counterparts.

Table 7

Descriptive Statistics for the LexTALE Scores by Group

Group	Mean	SD	Skewness	Kurtosis
Italian	46.80	7.92	-0.16	0.84
Spanish	47.30	6.36	-0.36	-0.52

4.1.3 Read-Aloud Task

In the production task, participants completed read-aloud trials targeting 4 experimental vowels. Accuracy scores were calculated as proportions of correct responses, and data distribution was examined through visual representations in the form of histograms and boxplots (see Appendix K), along with descriptive statistics. The distribution of vowel-specific accuracy scores per group was

visualised using histograms (see Appendices M and N) and boxplots (Appendix Q). Visual inspection suggested that scores were approximately normally distributed for both groups across vowels, with mild variability in symmetry and spread. The descriptive statistics in Table 8 show that overall accuracy in the Read-Aloud Task varied considerably across vowels. For the whole sample, the highest mean accuracy was found for /æ/, followed by /iː/ for which a similarly high mean score was reported. This indicates that participants found /æ/ and /iː/ relatively easier to produce as opposed to /ɪ/ and /ɑː/, which seemed to pose greater production challenges. These patterns were generally consistent across the two language groups.

The distributions of accuracy scores were approximately normal for most vowels, with skewness and kurtosis values falling within the conventional threshold of -2 to +2. However, one notable exception was the Spanish group's production of /æ/, which showed extreme kurtosis (12.08) and strong negative skew (–3.41), suggesting a highly peaked distribution with minimal variability, likely due to a ceiling effect in performance.

Table 8

Descriptive Statistics per Vowel and Group in the Read-Aloud Task

Vowel	Group	Mean	SD	Skewness	Kurtosis
/iː/	Whole sample	0.80	0.27	–1.18	0.32
	Italian	0.81	0.27	–1.21	0.30
	Spanish	0.78	0.30	–1.07	0.10
/ɪ/	Whole sample	0.56	0.31	–0.12	–1.00
	Italian	0.56	0.31	–0.09	–1.06
	Spanish	0.55	0.32	–0.18	–1.03
/ɑː/	Whole sample	0.58	0.34	–0.31	–1.08
	Italian	0.55	0.35	–0.16	–1.24

	Spanish	0.66	0.30	–0.60	–0.55
/æ/	Whole sample	0.86	0.19	–1.14	0.39
	Italian	0.82	0.20	–0.67	–0.60
	Spanish	0.95	0.15	–3.41	12.08

4.1.4 Perception Task

The Picture-Selection Task assessed perception accuracy across 7 English vowels. Accuracy scores were calculated as proportions of correct responses, and histograms for each vowel were used to assess normality (see Appendix L). Table 8 summarises the descriptive statistics, where variability in perception accuracy can be observed across different vowels for the whole sample and the two language groups. Mean scores were highest for /ʌ/, /əʊ/ and /æ/, and lower for /iː/, /ɪ/, and /ɒ/, which suggests greater perception difficulties with the latter vowels. These trends were generally consistent across groups. Slight group-level differences emerged for specific vowels: the Italian group outperformed the Spanish group on /ɪ/ and /ɑː/, while the Spanish group scored higher on /ɒ/. Both groups performed equally well on /iː/, /ʌ/, and /ɑʊ/, where accuracy was high and distributions indicated ceiling effects. Histograms for each vowel per group, and boxplots are provided in Appendices O, P, and R.

Skewness and kurtosis values mostly fell within the -2 to +2 range, indicating that the data was approximately normally distributed. Two exceptions slightly exceeded this range; namely, kurtosis values for /ʌ/ in both groups were just above +2. However, these values remain within an acceptable margin, as some researchers suggest that the normality threshold can be extended to ±3 (Kim, 2013).

Table 9

Descriptive Statistics per Vowel and Group in the Perception Task

Vowel	Group	Mean	SD	Skewness	Kurtosis
/iː/	Whole Sample	0.60	0.26	-0.10	-0.76
	Italian	0.60	0.27	-0.09	-0.91

	Spanish	0.59	0.25	–0.14	–0.48
/ɪ/	Whole Sample	0.63	0.29	-0.33	-0.97
	Italian	0.64	0.29	–0.54	–0.69
	Spanish	0.59	0.29	+0.18	–1.48
/ɑː/	Whole Sample	0.69	0.29	-0.79	-0.26
	Italian	0.70	0.29	–0.88	–0.09
	Spanish	0.64	0.30	–0.56	–0.68
/æ/	Whole Sample	0.81	0.18	-0.69	-0.29
	Italian	0.82	0.18	–0.89	+0.44
	Spanish	0.78	0.17	–0.25	+0.12
/ʌ/	Whole Sample	0.86	0.20	-1.76	2.73
	Italian	0.86	0.21	–1.75	+2.41
	Spanish	0.87	0.20	–1.69	+2.56
/ɒ/	Whole Sample	0.68	0.29	-0.47	-0.82
	Italian	0.64	0.30	–0.63	–0.86
	Spanish	0.77	0.26	–0.43	–0.81
/əʊ/	Whole Sample	0.81	0.24	-1.37	1.62
	Italian	0.81	0.23	–1.30	+1.54
	Spanish	0.82	0.26	–1.39	+1.29

4.2 RQ1 Which L2 vowels pose the greatest difficulty for L1 Italian and L1 Spanish learners of English in (a) perception and (b) production?

To address RQ1 and identify the vowels that are most problematic for L1 Italian and L1 Spanish learners of English in perception and production, a repeated measures ANOVA was conducted for both tasks. Sections 4.2.1 and 4.2.2 will present the analyses conducted on the perception data, followed by Sections 4.2.3 and 4.2.4, which will discuss the analyses run on the production data.

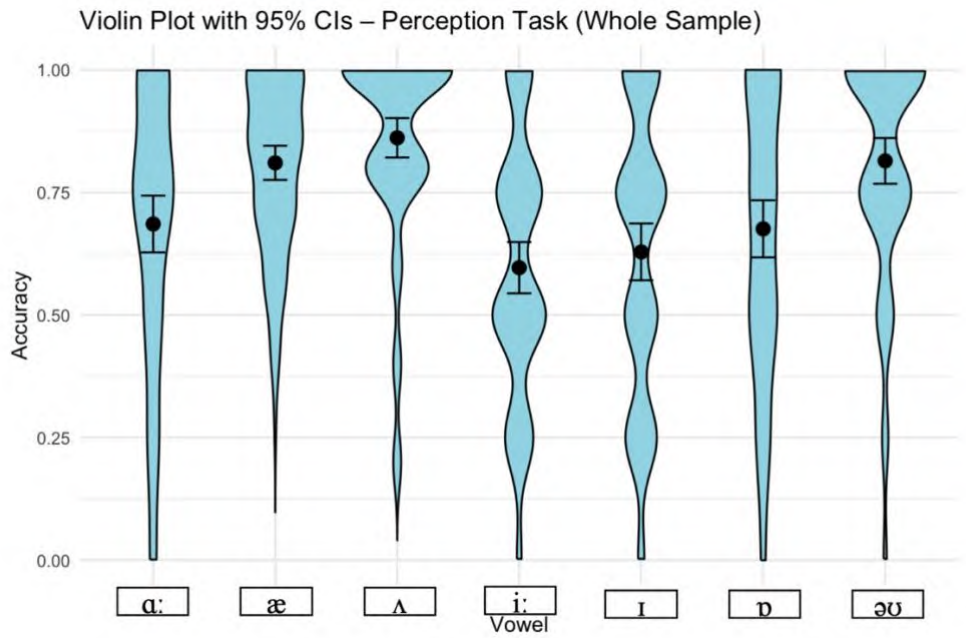
4.2.1 Perception Data – Repeated Measures ANOVA

To assess whether perception accuracy varied across the 7 English vowels tested by the Perception Task, a repeated measures ANOVA was conducted on the whole sample. Mauchly's test revealed that the assumption of sphericity was violated, $W = 0.567$, $p < .001$; therefore, Greenhouse-Geisser corrections were applied ($\varepsilon = 0.85$), resulting in an overall p-value that was still significant ($p < .001$). The ANOVA revealed a significant main effect of vowel $F(6, 600) = 21.45$, $p < .001$, which suggests that perception accuracy differed significantly across the 7 vowels. The partial eta squared (η^2) was 0.177, which, according to standard benchmarks ($\eta^2 > 0.14$), suggests a large effect size. This indicates that vowel type had a strong influence on participants' perception accuracy, accounting for approximately 17.7% of the variance in scores.

Post-hoc comparisons using Tukey adjustments revealed several significant differences in perception accuracy across vowel contrasts. Specifically, participants were significantly more accurate in perceiving /æ/ compared to /ɑː/ ($p = .0016$), and /ʌ/ compared to /ɑː/ ($p < .0001$). Additionally, perception accuracy for /ʌ/ was significantly higher than for /ɒ/ ($p < .0001$), and /əʊ/ was perceived significantly more accurately than /ɒ/ ($p = .0002$). However, a few contrasts did not reach statistical significance. Specifically, no significant differences were found between /æ/ and /ʌ/ ($p = .6671$), /ʌ/ and /əʊ/ ($p = .7474$), or /iː/ and /ɪ/ ($p = .9486$), which means that these vowels were perceived with similar levels of accuracy by learners. Figure 7 below displays the distribution of perception accuracy scores for each vowel using a violin plot with superimposed 95% confidence intervals.

Figure 7

Violin Plot of Perception Accuracy Across Vowels for the Whole Participant Sample



Note. See Appendix S for the corresponding point-and-line plot.

4.2.2 Perception Data – 2x7 Repeated Measures ANOVA

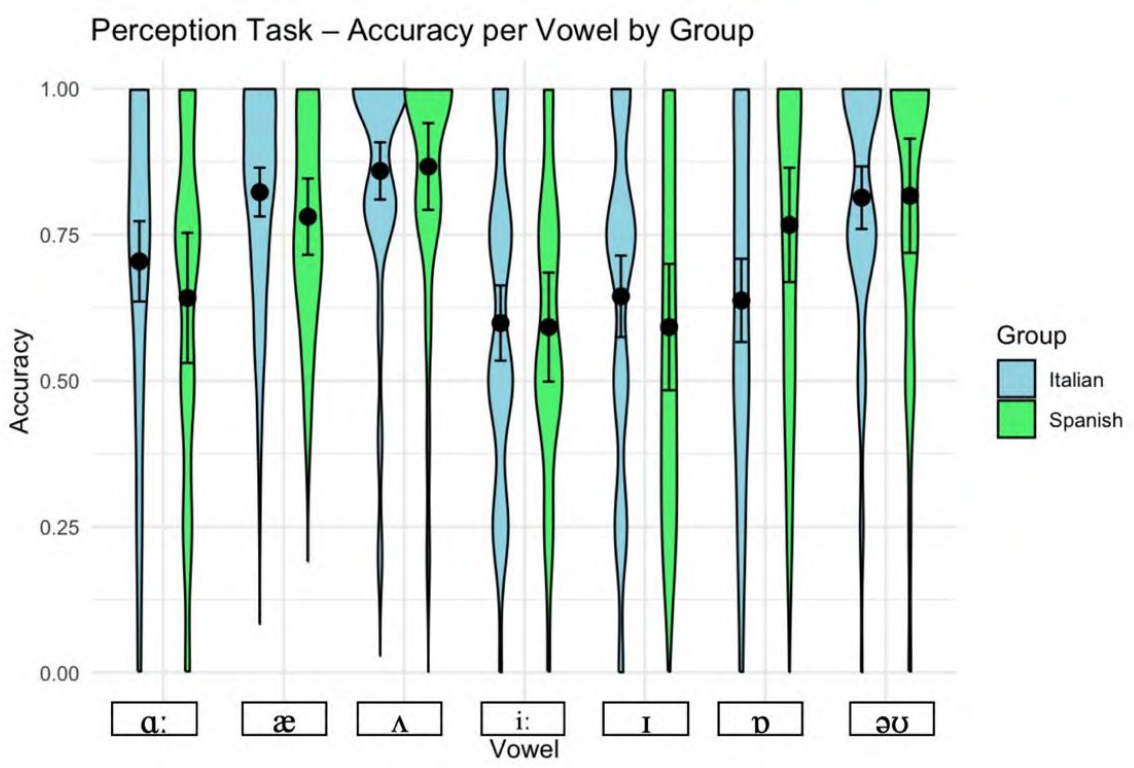
To further examine the effect of L1 background on perception accuracy, a 2x7 repeated measures ANOVA was conducted, with vowel as a within-subject factor (7 levels) and group (Italian vs Spanish) as a between-subjects factor. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($p < .001$), and the Greenhouse-Geisser correction was applied ($\varepsilon = 0.848$).

Results revealed a significant main effect of vowel ($F(6, 594) = 18.33$, $p < .001$), confirming that perception accuracy varied across vowel categories. This finding supports the results of the previous ANOVA and reinforces the notion that some vowels posed greater challenges than others for learners. The effect of group was not statistically significant ($F(1, 99) = 0.11$, $p = .917$), indicating that overall perception accuracy did not significantly differ between the Italian and Spanish groups. The interaction between vowel and group was also non-significant ($F(6, 594) = 1.79$, $p = .099$), which suggests that the pattern of vowel difficulty was broadly consistent across the two groups. For this reason, no post-hoc comparisons were conducted. Effect size calculations further confirm these findings: vowel type accounted for approximately 15.6% of the variance ($\eta^2 = .156$), a large effect, while group and the

vowel x group interaction accounted for less than 2% of the variance each ($\eta^2 = .0178$), both considered small effects.

Figure 8

Violin Plot of Perception Accuracy Across Vowels for the Italian and Spanish Learner Groups



Note. See Appendix T for the corresponding point-and-line plot.

4.2.3 Production Data – Repeated Measures ANOVA

To assess whether L2 vowel production accuracy varied across the 4 vowels included in the Read-Aloud Task, a repeated-measures ANOVA was conducted for the whole participant sample. Although the assumption of normality was violated for the Spanish group's production of /æ/ (see Section 4.1.3), this vowel was retained in the repeated-measures ANOVA based on empirical evidence supporting the robustness of this test under similar conditions (Blanca et al., 2017; Schmider et al., 2010). Given the fact that there was only one such violation, for one of the two groups in one condition, and in the interest of preserving the full range of learner responses in the dataset, the vowel was retained in the analysis.

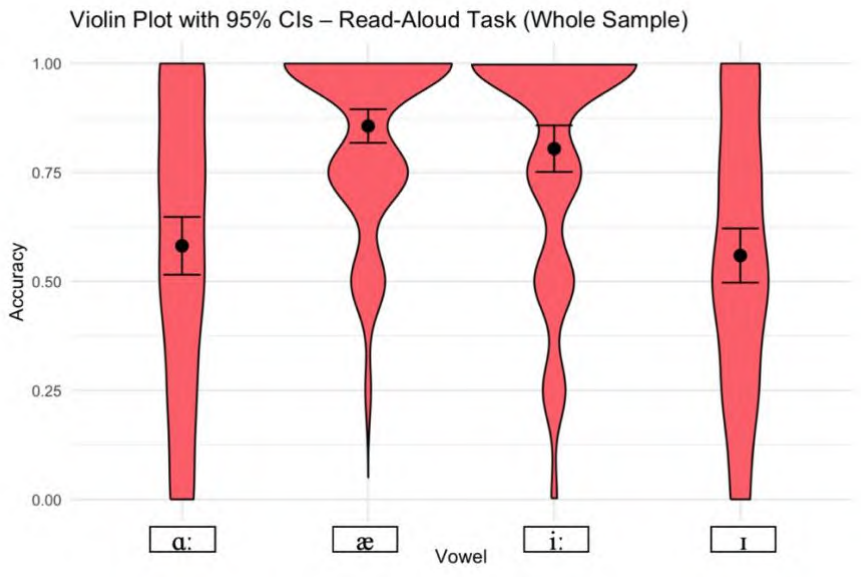
A significant main effect of vowel was found, $F(3, 300) = 35.84$, $p < .001$, indicating that production accuracy differed significantly depending on vowel type. As Mauchly's test of sphericity indicated a

violation of the sphericity assumption ($W = 0.6245$, $p < .001$), Greenhouse–Geisser corrections were applied ($\varepsilon = 0.7827$), with adjusted results still yielding statistical significance: $F(2.35, 235.6) = 35.84$, $p < .001$. The partial eta squared was $\eta^2 = 0.2639$, a large effect size. More than a quarter (26.4%) of the variance in vowel production scores can be explained by vowel identity, further supporting the notion that some vowels are consistently more challenging to produce than others for L2 learners.

Post-hoc comparisons with Tukey adjustments revealed that the participants were significantly more accurate in producing /æ/ compared to /ɑː/ ($p < .0001$), and /iː/ compared to /ɪ/ ($p < .0001$). These results highlight that, within this sample, /ɑː/ and /ɪ/ posed greater challenges in production than their counterparts. The violin plot in Figure 9 illustrates the distribution of accuracy scores across the four target vowels, clearly showing that the vowels /æ/ and /iː/ were produced more accurately than /ɪ/ and /ɑː/.

Figure 9

Violin Plot of Production Accuracy Across Vowels for the Whole Participant Sample



Note. See Appendix U for the corresponding point-and-line plot.

4.2.4 Production Data – 2x4 Repeated Measures ANOVA

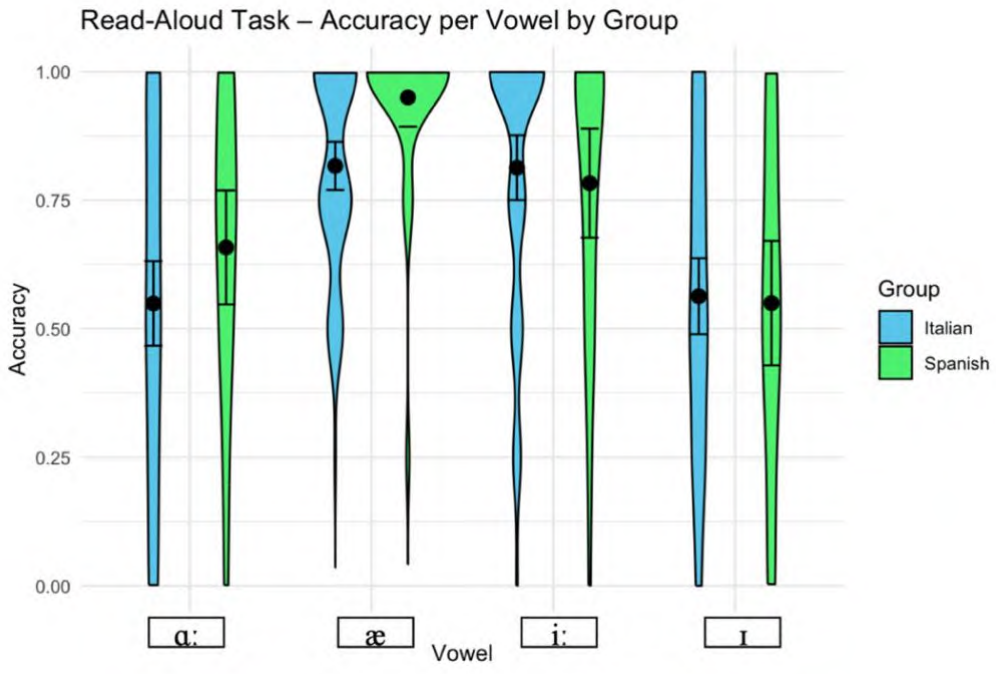
A 2×4 repeated measures ANOVA was conducted to examine whether the production accuracy of the 4 English vowels tested in the Read-Aloud Task varied significantly between Italian and Spanish learners. The within-subjects factor was vowel (4 levels), and the between-subjects factor was L1 group (Italian vs Spanish). Results revealed a significant main effect of vowel, F(2.28, ~225.7) = 31.86,

$p < .001$, with a large effect size ($\eta^2 = 0.2435$), indicating that vowel identity had a strong influence on production accuracy across both groups. The sphericity assumption was violated, $W = 0.5746$, $p < .001$, and Greenhouse-Geisser corrections were applied ($\varepsilon = 0.761$).

There was no significant main effect of group, $F(1, 99) = 1.64$, $p = .203$, $\eta^2 = 0.016$, suggesting that overall production accuracy did not significantly differ between Italian and Spanish learners. Additionally, the interaction between group and vowel was not statistically significant either, $F(2.28, \sim225.7) = 2.28$, $p = .097$, $\eta^2 = 0.023$, meaning that both groups followed similar patterns of vowel difficulty. These results support the findings from the previous ANOVA and confirm that vowel identity was the strongest predictor of performance. Since the effects of group and the group x vowel interaction were not significant, no post-hoc analyses were conducted. Figure 10 displays the violin plot with 95% confidence intervals for each vowel by group.

Figure 10

Violin Plot of Production Accuracy Across Vowels for the Italian and Spanish Learner Groups



Note. See Appendix V for the corresponding point-and-line plot.

4.3 RQ2 Does proficiency affect which vowels learners find problematic in (a) perception and (b) production?

This section explores the relationship between learners' proficiency levels and their performance in the perception and production of English vowels. To address RQ2, LexTALE scores were used as a predictor variable in a series of simple linear regression analyses, with vowel accuracy scores from the Picture-Selection Task and the Read-Aloud Task as outcome variables. Section 4.3.1 will present the findings for the perception task, whereas those related to the production task will be discussed in Section 4.3.2. All regression coefficients reported (b) are unstandardised estimates.

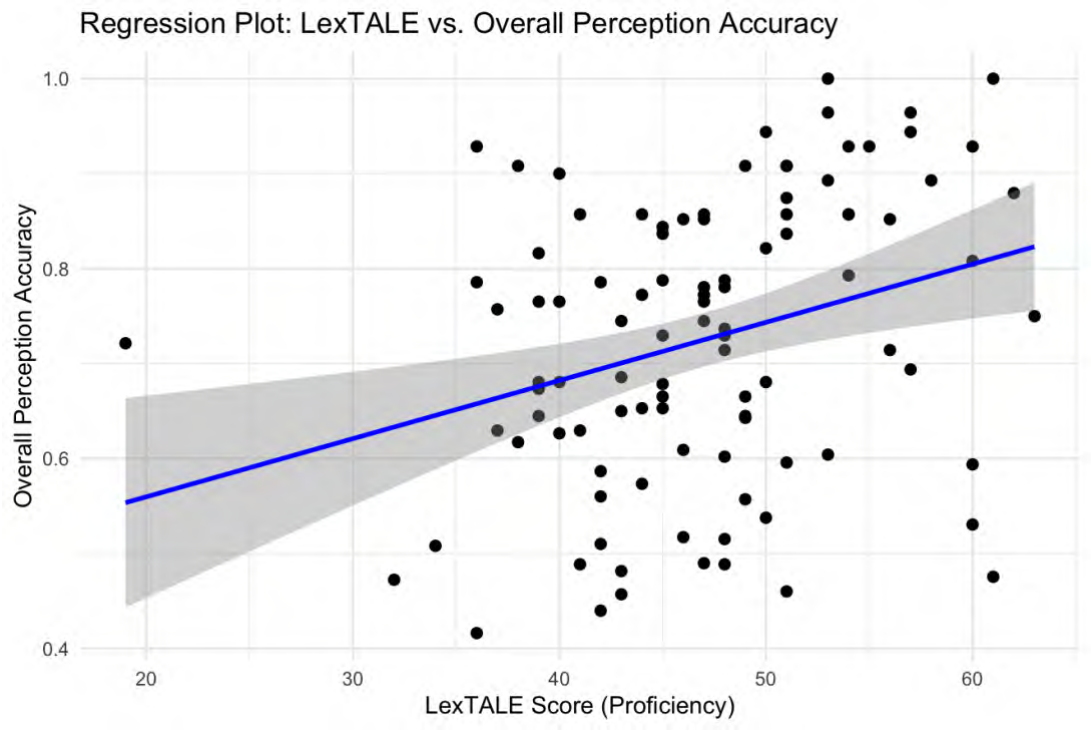
Prior to running the regression analyses, it was necessary to establish whether there was a correlation between the two variables. To do so, the assumption of normality for each variable had to be assessed. As discussed before, the LexTALE scores were found to be normally distributed, but the vowel accuracy scores were not. Therefore, Spearman's rank correlation tests were conducted instead of Pearson's correlations, as the latter requires the assumption of normality. These correlation tests were used to determine whether a significant relationship existed between proficiency and vowel performance, which would then justify running linear regressions on those pairs. All normality checks, including histograms, Shapiro-Wilk tests, and Q-Q plots, were conducted for each variable across both tasks, and the full set of visualisations is included in Appendix W.

4.3.1 Relationship Between Proficiency and Perception Accuracy

To investigate whether proficiency is associated with learners' ability to perceive L2 vowel contrasts, a Spearman's rank correlation was first conducted between participants' LexTALE scores and their overall perception accuracy scores from the perception task. The results showed a statistically significant moderate positive correlation, $\rho = 0.34$, $p = .0005$, indicating that higher LexTALE scores were associated with higher perception accuracy scores. This supports the idea that receptive vocabulary proficiency is meaningfully related to the ability to perceive L2 vowel contrasts. Following this, a simple linear regression analysis was performed with LexTALE as the predictor and overall perception accuracy as the outcome variable. The regression model was significant, $F(1, 99) = 10.17$, $p = .0019$, with LexTALE scores accounting for approximately 9.3% of the variance in perception accuracy ($R^2 = 0.093$). The unstandardised regression coefficient for LexTALE was $b = 0.0061$, $t(99) = 3.19$, $p = .0019$. These findings indicate that learners with higher levels of proficiency tend to perform more accurately in the perception of L2 English vowel contrasts.

Figure 11

Regression Scatterplot for LexTALE and Overall Perception Accuracy Scores



Note. See Appendix Y1 for regression scatterplots per-vowel.

In addition to the overall perception score, correlation analyses were conducted for each vowel in the Picture-Selection Task to explore whether the relationship between proficiency and perception accuracy varied by vowel. All correlation analyses were run using Spearman's rank correlation, as the vowel accuracy scores were not normally distributed. The corresponding scatterplots are included in Appendix X.

For the vowel /iː/, a significant positive correlation was found between LexTALE scores and perception accuracy, ρ = 0.224, p = .025. Although this relationship is weak, it suggests that learners with higher receptive vocabulary knowledge were slightly more accurate in perceiving /iː/. Similarly, perception accuracy for /ɑː/ also significantly correlated with LexTALE scores, ρ = 0.238, p = .017, indicating a weak-to-moderate positive association. This means that learners with higher proficiency levels tended to perform better in perceiving /ɑː/. For /ʌ/, a comparable pattern emerged: ρ = 0.240, p = .015, again showing a statistically significant weak-to-moderate positive relationship between proficiency and accuracy in perceiving this vowel. The strongest correlation was found for /ɒ/, with ρ = 0.329, p < .001. This result indicates a moderate positive correlation, suggesting that higher LexTALE scores were associated with greater accuracy in perceiving /ɒ/, more strongly than for any

other vowel. On the other hand, no significant correlations were observed for /ɪ/ (ρ = 0.159, *p* = .112), /æ/ (ρ = 0.187, *p* = .061), or /əʊ/ (ρ = 0.046, *p* = .649). While all three showed positive trends, their p-values exceeded the .05 threshold, indicating no reliable evidence of association in this sample.

A series of simple linear regressions was conducted to investigate whether LexTALE proficiency scores could predict participants' perception accuracy for individual English vowels that previously showed significant correlations. The results of these models are summarised in Table 10 below.

Table 10

Summary of Simple Linear Regression Results Predicting Vowel Perception Accuracy from LexTALE Scores

Vowel	b	t(99)	F(1,99)	*p*-value	R^2
/iː/	0.0064	1.83	3.34	.0706	.033
/ɑː/	0.0094	2.46	6.07	.0155	.058
/ʌ/	0.0070	2.66	7.08	.0091	.067
/ɒ/	0.0130	3.48	12.13	.0007	.109

As shown in Table 10, the model for /iː/ was not statistically significant, though the slope was positive. LexTALE scores explained just 3.3% of the variance in perception accuracy, suggesting only a weak, non-significant relationship between proficiency and perception of /iː/. By contrast, LexTALE scores significantly predicted perception accuracy for /ɑː/, /ʌ/, and /ɒ/, with the strongest effect observed for /ɒ/. Together, these findings indicate that proficiency predicts perception accuracy for some L2 English vowels, with the strongest effects observed for /ɒ/, /ʌ/, and /ɑː/.

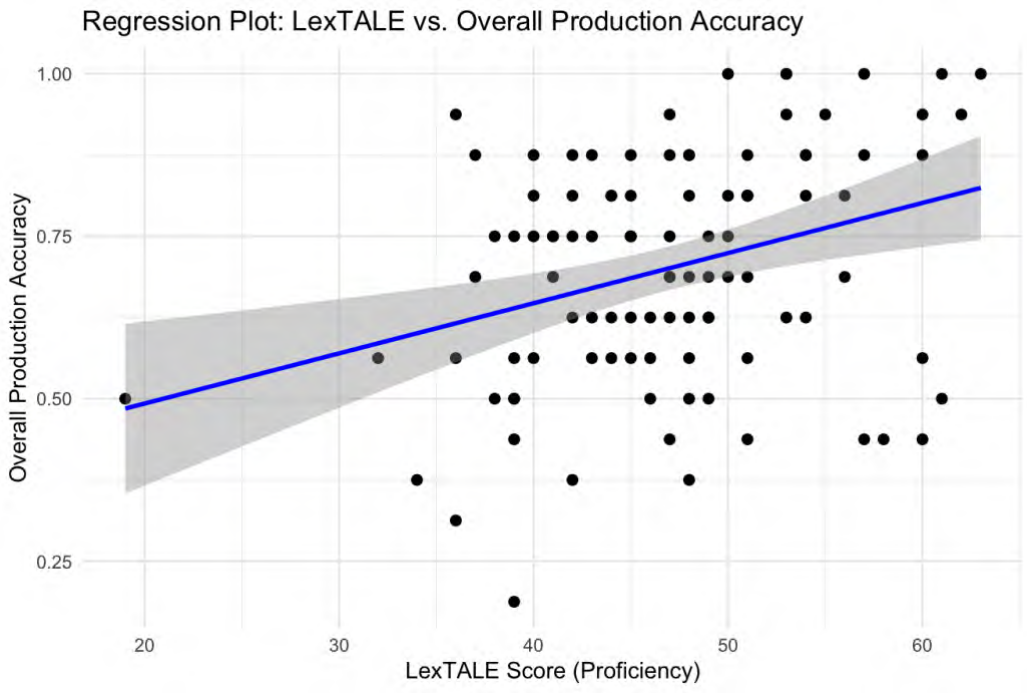
4.3.2 Relationship Between Proficiency and Production Accuracy

To explore whether participants' proficiency predicted their accuracy in producing English vowels, a correlation analysis was first conducted between LexTALE scores and participants' overall production accuracy. As in the previous section, a Spearman's rank correlation was used due to the non-normal

distribution of the vowel accuracy data. The analysis revealed a statistically significant positive correlation between LexTALE and overall Production accuracy, ρ = 0.30, p = .002 (see Appendix X for visualisation). To further investigate this relationship, a simple linear regression was conducted with LexTALE scores as the predictor and overall production accuracy as the outcome variable. The model was statistically significant, $F(1, 99) = 11.47$, p = .001, indicating that LexTALE scores reliably predicted participants' performance in the Read-Aloud Task (b = 0.0077, $t(99) = 3.39$). The model explained 10.4% of the variance in production scores (R^2 = .104), representing a small-to-moderate effect size that is meaningful within SLA research, where variability in production performance is expected. Together, these results indicate that higher proficiency levels are associated with greater accuracy in L2 vowel production.

Figure 12

Regression Scatterplot for LexTALE and Overall Production Accuracy Scores



Note. See Appendix Y2 for regression scatterplots per-vowel.

To explore whether proficiency influenced learners' ability to produce specific English vowel contrasts, correlation analyses were conducted between LexTALE scores and participants' accuracy scores for each vowel in the Read-Aloud Task. Spearman's rank correlation was used due to the non-normal distribution of the vowel-specific accuracy data. All correlation scatterplots are included in Appendix X. The results revealed that the correlation between LexTALE scores and production accuracy of /iː/

was statistically significant, ρ = 0.215, *p* = 0.031. Although the strength of the relationship was weak, this finding suggests that higher receptive vocabulary proficiency was positively associated with better production of this vowel. A similarly weak but statistically significant correlation was found for the short vowel /ɪ/, with ρ = 0.221 and *p* = 0.026, again indicating that increased proficiency was associated with more accurate production.

In contrast, the correlation between LexTALE scores and production accuracy of /ɑː/ did not reach statistical significance. While the observed ρ = 0.160 indicated a weak positive trend, the p-value of 0.110 suggests that this relationship was not reliable in the current sample. Lastly, a small to moderate significant correlation was observed between LexTALE scores and production accuracy of /æ/, with ρ = 0.253 and *p* = 0.011. This result supports the view that learners with higher proficiency were more successful in producing this vowel. Overall, the strength and presence of the correlations varied depending on the vowel, indicating that some contrasts may be more sensitive to learners' proficiency than others. While /iː/, /ɪ/, and /æ/ showed reliable associations with LexTALE scores, /ɑː/ did not.

Following the correlation analyses, simple linear regressions were conducted to examine whether LexTALE proficiency scores significantly predicted participants' production accuracy for the individual vowels that had previously shown statistically significant correlations. These included /iː/, /ɪ/, and /æ/.

Table 11

Summary of Simple Linear Regression Results Predicting Vowel Production Accuracy from LexTALE Scores

Vowel	**b**	**t(99)**	**F(1,99)**	***p*-value**	**R^2**
/iː/	0.0077	2.17	4.72	.0321	.046
/ɪ/	0.0093	2.26	5.09	.0263	.049
/æ/	0.0056	2.19	4.79	.0309	.046

As illustrated in Table 11, each model reached statistical significance with LexTALE scores positively predicting production accuracy in all cases; however, the explained variance was small. Together, these regression results demonstrate that proficiency is a statistically significant predictor of production accuracy for /iː/, /ɪ/, and /æ/, although the overall variance explained remains limited.

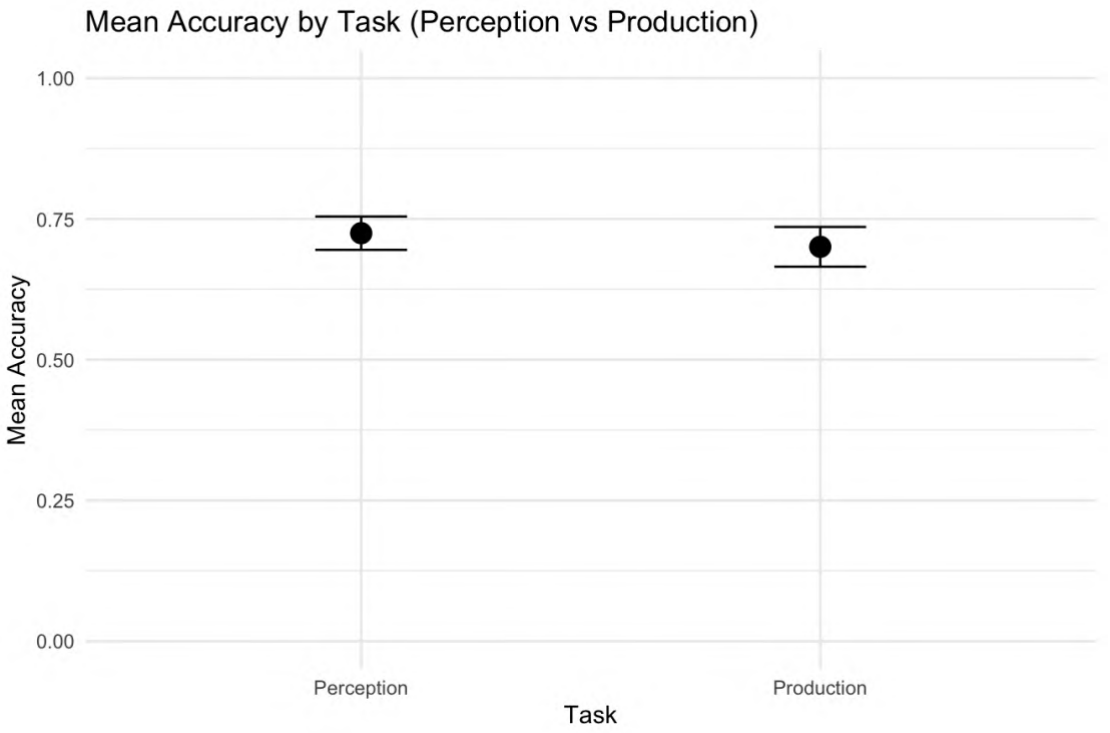
4.4 RQ3 Is the learners' perception of L2 vowels more accurate than their production of the same vowels?

To determine whether learners were more accurate at perceiving L2 vowels than producing them, a paired-samples t-test was conducted comparing the overall accuracy scores from the Perception Task and the Read-Aloud Task. Before running the t-test, the normality assumption was tested by applying a Shapiro-Wilk test on the difference scores, which were calculated for each participant by subtracting their mean production accuracy score from their mean perception accuracy score. The results showed that the data did not significantly deviate from normality, $W = 0.982$, $p = .185$. A Q-Q plot (Appendix Z) further confirmed the approximately normal distribution of the difference scores. Therefore, the assumption of normality was met, and a paired-samples t-test was deemed appropriate.

The t-test revealed that the mean perception accuracy (M = 0.72, SD = 0.15) was slightly higher than the mean production accuracy (M = 0.70, SD = 0.19). However, this difference was not statistically significant, t(100) = 1.53, p = .129, and the 95% confidence interval for the mean difference included zero ([-0.007, 0.056]). This indicates that there is no reliable evidence that participants performed better in perception than in production overall. The mean difference between tasks was small (0.024), suggesting a minimal advantage for perception, but not one that reached statistical significance. To further assess the magnitude of this difference, Cohen's *d* was calculated and yielded a value of 0.15. This falls within the range of negligible effect sizes according to conventional benchmarks. The 95% confidence interval for the effect size ([-0.043, 0.333]) also includes zero, reinforcing the conclusion that the observed difference between tasks is not meaningfully large or robust. A visualisation of the mean accuracy scores and their confidence intervals is provided in Figure 13. This plot illustrates the close overlap in performance between the two tasks, further supporting the statistical results.

Figure 13

Mean Perception and Production Accuracy Scores Across All Vowels with 95% Confidence Intervals



5. Discussion

This study aimed to investigate how L1 Italian and L1 Spanish learners of English perceive and produce selected English vowel contrasts, and whether their performance is affected by their level of L2 proficiency. By focusing on two typologically similar Romance languages, the study explored how subtle differences in learners' L1 phonological systems can influence the development of L2 interlanguage phonology.

Moreover, the study examined the relationship between perception and production, in light of ongoing debates in the literature. Some researchers argue that the two skills are interdependent and that better perception supports production (Flege, 1995), whereas others provide evidence for a dissociation between them, showing that learners may perceive contrasts they are not yet able to produce (De Leeuw et al., 2021; Kartushina & Frauenfelder, 2014). This study contributes to this debate by comparing perception and production scores for the same contrasts across learner groups.

Furthermore, the present dissertation also tested the usefulness of CA when used in tandem with EA. While CA makes predictions about which contrasts are likely to pose difficulties based on L1-L2 differences, EA offers an empirical account of the errors learners make. Therefore, one of this study's key aims was to evaluate whether learner performance data would confirm the predictions generated through CA. This chapter begins by stating the predictions associated with each research question. It then discusses and interprets the results of the perception and production tasks against these predictions and situates them within the existing literature.

5.1 RQ1: Which L2 vowels pose the greatest difficulty for L1 Italian and L1 Spanish learners of English in (a) perception and (b) production?

Based on the CA, it was predicted that both L1 Italian and L1 Spanish learners would encounter difficulties with English vowels absent from their native phonological inventories. In the perception task, the contrasts /ɪ/-/iː/, /æ/-/ʌ/, /ʌ/-/ɑː/, and /ɒ/-/əʊ/ were expected to be challenging, with particular difficulty anticipated for /ɪ/, /æ/, /ɑː/, and /ʌ/ due to the lack of clear equivalents in the L1s (Escudero & Boersma, 2004). In production, the vowels analysed were /iː/, /ɪ/, /ɑː/, and /æ/, among which /ɪ/, /ɑː/, and /æ/ were predicted to pose the greatest difficulty due to their absence in the L1 vowel space and the likelihood of assimilation to similar native categories (Bohn & Flege, 1990). Given the typological similarity between Italian and Spanish, both of which are Romance languages with comparable vowel systems, it was also expected that there would be no significant group effect. That is, both learner groups were predicted to exhibit similar patterns of difficulty in perception and production.

5.1.1 (a) Perception

The perception results showed that the vowels perceived with the lowest accuracy were /ɑː/ and /ɒ/, whereas /æ/, /ʌ/, and /əʊ/ were perceived significantly more accurately. Specifically, /æ/ was perceived significantly better than /ɑː/ ($p = .0016$), and /ʌ/ was perceived significantly better than both /ɑː/ ($p < .0001$) and /ɒ/ ($p < .0001$). Additionally, /əʊ/ was perceived more accurately than /ɒ/ ($p = .0002$). Crucially, no significant effect of L1 group emerged, confirming the prediction that Italian and Spanish learners would perform similarly in their perception of the examined contrasts. These findings largely confirm the CA-based predictions. The low perception accuracy for /ɑː/ was expected and aligns with literature reporting its consistent difficulty among L2 learners whose L1 lacks a back open vowel and vowel length contrast (Calvo Benzies, 2017; Wheelock, 2016).

For instance, Cebrian (2019), using a perceptual similarity task, found that English /æ/ and /ʌ/ were perceived more accurately than /ɑː/ by L1 Spanish listeners. This pattern was attributed to the fact that /æ/ and /ʌ/ were both predominantly assimilated to the same L1 category /a/, whereas /ɑː/, although also mapped to /a/, received significantly lower assimilation scores. Within the Perceptual Assimilation Model for L2 learners (PAM-L2; Best & Tyler, 2007), this reflects *category goodness assimilation* for /æ/ and /ʌ/, which may have facilitated perceptual discrimination. According to PAM-L2, L2 sounds are perceived in relation to existing L1 categories, and the ease or difficulty of perceiving L2 contrasts depends on how each sound maps onto the native phonological system. When two L2 vowels are assimilated to the same L1 category but differ in perceived quality, discrimination is possible but not optimal. These results resonate with the current findings, despite methodological differences: while Cebrian's (2019) study was based on perceived similarity ratings, the present error analysis similarly suggests that learners found /æ/ and /ʌ/ more perceptually accessible than /ɑː/, potentially due to shared assimilation to the L1 /a/ category but differing perceptual salience and phonetic features.

By contrast, the poor perception of /ɒ/ is an unexpected result. Given that both Italian and Spanish have a mid-back rounded vowel /o/, /ɒ/ should not pose significant perceptual or articulatory difficulty. Participants produced it successfully, confirming that its articulation is not problematic. However, in the perception task, /ɒ/ was tested exclusively in contrast with the diphthong /əʊ/ (e.g., in minimal pairs like *cot–coat*), a contrast that was predicted to be problematic for L1 Spanish and Italian learners, according to the CA and previous literature (Coe, 2001; Duguid, 2001). From this perspective, the low perception accuracy for /ɒ/ aligns with CA predictions not because of inherent difficulty with the vowel itself, but due to the perceptual challenge posed by the contrast /ɒ/–/əʊ/, where both may be mapped onto the L1 /o/ (Best & Tyler, 2007).

What is surprising, however, is that participants had lower perception scores for /ɒ/ than for /əʊ/. Considering previous evidence of poor production of the Southern Standard British English (SSBE) diphthong /əʊ/ by L1 Italian learners of English (Wheelock, 2016) and L1 Spanish learners of English (Calvo Benzies, 2017), and assuming that perception precedes production, /əʊ/ was predicted to be less accurately perceived than /ɒ/. Multiple factors may account for this unexpected finding. One possible explanation is the occurrence of a Type I error; that is, the incorrect rejection of a true null hypothesis. In this case, it would mean reporting a statistically significant difference in the perception accuracy of /ɒ/ and /əʊ/ when, in fact, such a difference may not exist (Winter, 2019). Given that the alpha level was set at 0.05, there remains a 5% probability of committing a Type I error, which is plausible considering the limited size of the L1 Spanish group and the multiple comparisons conducted.

Another surprising finding concerns the vowel /ɪ/, which did not emerge as significantly problematic in the perception task. This was unexpected given extensive prior evidence that /ɪ/ is one of the most challenging L2 vowels for both L1 Spanish and Italian learners to perceive accurately. For instance, Cebrian et al. (2021) and Veiga Pérez (2017) report consistent perceptual confusion between /ɪ/ and /iː/ among L1 Spanish learners, attributed to the lack of a corresponding short lax vowel in the L1 system. Similarly, Flege and MacKay (2004) and Flege (1995) note that L1 Italian learners tend to assimilate /ɪ/ to their native /i/, leading to reduced perceptual distinctiveness.

The absence of perceptual difficulty in the present study may be due to task design factors that may have increased the salience of /ɪ/. For instance, the use of images rather than written words, whilst avoiding any orthography-based effects, may have reduced cognitive load and encouraged participants to focus more directly on the aural stimulus, potentially enhancing perceptual accuracy (Morett, 2019). Moreover, the speakerʂ articulation may have made the vowels unusually salient through clear enunciation, which could have further facilitated the identification of /ɪ/. Hyperarticulated speech has been shown to enhance phonetic sensitivity in both native and non-native listeners (Uther et al., 2012). Alternatively, learners may have received prior instruction on the /ɪ/-/iː/ contrast. Regardless of the cause, this finding underscores the importance of not assuming perceptual difficulty solely based on theoretical predictions, and it calls for further investigation into the conditions under which /ɪ/ may become more or less perceptually accessible to L2 learners.

5.1.2 (b) Production

Production results revealed that /ɑ/ and /ɪ/ were produced significantly less accurately than their counterparts. Learners were significantly more accurate in producing /æ/ than /ɑː/ ($p < .0001$), and /iː/ than /ɪ/ ($p < .0001$). Again, no significant effect of group was found. These findings are consistent with

the CA. The vowel /ɪ/ is absent in both L1 Italian and Spanish phonemic inventories, and learners often assimilate it to the long /iː/, resulting in reduced contrast and production errors. Similarly, /ɑː/ is likely assimilated to the L1 /a/, which is typically central or front and short in both L1s, leading to inaccurate realisations. Interestingly, /æ/ was produced more accurately than /ɑː/, despite also lacking a direct L1 equivalent. One possible explanation is its relatively front articulatory position, which may make it more perceptually salient and easier to approximate using a modified L1 /a/ or /e/ sound. While CA predicted /æ/ to be problematic, it also allowed for the possibility that it might be easier to acquire than /ɑː/.

These findings also resonate with previous empirical evidence from EA studies. For instance, Wheelock (2016) found that /ɑː/ had the highest error rate (80.6%) among L1 Italian learners, and it was mispronounced at high rates even by higher-performing participants. Similarly, Calvo Benzies (2017) reported that /ɑː/ was one of the most frequently mispronounced vowels by L1 Spanish learners, and Conklin et al. (2019) reported lower error rates for /æ/ than /ɑː/, which was the most confusable vowel in production for L1 Colombian Spanish learners of English. In addition, Veiga Pérez (2017) identified /ɪ/ as one of the weakest vowels in learner production, with frequent confusion with /iː/, confirming the persistent difficulty of this contrast. Together, these studies corroborate the present findings and reinforce the importance of targeted instruction in these vowel contrasts.

Overall, these findings validate the joint use of contrastive and error analyses: while CA effectively anticipated general areas of difficulty, EA provided a finer-grained understanding of performance variation across specific contrasts and tasks.

5.2 RQ2: Does proficiency affect which vowels learners find problematic in (a) perception and (b) production?

It was predicted that learners with higher proficiency levels would be better at perceiving and producing L2 English vowels. This expectation aligns with previous research indicating that increased L2 proficiency tends to enhance phonological processing, possibly due to richer lexical representations and more extensive exposure to L2 input (Vafaee & Suzuki, 2020). Before interpreting the findings, it is important to clarify that proficiency was operationalised using the LexTALE test in this study (Lemhöfer & Broersma, 2012). This is a receptive vocabulary test that has been shown to correlate well with general proficiency but remains a proxy measure (Puig-Mayenco et al., 2023). As such, it captures only one dimension of L2 competence, receptive lexical knowledge, and may not fully reflect learners' abilities in English pronunciation.

5.2.1 (a) Perception

The results partly support the initial prediction. A statistically significant moderate correlation emerged between LexTALE scores and overall perception accuracy ($\rho = 0.34$, $p = .0005$), and a regression model confirmed that proficiency accounted for just over 9% of the variance ($R^2 = .093$). In other words, learners with stronger receptive vocabulary knowledge tended to perceive L2 vowel contrasts more accurately. This aligns with the idea that greater lexical knowledge contributes to sharper phonological discrimination skills in L2 learners (Vandergrift, 2006).

However, item-level analyses revealed that this relationship was not uniform across all vowels. Statistically significant regression models were found only for /ɑː/, /ʌ/, and /ɒ/, with /ɒ/ showing the strongest effect ($\rho = 0.329$, $p < .001$). /ɑː/ and /ʌ/ have previously been identified as especially problematic for L1 Italian and Spanish learners due to their articulatory features and limited overlap with L1 vowel spaces (Conklin et al., 2019; Wheelock, 2016). Thus, the fact that higher proficiency seems to facilitate perceptual discrimination of these two challenging vowels is not surprising. On the other hand, /ɒ/, which was not expected to be perceptually daunting for participants, emerged as particularly sensitive to proficiency level. As discussed in Section 5.1.1, this may be due to the perceptual challenge posed by its contrast with /əʊ/ in the perception task; perhaps, task-specific factors such as the use of minimal pairs and lexical familiarity may have interacted with proficiency to shape perceptual outcomes. That is, learners with higher vocabulary knowledge may have been better able to access detailed lexical representations, helping them discriminate between similar-sounding words like *cot–coat*. On the other hand, lower-proficiency participants may have lacked this lexical support, thus struggling to differentiate these two vowels.

By contrast, proficiency did not significantly predict perception accuracy for /ɪ/, /æ/, or /əʊ/. One reason for this may be that, in the current study, /ɪ/ did not appear to pose perceptual difficulty for the participants. This could explain why no effect of proficiency emerged: if most learners were already perceiving /ɪ/ relatively accurately, vocabulary knowledge may not have played a major role. As for /æ/, it may have shown little variability because learners performed relatively well across the board, which reflects a ceiling effect and, therefore, does not leave much room for proficiency to have an impact. Finally, factors such as speaker accent or phonetic context may have played a bigger role than overall L2 vocabulary knowledge in the perception of the diphthong /əʊ/. The speaker in this study had an SSBE accent, which may have been unfamiliar to some learners, potentially reducing perceptual accuracy (Adedeji, 2022). Additionally, phonetic context, such as surrounding consonants or prosodic features, can affect the salience of diphthongal cues, influencing identification independently of proficiency (Hardison, 2003).

5.2.2 (b) Production

The results for production point in a similar direction. As predicted, there was a significant positive correlation between LexTALE scores and overall vowel production accuracy ($\rho = 0.30$, $p = .002$), and the regression model confirmed that proficiency accounted for about 10% of the variance ($R^2 = .104$). This suggests that more proficient learners tend to be more accurate in producing English vowel sounds, a finding that supports the idea that exposure and lexical growth contribute to more robust phonological representations and greater production accuracy (Flege, 1995; Gurzynski-Weiss, 2022).

But again, the strength of this relationship varied depending on the vowel. Statistically significant associations were found for /iː/, /ɪ/, and /æ/, with regression models showing that proficiency explained a small but consistent proportion of the variance, between 4.6% and 4.9%. These vowels are known to pose challenges for L1 Italian and Spanish learners, but it seems that increased proficiency helps learners fine-tune their articulatory targets, leading to more accurate production. In particular, the vowel /ɪ/, notoriously difficult due to its phonetic proximity to /iː/ and the absence of a clear L1 equivalent, showed a significant correlation with proficiency. This finding aligns with previous studies. For instance, Wheelock (2016) reported that higher-performing L1 Italian learners of English produced /ɪ/ with greater accuracy than their lower-performing peers, suggesting that this vowel may be particularly sensitive to proficiency-driven improvements in phonological encoding and motor control.

By contrast, this pattern did not hold for /ɑː/. Despite being one of the most difficult vowels to perceive and produce for participants, its production accuracy did not significantly correlate with LexTALE scores. One possible explanation is that /ɑː/ presents persistent articulatory difficulties that may not be easily overcome through vocabulary growth alone. As a long, back vowel with no close L1 equivalent, it may require more explicit phonetic instruction or focused production training to be acquired accurately. This interpretation is supported by previous studies such as Wheelock (2016) and Calvo Benzies (2017), where /ɑː/ elicited high error rates even among higher-level learners, indicating that explicit instruction or focused articulatory training may be necessary for accurate acquisition.

5.3 RQ3. Is the learners' perception of L2 vowels more accurate than their production of the same vowels?

Based on previous research suggesting that perception tends to precede and support production (Flege, 1995; Sakai & Moorman, 2018; Thomson & Derwing, 2015), it was predicted that learners would perform significantly better in perception than in production. However, this hypothesis was not supported by the data.

A paired-samples t-test revealed no statistically significant difference between participants' mean perception (M = 0.72, SD = 0.15) and production (M = 0.70, SD = 0.19) scores, t(100) = 1.53, p = .129. Although the mean difference was in the expected direction, with perception scores slightly higher than production, the effect was small (mean difference = 0.024) and not statistically significant. This was further supported by a negligible Cohen's d of 0.15 and a 95% confidence interval for the effect size that crossed zero ([-0.043, 0.333]), suggesting that the result may well be due to chance.

There are several possible explanations for this unexpected finding. One is that the difference may not have been large enough in this sample to reach statistical significance. Another possibility is that participants found the perception task more challenging than anticipated, possibly due to the use of SSBE stimuli. Learners may have performed better with a more familiar accent, such as General American (GA) English, which might better match their exposure or mental representations of English vowels (Escudero & Boersma, 2004).

It is also worth noting that this result only provides a global comparison. When we zoom in on individual vowels, as explored in RQ1 and RQ2, a more nuanced picture emerges: some vowels were perceived more accurately than produced (e.g., /ɪ/), while others showed the opposite pattern. This suggests that, although the group-level difference was not significant, there may still be a dissociation between the two skills depending on the specific contrast. In this sense, the result does not necessarily contradict the idea that perception and production are distinct yet related processes; it simply highlights that their relationship may vary depending on the phoneme, the learner, and the linguistic context. While the overall comparison did not show a significant advantage for perception over production, item-level and task-specific analyses reveal a more complex interplay between the two modalities, one that warrants further exploration in future research.

6. Conclusions

6.1 Implications for EFL Teaching

The findings of this study underline the need to bring explicit phonetic training into the EFL classroom. While pronunciation instruction is often overlooked or reduced to articulatory drills, recent research points to the value of developing learners' awareness of what they can (and cannot yet) perceive and produce. As Saito and Plonsky (2019) explain, there are two main approaches to L2 pronunciation instruction: articulatory-based training, which aims at the simultaneous development of perception and production, and auditory-based training, which first focuses on enhancing learners' ability to discriminate between sounds. The latter draws on the well-established view that accurate perception typically precedes accurate production (Flege, 1995) and is supported by robust empirical evidence showing that perception training alone can lead to significant gains in production accuracy (Thomson, 2011, 2018; Thomson & Derwing, 2016).

The present study contributes to this debate by showing that although perception and production are distinct skills, they are also deeply interconnected. Learners who performed better in the perception task generally showed better production performance, too. This suggests that improving perception may be an effective first step towards better pronunciation. In this regard, Gordon and Arias (2024) argue that one of the most important roles of pronunciation instructors is to raise learners' phonological awareness, an observation that aligns with the present findings. When learners become aware of which contrasts they fail to perceive or produce, this often leads to substantial improvement in both areas. Numerous studies provide empirical evidence for a significant positive impact of perception-led training not only on L2 pronunciation but also on listening comprehension skills (Choe et al., 2020; Khaghaninejad & Maleki, 2015; Trinh et al., 2022; Zhong et al., 2023).

However, perceptual training may not be enough in every case. Some sounds, such as /ɪ/ in this study, continue to be mispronounced even when they were successfully perceived. This indicates that certain vowels may require specific articulatory-based instruction in addition to perceptual training (Sakai, 2016). Pronunciation teachers should, therefore, combine perceptual and articulatory practice depending on the nature of the phoneme considered rather than relying on a single approach. Some segments may benefit from targeted production training, while for others, improved perception may suffice. Adopting a selective, contrast-specific approach can allow teachers to maximise instructional efficiency and avoid spending time on less relevant targets. This is precisely where research-informed

teaching becomes essential. By drawing on CA and EA, informed decisions can be made regarding which L2 sounds to prioritise and focus on in the classroom.

To achieve this goal, several practical strategies can be adopted. One popular technique is shadowing, where learners repeat audio input almost simultaneously with the speaker. Hamada (2016) found that shadowing enhances perception skills across proficiency levels and is especially useful for developing listening comprehension among lower-level learners. A different technique is High Variability Phonetic Training (HVPT), which exposes learners to multiple speakers and varied phonetic contexts for a given L2 sound. It has been shown to improve perception, and often production, by promoting more robust phonemic categories. As Thomson (2018) points out in his review of 32 HVPT studies, this method is particularly effective for adult learners, as it helps them move beyond speaker-dependent listening and build more target-like perceptual categories. The variability introduced by HVPT can also result in increased production accuracy of specific L2 segments and allow learners to apply their perceptual gains in real-world scenarios (Barriuso & Hayes-Harb, 2018). Although primarily a materials-based approach, HVPT can be implemented through curated multi-speaker recordings or specialised training software (Thomson, 2018).

However, implementing these strategies requires a certain degree of phonetic knowledge and pedagogical expertise. Therefore, EFL teachers need better training in phonetics and phonology so that they can effectively assess learners' needs and select the most appropriate instructional solutions. Such training should not be grounded in native speaker norms, but in the broader aim of helping learners become intelligible and communicatively effective in diverse contexts. The goal of pronunciation teaching should not be to sound like a native speaker, but to foster international intelligibility and communicative competence (Smith & Nelson, 1985). Learners should be empowered to speak with a comprehensible, rather than a native-like, accent, and to be intelligible to a range of interlocutors.

6.2 Limitations of This Study and Recommendations for Future Research

Despite its interesting findings and pedagogical implications, this study has some limitations that should be acknowledged. To begin with, although the sample size was appropriate for the design and statistical analyses used, the Spanish group was considerably smaller than its Italian counterpart. As a result, the statistical power for analyses involving only Spanish participants was lower, increasing the risk of Type II errors and making it more difficult to detect meaningful effects within that subgroup. Additionally, because the overall descriptive and inferential results reflect pooled data, the larger size of the Italian group may have disproportionately influenced the group means, limiting the extent to which the findings can be generalised equally to both L1 backgrounds.

Moreover, most Italian participants were from Sardinia, and the most popular dialect among participants was Sardinian Campidanese. Although not initially planned, this regional concentration could have introduced variability not representative of the broader L1 Italian-speaking population. Future studies should control more rigorously for dialectal variation, possibly selecting speakers from a single dialect or including dialect as a variable of analysis (see Escudero et al., 2014, for similar work with Salento speakers).

Another limitation concerns the operationalisation of L2 proficiency. While LexTALE provided a reliable and efficient indication of receptive vocabulary knowledge, it did not capture other key aspects of language competence that could have been relevant to this study's experiments, such as oral accuracy, fluency, and listening skills. Nonetheless, receptive vocabulary is widely recognised as a robust predictor of general L2 proficiency and correlates strongly with listening comprehension (Atas, 2018; Ha, 2021), and with broader skills including speaking, reading, and writing (Miralpeix & Mu ñoz, 2018). As such, the observed associations between proficiency and phonological performance may reflect a meaningful, though partial, component of a more complex interaction.

Furthermore, although participants were informed that the target variety was SSBE, many may have been more familiar with other English varieties, particularly GA, which is a popular variety of English in South America, from which most of the Spanish-speaking participants came. Since vowel realisations often vary across dialects (e.g., /ɑː/ vs. /æ/), this may have influenced participants' perception and production patterns.

Finally, it was not possible to control for numerous individual learner variables. Although background questionnaires gathered information on some (e.g., age, gender, L2 exposure, and L1 dialect use), these factors were not incorporated into the statistical models. This was a deliberate choice based on both the nature of the data collected and the focus of the study. Age and gender were not expected to affect performance, given that all participants were adults and completed tightly controlled tasks (see Flege et al., 1999). L2 exposure could have been a relevant factor, but the measure used was relatively crude, and participants were broadly similar in their language learning contexts, suggesting that any variation in exposure was likely to have a limited effect. As for L1 dialect use, although this was recorded, the number of participants from each dialect group was insufficient to allow for meaningful comparisons. Future studies would benefit from a more systematic consideration of such variables, particularly given their potential influence on both perception and production.

Multiple directions for future research emerge from the present study. First, it would be valuable to replicate this design with a larger and more geographically diverse participant pool. This is particularly

important among Italian learners, given the regional concentration of the current sample in Sardinia, to examine the extent to which dialectal variation influences L2 phonological performance.

Additionally, intervention-based studies are needed to explore whether targeted instruction, especially perception-led or articulatory-focused approaches, results in measurable improvements in learners' ability to perceive and produce challenging contrasts. Pre-/post-test designs could also provide insight into the sequence of development across modalities. Longitudinal studies are also essential for understanding how L2 phonological competence develops over time. Few studies have explored how both perception and production evolve together or separately across extended periods of learning, which could inform curriculum design and classroom practice (Sakai & Moorman, 2018; Thomson, 2011; Thomson & Derwing, 2015).

Concluding, researchers should examine the role of individual differences, such as age, working memory, and motivation. Even within the same L1 group, learners follow diverse acquisition trajectories; understanding these pathways could support the design of more personalised pronunciation instruction.

References

Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Adedeji, K. A. (2022). The effect of accent familiarity on the intelligibility of native speaker speech. *NIU Journal of Social Sciences*, *8*(3), 195-200. https://doi.org/10.58709/niujss.v8i3.1686

Anderson, R. T. & Centeno, J. G. (2018). Contrastive analysis between Spanish and English. In *Communication Disorders in Spanish Speakers* (pp. 11–33). Multilingual Matters. https://doi.org/10.21832/9781853599736-004

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. Behavior Research Methods, 52(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arjmandi, M. K. & Behroozmand, R. (2024). On the interplay between speech perception and production: insights from research and theories. *Frontiers in Neuroscience*, *18*, Article 1347614. https://doi.org/10.3389/fnins.2024.1347614

Asher, J. J. & García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal (Boulder, Colo.)*, *53*(5), 334–341. https://doi.org/10.1111/j.1540-4781.1969.tb04603.x

Atas, U. (2018). The role of receptive vocabulary knowledge in advanced EFL listening comprehension. *TESL-EJ, 21*(4). https://tesl-ej.org/wordpress/issues/volume21/ej84/ej84a4/

Balari, S. & Lorenzo, G. (2015). Should it stay or should it go? A critical reflection on the critical period for language. *Biolinguistics*, *9*, 8–42. https://doi.org/10.5964/bioling.9027

Bardovi-Harlig, K. & Sprouse, R. A. (2024). Negative versus positive transfer. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0084.pub2

Barriuso, T. A. & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, *30*(1). https://doi.org/10.5070/B5.35970

Bassetti, B., Cerni, T., & Masterson, J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics*, *43*(3), 683–705. https://doi.org/10.1017/S014271642200008X

Bassetti, B., Mairano, P., Masterson, J., & Cerni, T. (2020). Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning*, *70*(4), 1218–1256. https://doi.org/10.1111/lang.12423

Bassetti, B., Masterson, J., Cerni, T., & Mairano, P. (2021). Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1583–1603. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1037/xhp0000949

Berruto, G., Ayres-Bennett, W., & Carruthers, J. (2018). The languages and dialects of Italy. In *Manual of Romance Sociolinguistics* (Vol. 18, pp. 494–525). Walter de Gruyter GmbH. https://doi.org/10.1515/9783110365955-019

Berti, S. (2015). *Contrastive analysis of the Italian and English consonant phoneme systems* (Unpublished master's thesis). Ca' Foscari University of Venice.

Bertinetto, P. M. & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151. https://doi.org/10.1017/S0025100305002148

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, *167*(224), 233-277. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.7551/mitpress/2387.003.0011

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M. J. & Bohn, O.-S. (Eds.), *Language Experience in Second Language Speech Learning* (Vol. 17, pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552. https://doi.org/10.7334/psicothema2016.383

Bohn, O. S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303–328. https://doi.org/10.1017/S0142716400008912

Calvo Benzies, Y. J. (2017). Advanced Spanish university students problems with the pronunciation of English vowels. Identification, analysis and teaching implications. *Odisea (Almería, Spain)*, *14*, 37–56. https://doi.org/10.25115/odisea.v0i14.256

Campos-Astorkiza, R. (2013). The phonemes of Spanish. In Hualde, J. I., Olarrea, A., & O'Rourke, E. (Eds.), *The Handbook of Hispanic Linguistics* (1. Aufl., Vol. 93). (pp. 89 – 110). Wiley-Blackwell. https://doi.org/10.1002/9781118228098

Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, *145*(1), EL52–EL58. https://doi.org/10.1121/1.5087645

Cebrian, J., Gorba, C., & Gavaldà, N. (2021). When the easy becomes difficult: Factors affecting the acquisition of the English/iː/-/ɪ/contrast. *Frontiers in Communication*, *6*, Article 660917. https://doi.org/10.3389/fcomm.2021.660917

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (2023). The elusive impact of L2 immersion on translation priming. *Studies in Second Language Acquisition*, *45*(2), Article 0272263122000249. https://doi.org/10.1017/S0272263122000249

Choe, S., Lee, K., & So, Y. (2020). The effects of phonemic awareness instructions on L2 listening comprehension: A meta-analysis. *The Journal of Asia TEFL*, *17*(4), 1294–1309. https://doi.org/10.18823/asiatefl.2020.17.4.9.1294

Cilibrasi, L. & Marková, D. (2022). The development of vocabulary and overall proficiency in English as an additional (third) language. What matters more, daily use or age of onset? *Theory and Practice of Second Language Acquisition*, *8*(2), 1–22. https://doi.org/10.31261/TAPSLA.11862

Coe, N. (2001). Speakers of Spanish and Catalan. In Swan, M. & Smith, B. (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems. 2nd ed.* (pp. 90-112). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667121.008

Cohen, L., Manion, L., & Morrison, K. (2018). Experiments. In *Research Methods in Education* (8th ed., Vol. 1, pp. 391–426). Routledge. https://doi.org/10.4324/9781315456539-20

Colenbrander, D., Miles, K. P., & Ricketts, J. (2019). To see or not to see: How does seeing spellings support vocabulary learning?. *Language, Speech, and Hearing Services in Schools*, *50*(4), 609-628. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0135

Conklin, J., Dmitrieva, O., & Pineda Mora, G. M. (2019). Production and perception of two English vowels by advanced Colombian Spanish learners of English. *The Journal of the Acoustical Society of America*, *145*(3), 1827–1827. https://doi.org/10.1121/1.5101677

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching, IRAL*, *5*(1–4), 161–170. https://doi.org/10.1515/iral.1967.5.1-4.161

Corder, S. P. (1982). *Error analysis and interlanguage* (Vol. 198, No. 1). London: Oxford University Press.

De Leeuw, E., Stockall, L., Lazaridou-Chatzigoga, D., & Gorba Masip, C. (2021). Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research*, *37*(4), 587–618. https://doi.org/10.1177/0267658319886623

De Mauro, T. (2011). Per la storia linguistica dell'Italia repubblicana. *Italica*, *88*(1), 40–58. http://www.jstor.org/stable/23070857

Declerck, M. & Kirk, N. W. (2023). No evidence for a mixing benefit—A registered report of voluntary dialect switching. Plos one, 18(5), e0282086. https://doi.org/10.1371/journal.pone.0282086

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.42

Duguid, A. (2001). Italian Speakers. In Swan, M. & Smith, B. (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems. 2nd ed.* (pp. 73-89). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667121.007

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 151-190). Berlin, New York: De Gruyter Mouton. https://doiorg.ezproxyprd.bodleian.ox.ac.uk/10.1515/9783110219234.151

Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Escudero, P., Sisinni, B., & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America*, *135*(3), 1577–1584. https://doi.org/10.1121/1.4864477

EF Education First. (2024). *EF English Proficiency Index 2024*. https://www.ef.com/epi/

European Commission: European Education and Culture Executive Agency, Birch, P., Baïdak, N., De Coster, I., & Kocanova, D. (2023). *Key data on teaching languages at school in Europe: 2023 edition*, (P..Birch, Ed.) Publications Office of the European Union. https://data.europa.eu/doi/10.2797/529032

Fernández, F. M. & Roth, J. O. (2015) Capítulo 2: Cartografía del Español en el mundo. CVC. In *El español, lengua global. La economía.* Instituto Cervantes. https://cvc.cervantes.es/lengua/espanol_economia/02.htm

Field, J. (2019). Second language listening: Current ideas, current issues. In Schwieter, J. & Benati, A. (Eds.), *The Cambridge Handbook of Language Learning* (pp. 283–319). Cambridge University Press. https://doi.org/10.1017/9781108333603.013

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology. A Human Experimental Psychology*, *43*(3), 701–731. https://doi.org/10.1080/14640749108400993

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*(1), 233-277. https://www.researchgate.net/publication/333815781_Second_language_speech_learning_Theory_findings_and_problems

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In Schiller, N. O. & Meyer, A. S (Eds.), *Phonetics and Phonology in Language*

Comprehension and Production (Vol. 6, pp. 319–358). De Gruyter, Inc. https://doi.org/10.1515/9783110895094.319

Flege, J. E. (2019). A non-critical period for second-language learning. *A sound approach to language matters: In honor of Ocke-Schwen Bohn*, 501-541. Department of English, Aarhus University. https://www.researchgate.net/publication/328661276_A_non-critical_period_for_second-language_learning

Flege, J. E. & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*(1), 1–34. https://doi.org/10.1017/S0272263104261010

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987. https://doi.org/10.1121/1.428116

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Gallardo del Puerto, F. & Gómez Lacabex, E. (2008). La enseñanza de las vocales inglesas a los hablantes de español. *Pulso (Alcalá de Henares. Internet)*, *31*, 37–66. https://doi.org/10.58265/pulso.4963

Gass, S. M. & Mackey, A. (2005). Second language research: Methodology and design. In *Second Language Research*. Taylor & Francis Group.

Gómez González, M. A. & Sánchez Roura, T. (2021). *English pronunciation for speakers of Spanish: From theory to practice* (1st ed.). Universidad de Santiago de Compostela. https://doi.org/10.1515/9781501510977

Gordon, J. & Arias, R. S. (2024). "The most important thing is to make them aware": A case study of teacher metalanguage knowledge and explicit L2 pronunciation instruction. *TESOL Quarterly*, *58*(4), 1666–1702. https://doi.org/10.1002/tesq.3301

Gurzynski-Weiss, L. (2022). Individual difference factors for second language speaking. In *The Routledge Handbook of Second Language Acquisition and Individual Differences* (1st ed., Vol. 1, pp. 347–363). Routledge. https://doi.org/10.4324/9781003270546-28

Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, *11*(1), 20. https://doi.org/10.1186/s40468-021-00131-8

Hamada, Y. (2016). Shadowing: Who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research : LTR*, *20*(1), 35–52. https://doi.org/10.1177/1362168815597504

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hattori, K. & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception and production by adult Japanese speakers. In *Interspeech workshop on second language studies: Acquisition, learning, education and technology* (pp. 22-24). Tokyo: Waseda University.

Instituto Cervantes. (2023, July 17). *El español: una lengua viva. Informe 2023.* https://cvc.cervantes.es/lengua/anuario/anuario_23/informes_ic/p01.htm

Iverson, P. & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://doi.org/10.1121/1.2783198

Iverson, P. & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. https://doi.org/10.1121/1.3148196

James, C. (2013). *Errors in language learning and use: Exploring error analysis*. Routledge. https://doi.org/10.4324/9781315842912

Jarosz, A. (2023). Key aspects of pronunciation learning and teaching. In *English Pronunciation Teaching* (Vol. 160, pp. 8–18). Multilingual Matters. https://doi.org/10.21832/9781800410503-006

Kartushina, N. & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246–1246. https://doi.org/10.3389/fpsyg.2014.01246

Khaghaninejad, M. S. & Maleki, A. (2015). The effect of explicit pronunciation instruction on listening comprehension: Evidence from Iranian English learners. *Theory and Practice in Language Studies*, *5*(6), 1249. https://doi.org/10.17507/tpls.0506.18

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics*, *38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Krämer, M. (2009). *The phonology of Italian* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780199290796.001.0001

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenneberg, E. (1967). *The biological foundations of language.* New York: Wiley.

Lennon, P. (2008). Contrastive analysis, error analysis, interlanguage. *Bielefeld Introduction to Applied Linguistics. A Course Book. Bielefeld: Aisthesis Verlag*, 51-60.

Lipski, J. M. (2013). Geographical and social varieties of Spanish: An overview. In Hualde, J. I., Olarrea, A., & O'Rourke, E. (Eds.), *The Handbook of Hispanic Linguistics* (pp. 1-26). Wiley-Blackwell. https://doi.org/10.1002/9781118228098

Mackey, A. & Gass, S. M. (2015). *Second language research : methodology and design* (2nd ed.). Routledge, an imprint of Taylor and Francis. https://doi.org/10.4324/9781315750606

Manirakiza, E., Mugirase, G., & Hakizimana, I. (2021). Reflecting on the nature and causes of errors in second language learning and their classroom implications. *European Journal of Teaching and Education*, *3*(4), 18-28. https://doi.org/10.33422/ejte.v3i4.549

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2021). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, *38*(3), 581-605. https://doi.org/10.1177/0267658320988061

Meyerstein, R. S. (2019). *Functional load: Descriptive limitations, alternatives of assessment and extensions of application* (Reprint ed.). De Gruyter Mouton. https://doi.org/10.1515/9783111354163

Miralpeix, I. & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching, IRAL*, *56*(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Morett, L. M. (2019). The power of an image: Images, not glosses, enhance learning of concrete L2 words in beginning learners. *Journal of Psycholinguistic Research*, *48*(3), 643-664. https://doi.org/10.1007/s10936-018-9623-2

Moyer, A. (2007). Empirical Considerations on the Age Factor in L2 Phonology. *Issues in Applied Linguistics*, 15(2). http://dx.doi.org/10.5070/L4152005087

Munro, M. J. (2018). How well can we predict second language learners' pronunciation difficulties? *The CATESOL Journal*, *30*(1), 267-281. https://doi.org/10.5070/B5.35975

Munro, M. (2008). Foreign accent and speech intelligibility. In Edwards J. G. H. & Zampini M. L. (Eds.), *Phonology and Second Language Acquisition* (pp. 193-218). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.36.10mun

Munro, M. J., Flege, J. E., & Mackay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313–334. https://doi.org/10.1017/S0142716400007967

Nakata, T., Tamura, Y., & Scott, A. (2020). Examining the Validity of the LexTALE Test for Japanese College Students. *Journal of Asia TEFL, 17*(2), 335-348. https://doi.org/10.18823/asiatefl.2020.17.2.2.335

Núñez-Méndez, E. (2022). Variation in Spanish /s/: Overview and new perspectives. *Languages (Basel)*, *7*(2), 77. https://doi.org/10.3390/languages7020077

Nzerem, J-K. K. & Bob, P. O. (2021). The Role and significance of error linguistics in a second language environment. *Ansu Journal of Language and Literary Studies*, *1*(5).

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283. https://doi.org/10.1007/BF01067377

Peperkamp, S. & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Interspeech* (pp. 161-164).

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martin-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, *13*(3), 299–314. https://doi.org/10.1075/lab.22048.pui

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Roca, I. (2016). Phonology and English spelling. In *The Routledge Handbook of the English Writing System* (1st ed., pp. 65–91). Routledge. https://doi.org/10.4324/9781315670003-6

Rogerson-Revell, P. (2018). English vowels and consonants. In *The Routledge Handbook of Contemporary English Pronunciation* (1st ed., Vol. 1, pp. 92–121). Routledge. https://doi.org/10.4324/9781315145006-7

Rost Bagudanch, A. (2017). Variation and phonological change: The case of yeísmo in Spanish. *Folia Linguistica*, *51*(1), 169–206. https://doi.org/10.1515/flin-2017-0005

Rustipa, K. (2011). *Contrastive analysis, error analysis, interlanguage and the implication to language teaching*. *11*(1).

Sakai, M. (2016). *(Dis)connecting perception and production: Training native speakers of Spanish on the English /i/-/I/ distinction*. ProQuest Dissertations & Theses. https://www.proquest.com/dissertations-theses/dis-connecting-perception-production-training/docview/1853102967/se-2

Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224. https://doi.org/10.1017/S0142716417000418

Saito, K. & Plonsky, L. (2019). Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6* (4), 147-151. https://doi.org/10.1027/1614-2241/a000016

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223. https://doi.org/10.1017/S0267190500200135

Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2012). Pronominal objects in English–Italian and Spanish–Italian bilingual children. *Applied Psycholinguistics*, *33*(4), 725–751. https://doi.org/10.1017/S0142716411000543

Selinker, L. (2015). Interlanguage. In *Error analysis* (pp. 31-54). Routledge.

Silva Valencia, J. C. (2022). A comparative linguistic analysis of English and Spanish phonological system. *GIST – Education and Learning Research Journal*, *25*, 140–156. https://doi.org/10.26817/16925777.1152

Singleton, D. & Lengyel, Z. (1995). *The Age factor in second language acquisition*. Multilingual Matters. https://research.ebsco.com/linkprocessor/plink?id=e410f4bc-9b56-3124-a2a1-d2c5ea120f3c

Smith, L. E. & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), 333-342. https://doi.org/10.1111/j.1467-971X.1985.tb00423.x

Stan, I. (2024). Lending ears to EFL learners: perceived difficulties and leading causes. *Studi di glottodidattica*, *9*(1), 49-62. https://doi.org/10.15162/1970-1861/2028

Thomson, R. I. (2011). Computer Assisted Pronunciation Training. *CALICO Journal*, *28*(3), 744–765. https://doi.org/10.11139/cj.28.3.744-765

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT) A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208-231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R. I. (2022). The relationship between L2 speech perception and production. In *The Routledge handbook of second language acquisition and speaking* (pp. 372-385). Routledge. https://doi.org/10.4324/9781003022497-32

Thomson, R. I. & Derwing, T. M. (2015). Is phonemic training using nonsense or real words more effective?. *Pronunciation in Second Language Learning and Teaching Proceedings*, *7*(1).

Trinh, Q. L., Nguyen, T. D. L., & Le, T. T. (2022). Using explicit instruction of the International Phonetic Alphabet system in English as a foreign language adult classes. *European Journal of Educational Research*, *11(2)*, 749–761. https://doi.org/10.12973/eu-jer.11.2.749

Uribe-Enciso, O. L., Fuentes Hernandez, S. S., Vargas Pita, K. L., & Rey Pabón, A. S. (2019). Problematic phonemes for Spanish-speakers' learners of English. *GIST – Education and Learning Research Journal*, *19*, 215–238. https://doi.org/10.26817/16925777.701

Uther, M., Giannakopoulou, A., & Iverson, P. (2012). Hyperarticulation of vowels enhances phonetic change responses in both native and non-native speakers of English: Evidence from an auditory event-related potential study. *Brain research, 1470*, 52-58. https://doi.org/10.1016/j.brainres.2012.06.041

Vafaee, P. & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, *42*(2), 383–410. https://doi.org/10.1017/S0272263119000676

Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency? *The Modern Language Journal (Boulder, Colo.)*, *90*(1), 6–18. https://doi.org/10.1111/j.1540-4781.2006.00381.x

Vargas Jiménez, G. (2015). Three communication difficulties of EFL students. *Revista De Lenguas Modernas*, (23). https://doi.org/10.15517/rlm.v0i23.22347

Veiga-Pérez, C. (2017). Spanish teenager's pronunciation of English as a second language. *Revista de Educación de la Universidad de Granada*, *24*, 273-292.

Vera Diettes, K. J. (2014). Fenómenos de reducción vocálica por hablantes colombianos de inglés como L2: Un estudio acústico. *Forma y Función*, *27*(1), 11–43. https://doi.org/10.15446/fyf.v27n1.46940

Vietti, A. (2019). Phonological variation and change in Italian. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.494

Vincent, N. (1988). Italian. In Vincent, N. (Ed.) *The Romance Languages* (pp. 279–313). Routledge. https://doi.org/10.4324/9780203426531-8

Wardhaugh, R. (1970). The Contrastive analysis hypothesis. *TESOL Quarterly*, *4*(2), 123–130. https://doi.org/10.2307/3586182

Wheelock, A. (2016). Phonological difficulties encountered by Italian learners of English: An error analysis. *Hawaii Pacific University TESOL Working Paper Series*, *14*, 41-61.

Winter, B. (2019). Statistics for linguists: An introduction using R (1st ed.). Routledge. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.4324/9781315165547

World Population Review. (2024, July 6). *Italian Speaking Countries 2024.* https://worldpopulationreview.com/country-rankings/italian-speaking-countries

World Population Review. (2024, July 17). *Spanish Speaking Countries 2024.* https://worldpopulationreview.com/country-rankings/spanish-speaking-countries

Zhong, W., Dou, W., & Chen, J. (2023). The effect of explicit pronunciation instruction on enlarging listening vocabulary size. *Theory and Practice in Language Studies*, *13*(3), 718–727. https://doi.org/10.17507/tpls.1303.21

Appendices

Appendix A: Dialects spoken by Italian participants

Dialect	Count
Sardinian Campidanese	38
Abruzzese	2
Sicilian	2
Sardinian	2
Spezzino	1
Beneventano	1
Sardinian Ogliastra	1
Tuscan	1
Venetian	2
Campobassano	1
Romagnolo	1
Lucano	1
Calabrese	1
Florentine	1
Trentinian	2
Comasco	1
Sardinian Logudorese	1

Neapolitan	1
Salentino	1
Sardinian Nuorese	1
Brianzolo	1

Appendix B: Participant regional and national origins

Italian Regions	Count	Spanish Autonomous Communities	Count	Other Spanish-Speaking Countries	Count
Sardinia	44	Castilla y León	9	Colombia	8
Veneto	3	Extremadura	2	Mexico	3
Tuscany	2	Andalucía	2	Argentina	1
Abruzzo	2	Galicia	1	Puerto Rico	1
Sicily	2	La Rioja	1	Chile	1
Campania	2			Ecuador	1
Lombardia	2				
Trentino	2				
Liguria	2				
Piemonte	1				
Puglia	1				
Basilicata	1				
Calabria	1				
Emilia-Romagna	1				
Molise	1				

Appendix C: Demographics and Language Background Questionnaire

In this questionnaire you will be asked to answer questions about your language learning background as well as to provide demographic information necessary for this project. Please answer the following questions:

truthfully
carefully
honestly

Section 1: Demographic Information

1. **How old are you?**

2. **Please, select your gender:**

   ☐ Male

   ☐ Female

   ☐ Other

3. **What's your nationality?**

4. **Where are you from originally?**

   ☐ Italy

   ☐ Spain

   ☐ Other Italian-speaking country

   ☐ Other Spanish-speaking country

5. **If you answered Italy, please specify the Italian region: _____**

6. **If you answered Spain, please specify the Autonomous Community:_____**

7. **If you answered other Italian- or Spanish-speaking country, please specify:_____**

8. **Where do you currently live? Please, specify the country and city. _____**

Section 2: Language Background

1. **What is your first language (L1)?:**

☐ Italian

☐ Spanish

2. **What regional variety or dialect do you speak**? If there is more than one in your region, please specify the one you use or are most familiar with (e.g., Sardinian Campidanese). **_____**

3. **How much do you speak your regional variety/dialect?**

☐ Never

☐ Occasionally

☐ Frequently

☐ Daily

4. **At what age did you begin learning English?: ____**

5. **How long have you studied English (in years)?: ____**

6. **Where did you learn English? (Select all that apply):**

☐ School

☐ Language courses

☐ Abroad

☐ Other – please specify: ____

7. **Rate your English proficiency (CEFR level):**

☐ Beginner (A1-A2)

☐ Intermediate (B1-B2)

☐ Advanced (C1-C2)

8. **Do you speak any other languages?**

☐ Yes

☐ No

If yes, write the languages you speak in the order you learned them. If you grew up speaking two languages at the same time, write both as your first language (e.g., 1. Italian and French, 2. English, 3. Portuguese). _____

9. How often do you use your additional language (the language you speak the most besides your L1(s) and English)?

☐ Never

☐ Occasionally

☐ Frequently

☐ Daily

Section 3: Language Exposure

1. How often do you use English in your daily life?:

☐ Never

☐ Occasionally

☐ Frequently

☐ Daily

2. Have you ever lived in an English-speaking country?:

☐ Yes (If yes, specify where and for how long): ____

☐ No

3. Rate your exposure to English-speaking media (TV, radio, music, etc.):

☐ Never

☐ Occasionally

☐ Frequently

☐ Daily

Section 4: Task Readiness

Do you have any hearing or speaking difficulties that could affect your participation in this study (e.g., trouble hearing sounds or pronouncing words)?

☐ Yes

☐ No

☐ (If yes, please explain): _____

Appendix D: Sentences Used in the Production Task

The 24 sentences used in the production task are listed below. They are divided into two categories, Experimental and Control, and organised by target vowel.

Vowel	Experimental	Control
/iː/	The sheep is grazing in the field. I found a bean under the table. They cooked a leek in the kitchen. The room was filled with warm heat.	
/ɪ/	The ship sails slowly across the ocean. There is a bin outside the bathroom. The boy licked the envelope to seal it. I accidentally hit my hand on the table.	
/ɑː/	Let’s go to the park after dinner. Please stay calm while we wait outside. The dog began to bark at the cat. She has a big and kind heart.	
/æ/	He carried a heavy pack to school. The cam broke down in the machine.	

	My back hurts from sitting all day. She wore a bright red hat today.	
/e/		The family bought a new pet rabbit. She placed a big bet on the match. He paid his debt on time yesterday. The bank sent her a cheque last week.
/ɒ/		The soup boiled over in the pot. A helpful bot answered all my questions. There is a small black dot on the map. I lost one blue sock this morning.

Appendix E: Perception Task: Vowel Contrasts, Stimuli, and Conditions

Contrast	Aural Stimulus	MP Word	MP Word	Repetition	Reverse Condition
/iː/ - /ɪ/	/ʃiːp/	sheep	ship	No	Yes – /ʃɪp/
/iː/ - /ɪ/	/liːk/	leek	lick	Yes	No
/iː/ - /ɪ/	/biːn/	bean	bin	No	Yes – /bɪn/
/iː/ - /ɪ/	/hiːt/	heat	hit	Yes	No
/ɑː/ - /æ/	/pɑːk/	park	pack	Yes	No
/ɑː/ - /æ/	/hɑːt/	heart	hat	No	Yes – /hæt/
/ɑː/ - /ʌ/	/dɑːk/	dark	duck	No	Yes – /dʌk/
/ɑː/ - /ʌ/	/kɑːm/	calm	come	Yes	No
/æ/ - /ʌ/	/kʌt/	cat	cut	Yes	No
/æ/ - /ʌ/	/kæp/	cap	cup	No	Yes – /kʌp/
/æ/ - /ʌ/	/kæm/	cam	come	Yes	No
/æ/ - /ʌ/	/bʌg/	bag	bug	No	Yes – /bæg/
/əʊ/ - /ɒ/	/gəʊt/	goat	got	No	Yes – /gɒt/
/əʊ/ - /ɒ/	/bəʊt/	boat	bot	Yes	No
/əʊ/ - /ɒ/	/kəʊt/	coat	cot	Yes	No
/əʊ/ - /ɒ/	/rəʊd/	road	rod	No	Yes – /rɒd/

Appendix F: Rating Instructions, Coding Template, and Participant Labels for the Production Task

The following materials were provided to the two trained raters who independently evaluated the vowel productions from the Read-Aloud task:

F1. Rating Instructions (as shared with raters)

Criteria for Data Analysis

Dear raters,

This document provides detailed instructions for analysing the data from the read-aloud task in the experiment. You have received this document along with an Excel template, which you will need to use for the analysis. Please follow the instructions below to ensure consistency in your evaluation.

✓ Each participant's analysis should take approximately 15 minutes.

✓ You will analyse the data for 24 participants in total, which is 20% of the total sample.

In this experiment, participants were asked to read sentences aloud. However, for this analysis, you are required to focus only on the target words within these sentences and not on the entire sentence. The goal is to assess the accuracy of vowel production in the specified target words.

Instructions & Criteria for Analysis

1. Listen carefully to the target word in the sentence.
2. Compare the vowel produced with the target vowel listed in the table.
3. Decide if the pronunciation is target-like (matching the expected vowel).

• If fully accurate, mark it as 'Correct' under the Accuracy column.

• If the vowel differs from the target, mark it as 'Incorrect' and provide a description of the error in the Error Analysis column.

• If the vowel falls between two categories, decide which one it leans toward and classify accordingly.

Target Vowel Characteristics

/iː/ = Long, close front unrounded vowel

/ɪ/ = Short, near-close near-front unrounded vowel

Key criterion: vowel length

/ɑː/ = Open back unrounded vowel

/æ/ = Near-open front unrounded vowel

Key criterion: tongue position (frontness/backness), openness

Spreadsheet Columns

Participant

Target Vowel

Target Word

Accuracy (Correct / Incorrect)

Error Analysis (describe deviation, use IPA if possible)

Comment (e.g., if deviant but acceptable)

F2. Coding Template Structure

The Excel spreadsheet included the following column headings:

Participant	Target Vowel	Target Word	Accuracy	Error Analysis	Comment

Each rater was asked to label participants according to the relabelling system below.

F3. Participant Relabels

Each participant was relabelled with an anonymised code for the purpose of the rating task.

Participants for Rater 1

(It = Italian, Sp = Spanish)

12339132 → It1

12366634 → It2

12366687 → It3

12367428 → It4

12459957 → It5

12460107 → Sp1

12460198 → Sp2

12460468 → It6

12460590 → Sp5

12460694 → Sp6

12460746 → Sp7

12461303 → Sp8

12461555 → It12

12461585 → It13

12461720 → It14

12461784 → It16

12461882 → Sp9

12462232 → Sp10

12462767 → Sp11

12463636 → It19

12463731 → It20

12463769 → It21

12463924 → It22

12464110 → Sp13

Participants for Rater 2

12464786 → It24

12464806 → It25

12465170 → Sp14

12467812 → It27

12468286 → It29

12468353 → It30

12468489 → It31

12468657 → It33

12471481 → It34

12472114 → It35

12474844 → It36

12475591 → It37

12475680 → It38

12476081 → It39

12476169 → It40

12476229 → It41

12476455 → Sp16

12478359 → It43

12485353 → It45

12488801 → It46

12491389 → It47

12494962 → It48

12495921 → It49

12497116 → It50

Appendix G: Observed Agreement Calculations for Interrater Reliability

The following table presents the observed agreement (Po) between the primary researcher and each of the two external raters who independently evaluated vowel accuracy in the Read-Aloud Task.

Observed agreement (Po) was calculated using the formula:

Po = (Number of agreements) / (Total number of items rated)

Each rater analysed the production data of 24 participants, corresponding to a total of 384 tokens per rater (16 items × 24 participants). The observed agreement values were used as the primary reliability metric.

Rater Pair	Number of Agreements	Total Items Rated	Observed Agreement (Po)
Researcher – Rater 1	314 (228 + 86)	384	0.82
Researcher – Rater 2	254 (223 + 31)	384	0.66 (= 0.7)

Note. The observed agreement with Rater 2 was 0.66 and is reported in the main text as 70% for simplicity. Disagreements were limited in scope and did not indicate systematic inconsistencies. Based on this, the researcher proceeded to code the remainder of the dataset independently using the same rubric.

Appendix H: Information Sheet and Consent Form

The following is the full information sheet and consent form as presented to participants at the beginning of the study. Participants were required to read the English version and could also consult brief summaries in Italian or Spanish. They provided informed consent before proceeding.

DEPARTMENT OF EDUCATION

Dr Faidra Faitaki
University Direct Line: 01865 274024
University e-mail: faidra.faitaki@education.ox.ac.uk

Information Sheet for Participants

The production and perception of L2 phonological vowel contrasts by L1 Italian and L1 Spanish learners of English.

CUREC Approval Reference: EDUC_663263

Dear participant,

Thank you for your interest in participating in this online research study.

The aim of this research is to explore how native speakers of Italian and Spanish who are learning or have learnt English as a second language perceive and produce specific phonological contrasts in English. This will help us understand the difficulties learners face with certain sounds and how this affects their intelligibility in English.

You have been invited to participate as you are a native speaker of either Italian or Spanish, aged 18 or above, and are currently learning or have learnt English as a second language. Participation

is completely voluntary, and you are under no obligation to take part in the research. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below. You may ask any questions before deciding to take part by contacting the researcher (details below).

The Principal Researcher is Claudia Concas who is attached to the Department of Education at the University of Oxford. This research is being completed under the supervision of Dr Faidra Faitaki.

What will I be asked to do?

As part of the study, you will be asked to complete an online questionnaire regarding your language learning background and a quick online proficiency test. Following this, you will perform two tasks:

1. **Production Task**: you will be presented with short sentences in English, which you will be asked to read aloud while recording yourself. Since this is an audio recording task, please ensure that you complete the experiment in a quiet room.
2. **Perception Task:** after hearing audio stimuli, you will be asked to identify the correct image that corresponds to the word you hear.

The experiment should take about 20 minutes to complete. No background knowledge is required.

Please note that there is a chance your screen may freeze mid-task or mid-experiment. In the rare event that this happens (e.g., the experiment screen freezes or a blank screen appears), please wait at least one minute before closing your browser and leaving the experiment. If the

issue persists, this is a technical fault on our side, for which we apologise profusely. You are welcome to try again if you wish to do so; otherwise, we greatly appreciate your participation.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. However, you will not be able to withdraw your consent once the experiment is complete.

How will my data be used?

We will not collect any data that could directly identify you, aside from an optional email address which you may provide upon completion of the experiment if you wish to enter a lottery draw for a chance to win a £25 voucher. Providing your email address is entirely optional. Your IP address will not be stored. We will take all reasonable measures to ensure that data remain confidential.

The responses you provide will be stored in a password-protected electronic file on University of Oxford secure servers and may be used in academic publications and reports for external sources. Identifiable information will be deleted as soon as it is no longer required for the research.

Who will have access to my data?

The University of Oxford is the data controller with respect to your personal data and, as such, will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform

in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

The results will be written up for an MSc degree.

Who has reviewed this research?

The application was reviewed and approved by my supervisor on behalf of the Department of Education's Research Ethics Committee [EDUC_663263].

Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this research, please contact Claudia Concas at claudia.concas@st-annes.ox.ac.uk or Dr Faidra Faitaki at faidra.faitaki@education.ox.ac.uk, and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the University of Oxford Research Governance, Ethics & Assurance (RGEA) team at rgea.complaints@admin.ox.ac.uk or on 01865 616480.

Scheda informativa (Italian speakers)

Questo studio esplora come i madrelingua italiani e spagnoli che stanno imparando o hanno imparato l'inglese percepiscono e producono determinati contrasti fonologici in inglese. Ti verrà chiesto di compilare un questionario online, completare un breve test di competenza linguistica e partecipare a due attività: leggere ad alta voce frasi brevi e identificare immagini basate su stimoli audio. La partecipazione è volontaria, anonima e puoi ritirarti in qualsiasi momento prima di inviare le tue risposte. Non saranno raccolti dati personali identificativi, tranne un

indirizzo email che è opzionale fornire se desideri partecipare a un sorteggio per vincere un buono da 25 sterline. I dati saranno conservati in modo sicuro e utilizzati solo a fini di ricerca. Per partecipare, è necessario compilare le caselle di controllo qui sotto, certificando di avere almeno 18 anni e accettando le informazioni sul trattamento e la protezione dei dati descritte nel modulo informativo. Compilando le caselle, darai il tuo consenso a partecipare allo studio. Per qualsiasi dubbio, puoi contattare la ricercatrice principale, Claudia Concas.

Hoja informativa (Spanish speakers)

Este estudio explora cómo los hablantes nativos de italiano y español que están aprendiendo o han aprendido inglés perciben y producen ciertos contrastes fonológicos en inglés. Se te pedirá que completes un cuestionario en línea, realices una breve prueba de competencia lingüística y participes en dos actividades: leer en voz alta oraciones cortas e identificar imágenes basadas en estímulos auditivos. La participación es voluntaria, anónima y puedes retirarte en cualquier momento antes de enviar tus respuestas. No se recopilarán datos personales identificativos, excepto una dirección de correo electrónico que es opcional proporcionar si deseas participar en un sorteo para ganar un vale de 25 libras. Los datos se almacenarán de manera segura y se utilizarán solo con fines de investigación. Para participar, es necesario completar las casillas de verificación a continuación, certificando que tienes al menos 18 años y aceptando la información sobre el tratamiento y la protección de datos descrita en este documento. Al hacerlo, estarás dando tu consentimiento para participar en el estudio. Para cualquier consulta, puedes contactar con la investigadora principal, Claudia Concas.

Please note that you may only participate in this online experiment if you are 18 years of age or over.

☐ I certify that I am 18 years of age or over

If you have read the information above and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please tick the box below to start.

☐ Yes, I agree to take part

Appendix I: Ethics Approval Letter



Education (Educ) DREC
15 Norham Gardens, Oxford, OX2 6PY

Applicant: Claudia Concas
Principal Investigator: Dr Faidra Faitaki
Department: Education

Study title: The production and perception of L2 phonological vowel contrasts by L1 Italian and L1 Spanish learners of English.
(Version: 1.1)

Ethics reference: Education (Educ) DREC - 663263

Dear Dr Faidra Faitaki,

On behalf of the Committee, I confirm that the above research study described in the application and other supporting documentation submitted to the committee has been carefully considered on behalf of the Education (Educ) DREC in accordance with the University's regulations and policy for ethics approval of research involving human participants, human tissue and/or personal data. The opinion is as follows:

Opinion of Research Ethics Committee: Amendment Submission

Subject to the following conditions:

Decision Date: 16 Dec 2024, 18:24

Opinion End Date:

If favourable, insurance-provided indemnity arrangements will be in place between the decision date and opinion end date and you may now commence your study activities. Should you plan to continue the research beyond the end date above, it is your responsibility to ensure that you request, and receive, an extension (via amendment) from the committee for indemnity to remain in place. You may be required to provide a justification.

Please note the following:

Amendments: Should there be any subsequent changes to the reviewed study, applications for amendments can be made via the Oxford Ethics Application System (Worktribe Ethics).

Reports: Studies considered by OxTREC are expected to submit an *annual progress report* on each anniversary of study approval, until the study is completed. An end of study report is also required.

Audit: This study may be selected for audit at the discretion of the Research Governance, Ethics and Assurance Team.

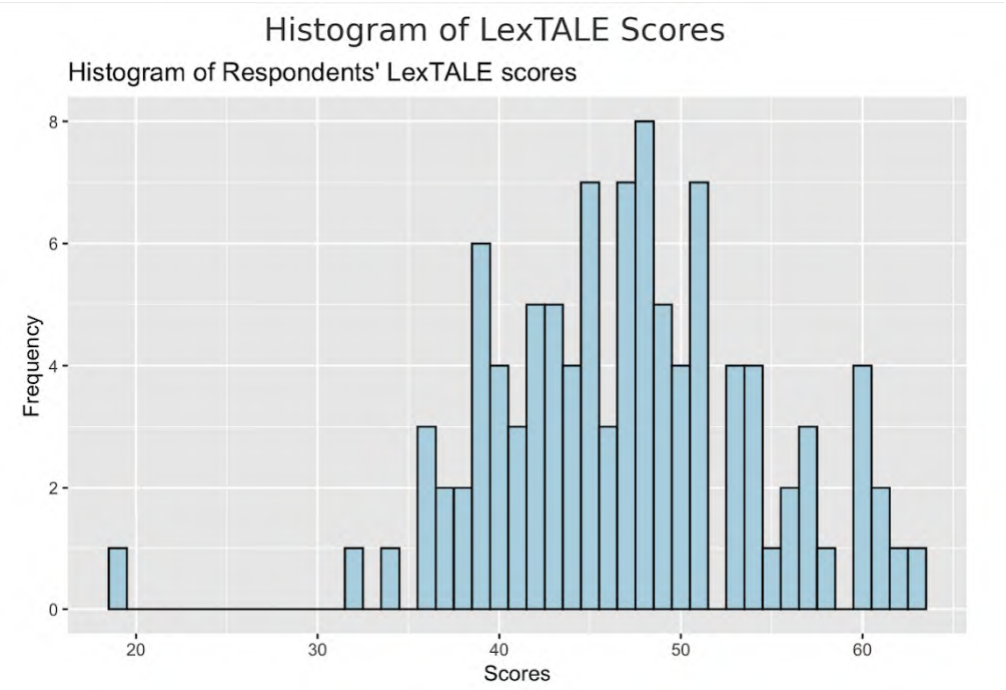
Data safety: It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely in accordance with University requirements. Further guidance and advice are available from the Research Data Team. Additional information is

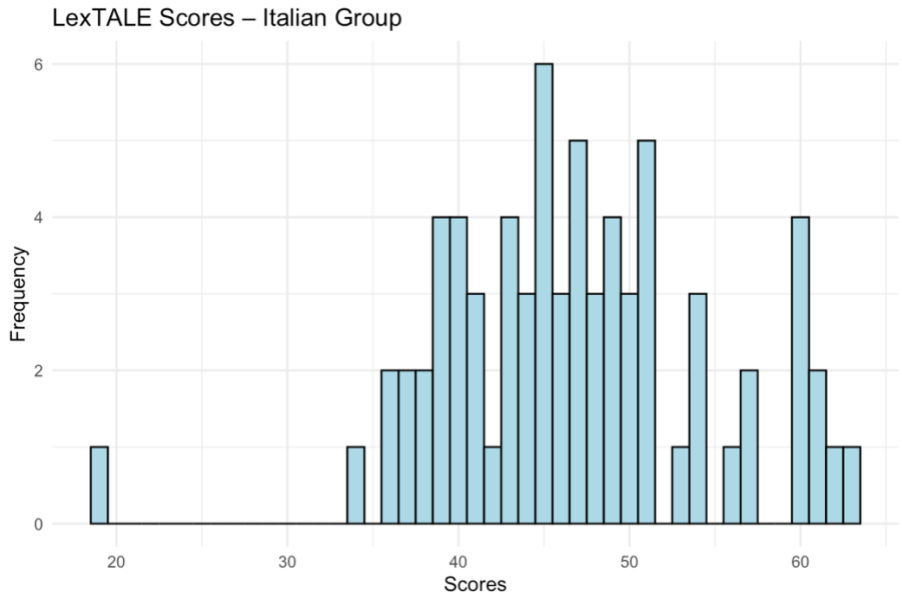
available at https://researchsupport.web.ox.ac.uk/governance/ethics

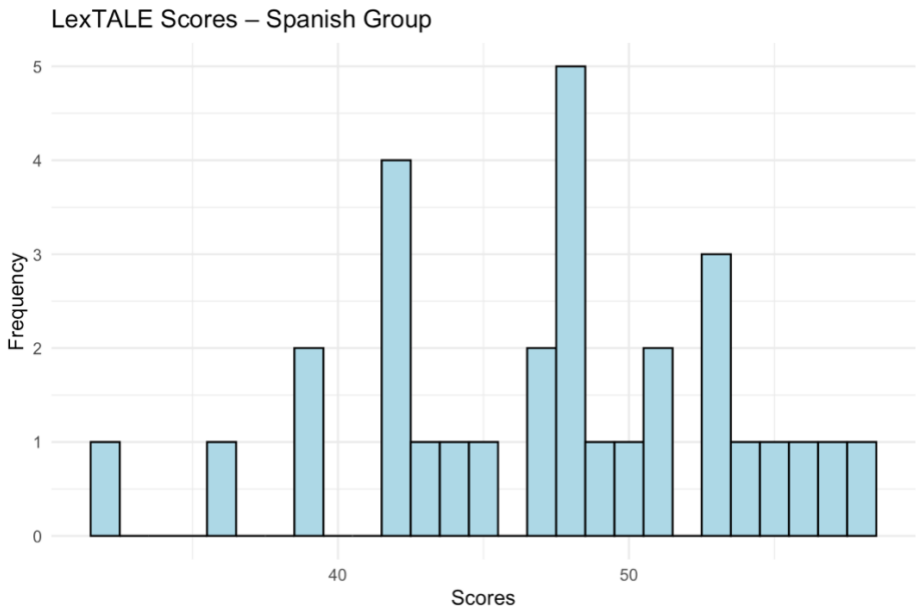
Yours Sincerely

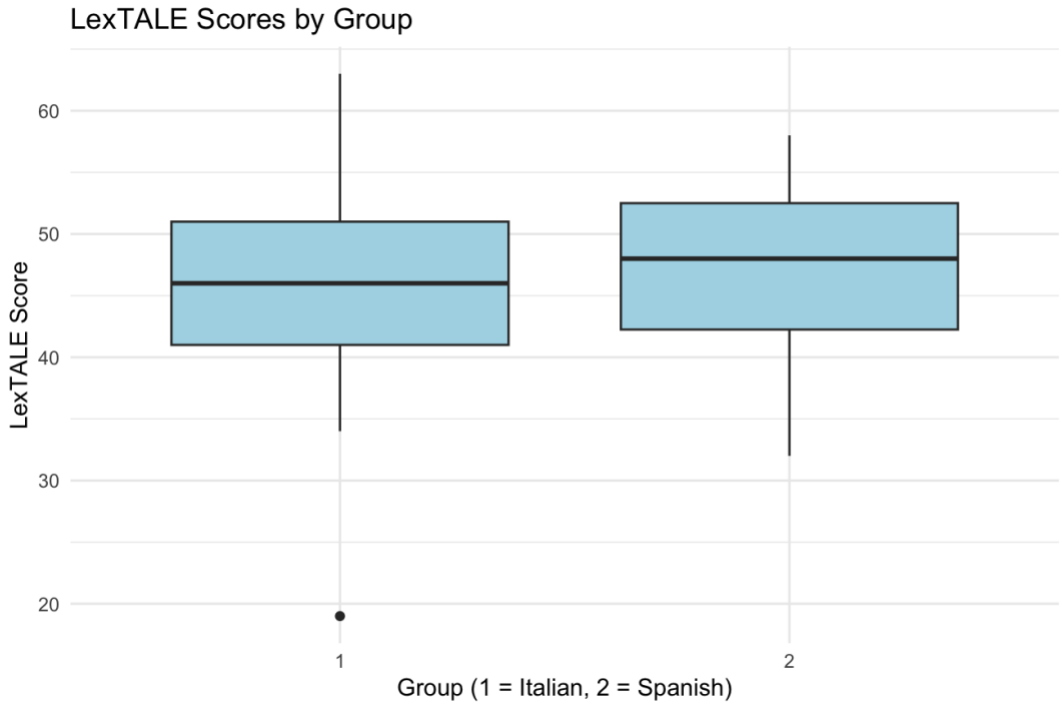
Education Ethics Officer

Appendix J: Distribution of LexTALE scores (whole sample and by group)

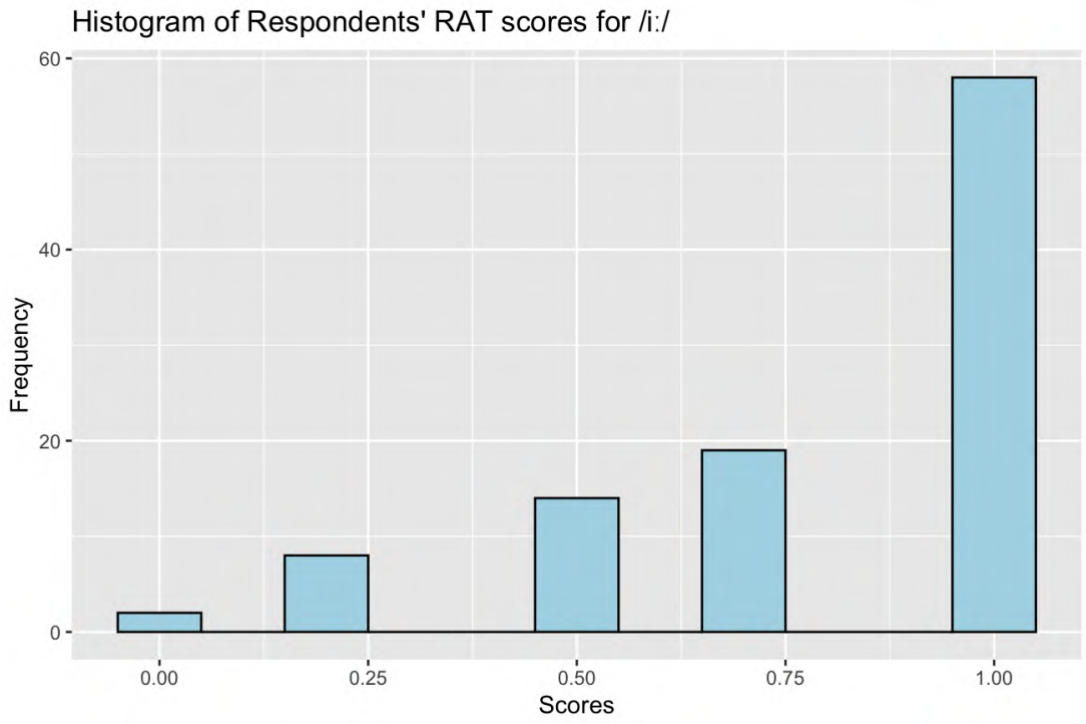


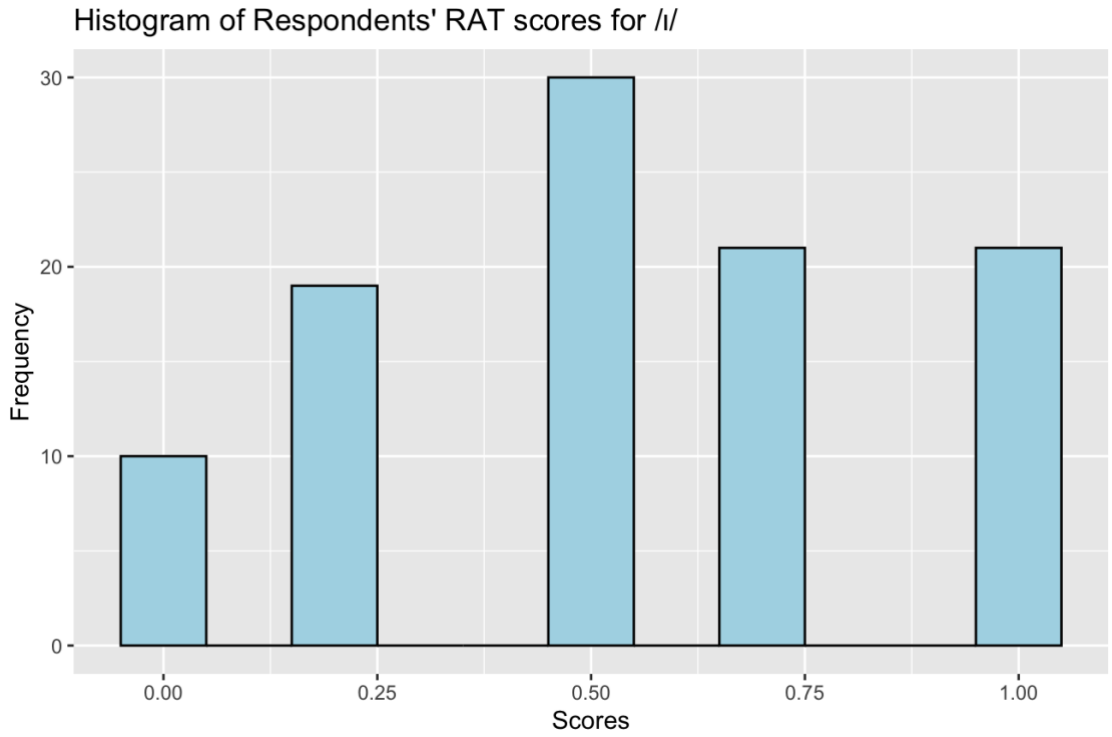


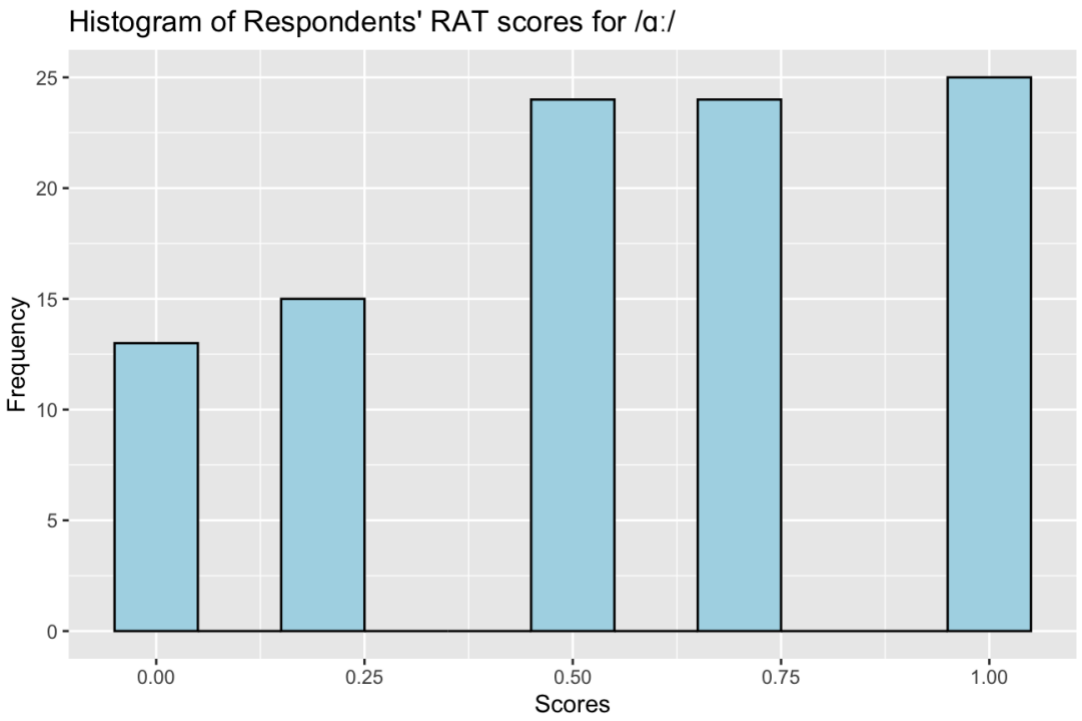

LexTALE Scores – Spanish Group
Frequency
0
1
2
3
4
5
40
50
Scores

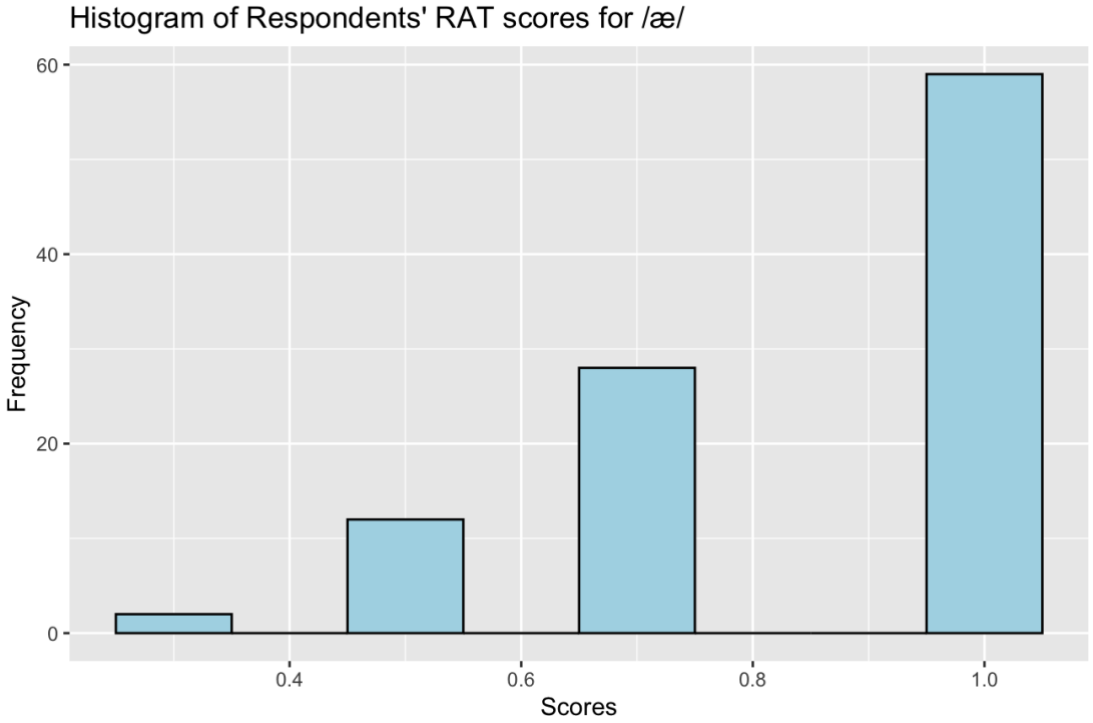

LexTALE Scores by Group
LexTALE Score
20
30
40
50
60
1
2
Group (1 = Italian, 2 = Spanish)

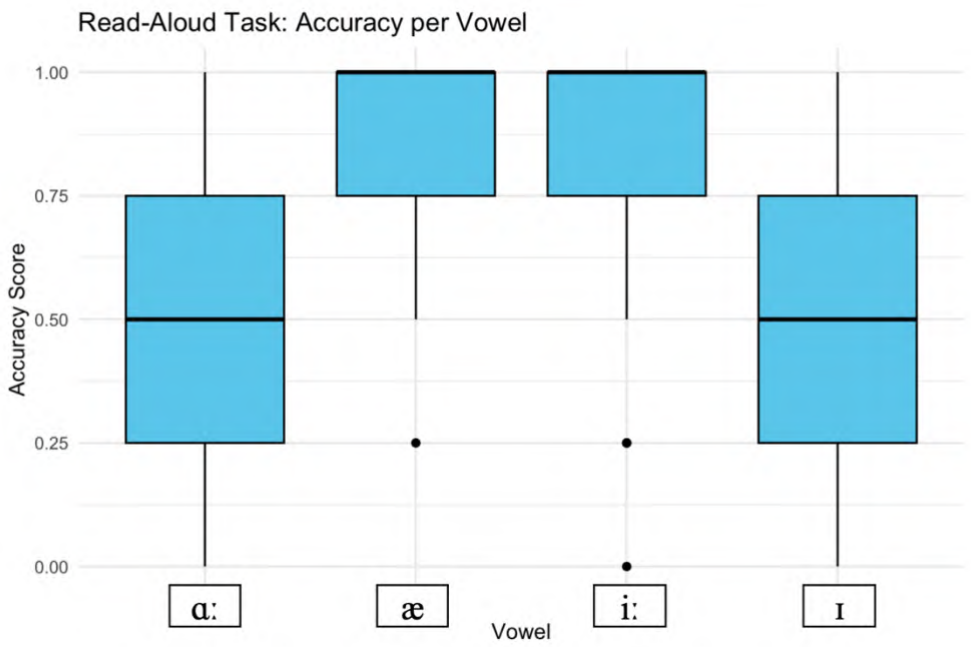
Appendix K: Distribution of Read-Aloud Task accuracy scores per vowel (whole sample)



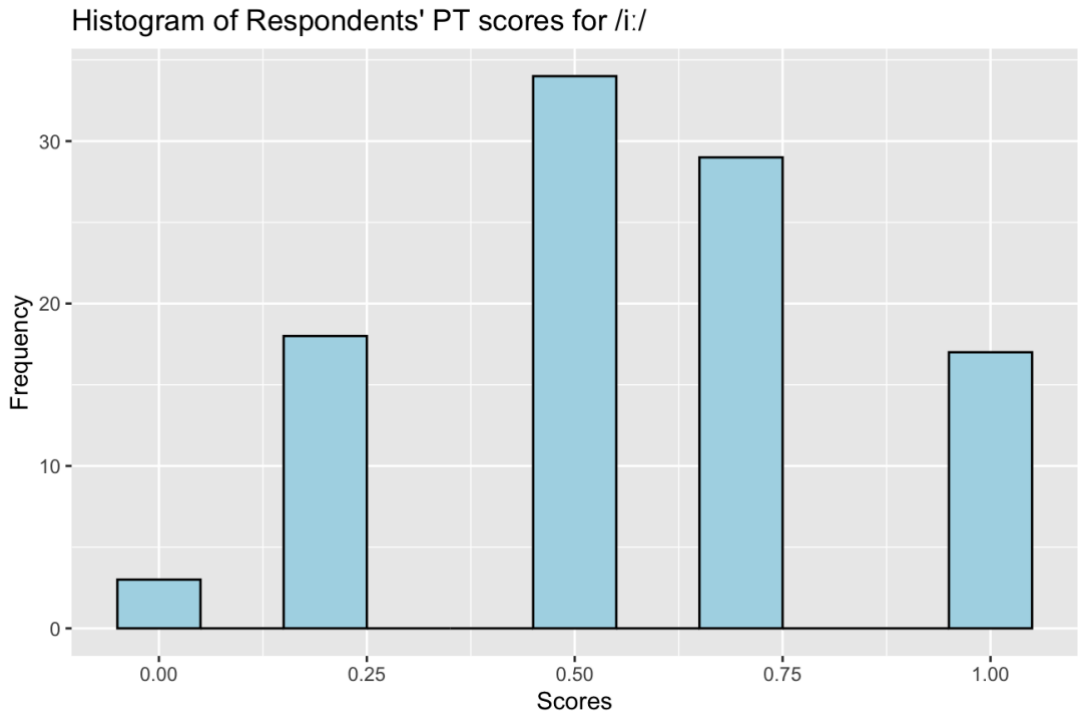


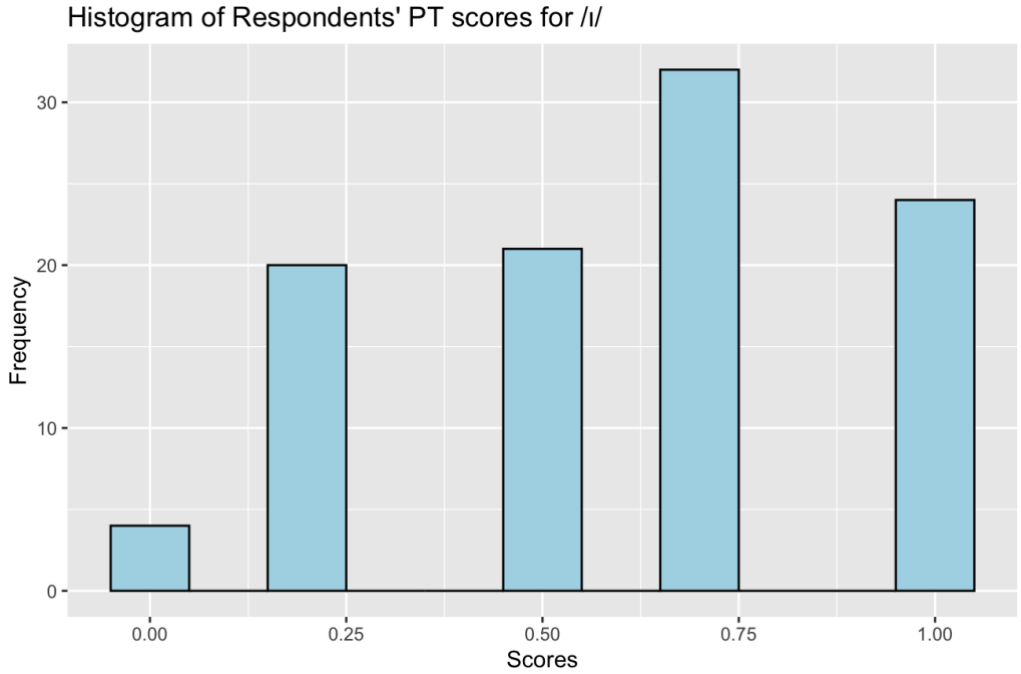

Histogram of Respondents' RAT scores for /ɑ:/
Frequency
25
20
15
10
5
0
0.00
0.25
0.50
0.75
1.00
Scores

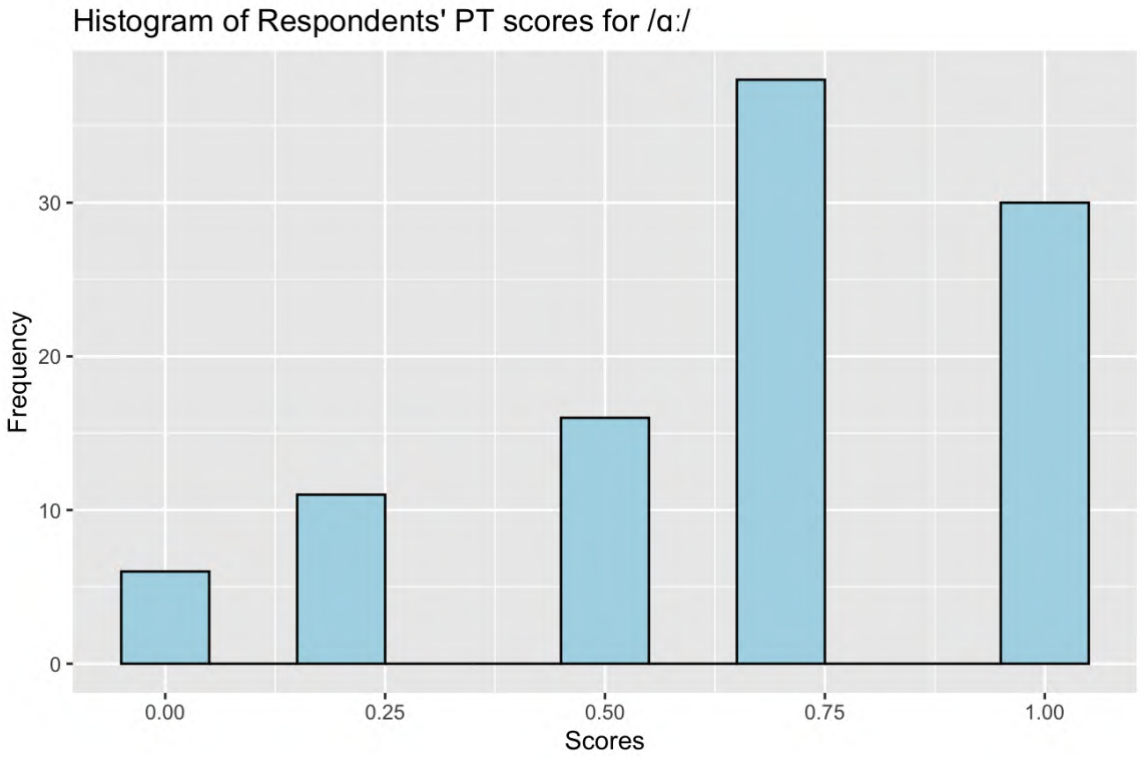

Histogram of Respondents' RAT scores for /æ/
Frequency
60
40
20
0
0.4
0.6
0.8
1.0
Scores

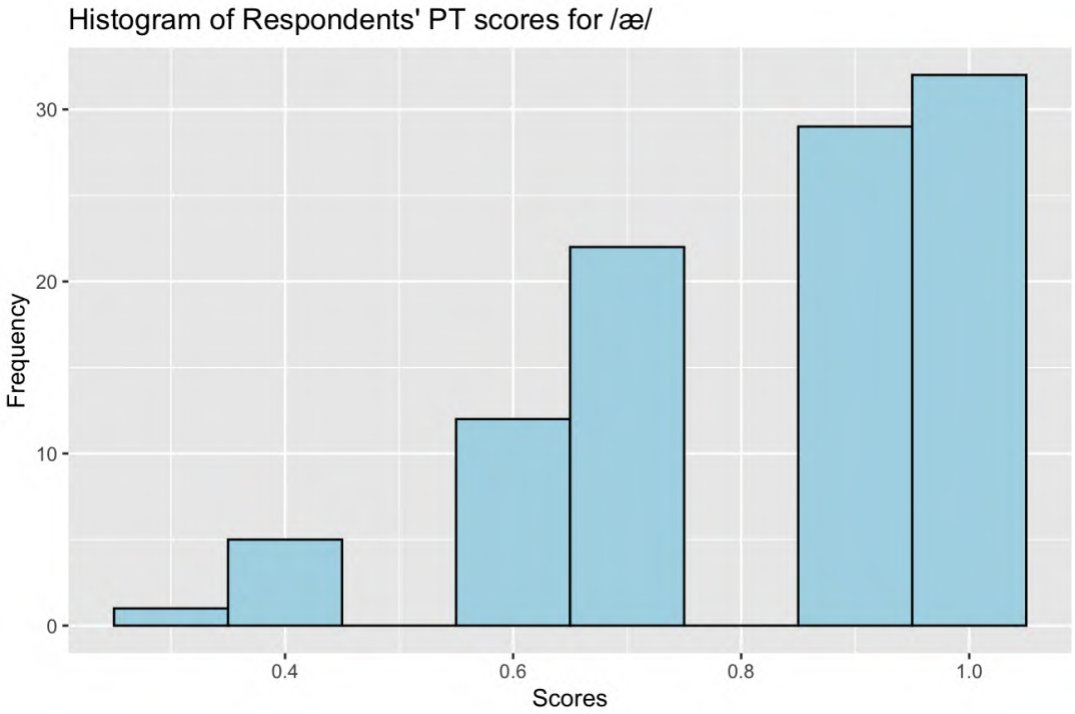

Read-Aloud Task: Accuracy per Vowel
Accuracy Score
1.00
0.75
0.50
0.25
0.00
ɑː
æ
iː
ɪ
Vowel

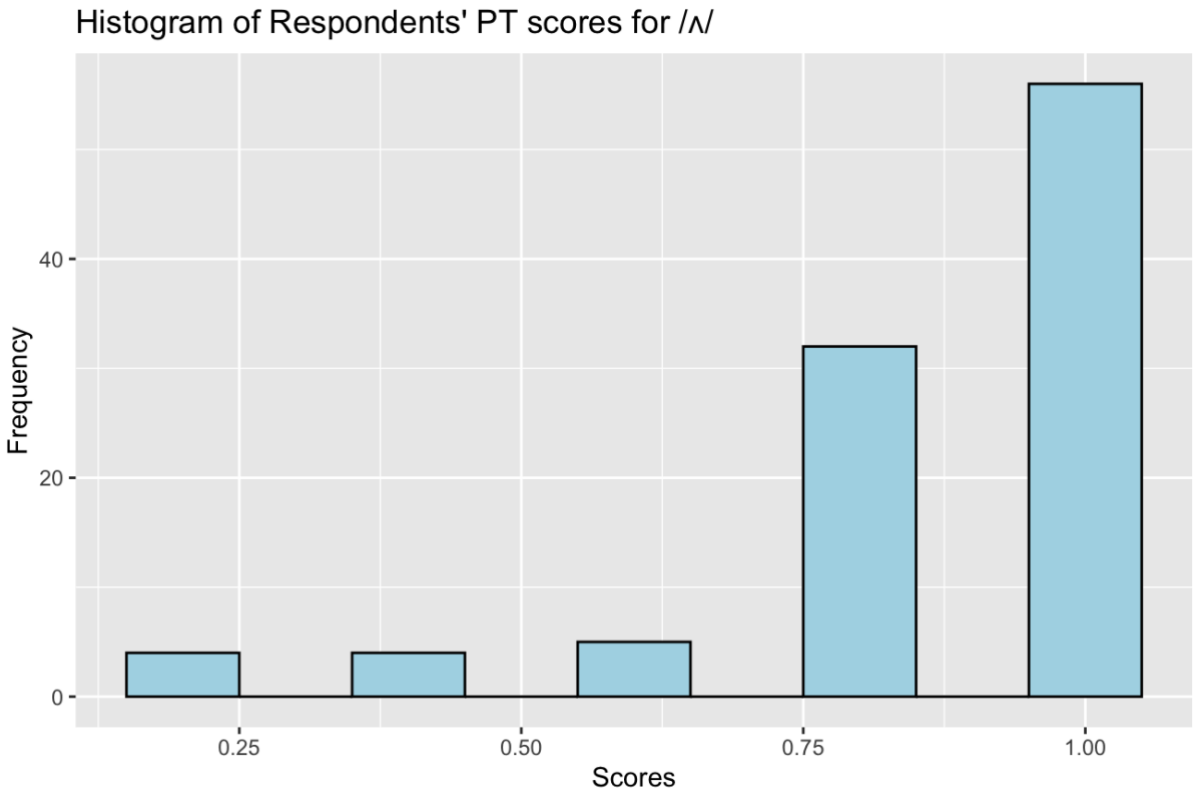
Appendix L: Distribution of Perception Task accuracy scores per vowel (whole sample)

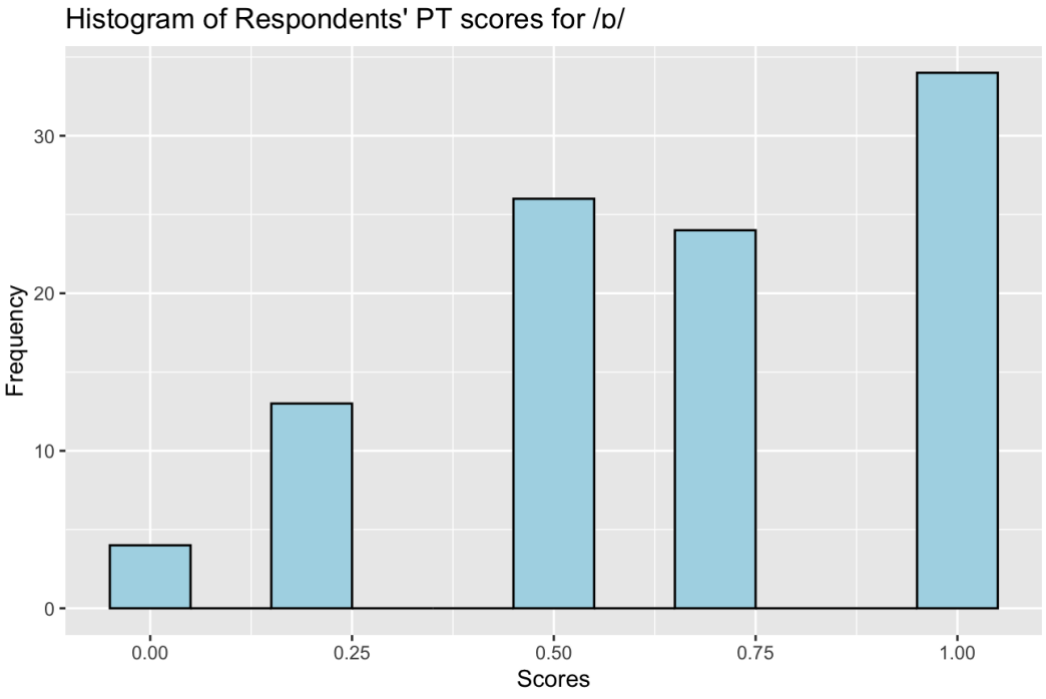


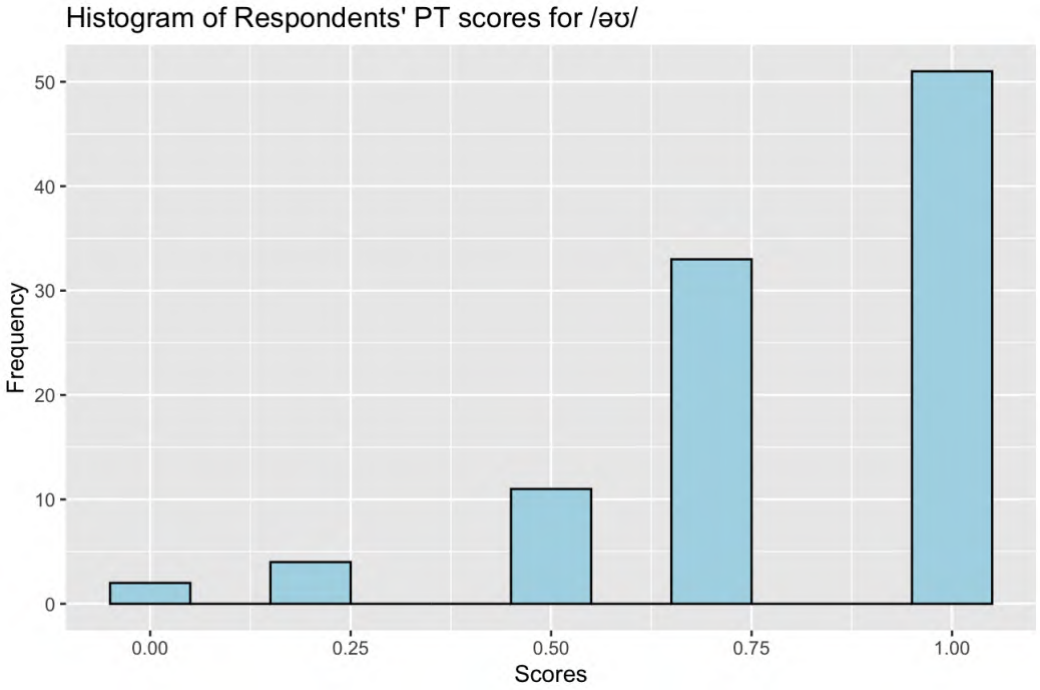


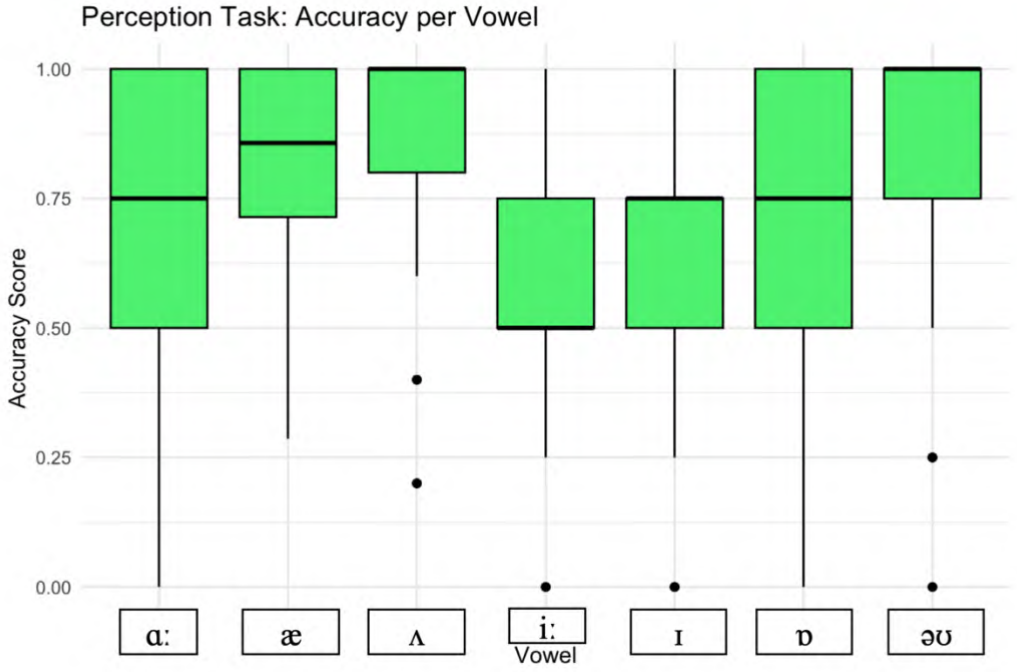

Histogram of Respondents' PT scores for /ɑː/
Frequency
30
20
10
0
0.00
0.25
0.50
0.75
1.00
Scores


Histogram of Respondents' PT scores for /æ/
Frequency
30
20
10
0
0.4
0.6
0.8
1.0
Scores

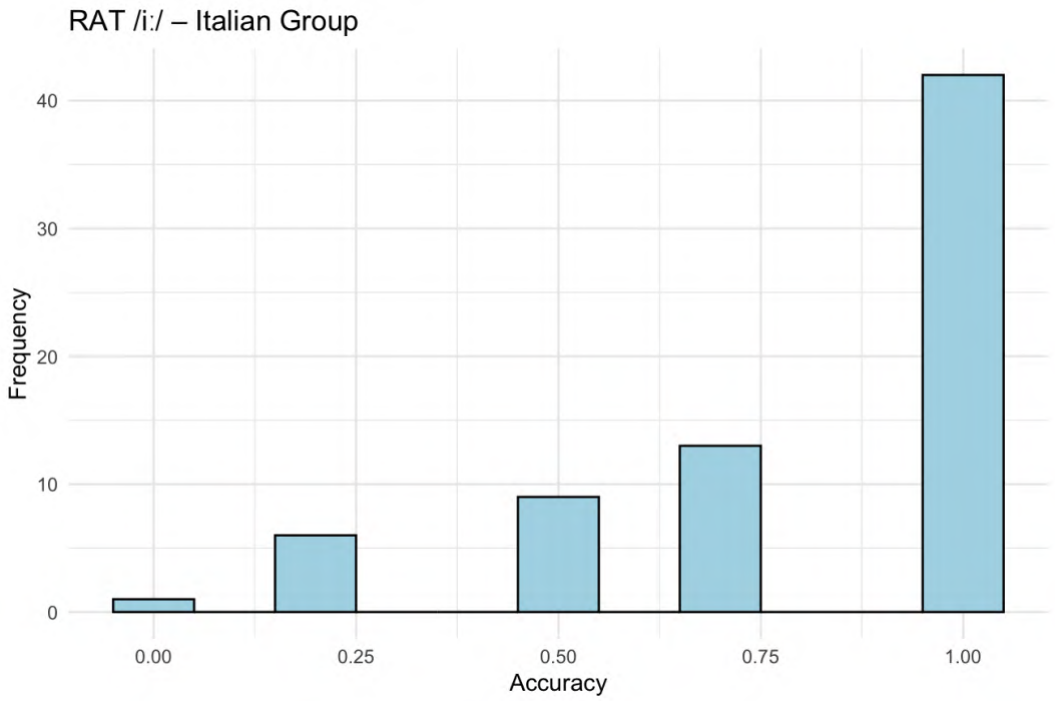

Histogram of Respondents' PT scores for /ʌ/
Frequency
40
20
0
0.25
0.50
0.75
1.00
Scores

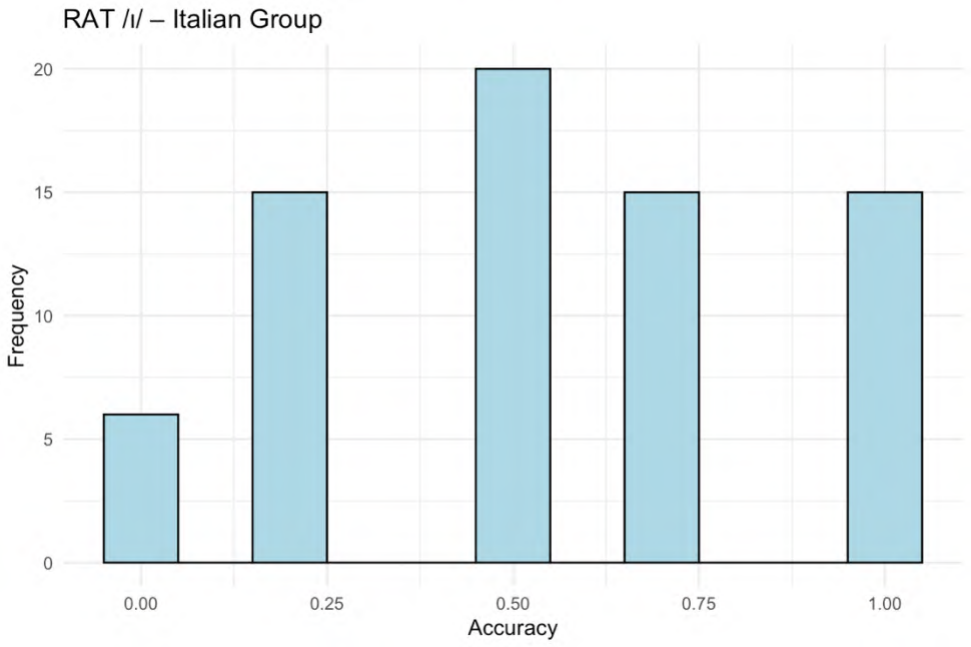

Histogram of Respondents' PT scores for /ɒ/
Frequency
30
20
10
0
0.00
0.25
0.50
0.75
1.00
Scores

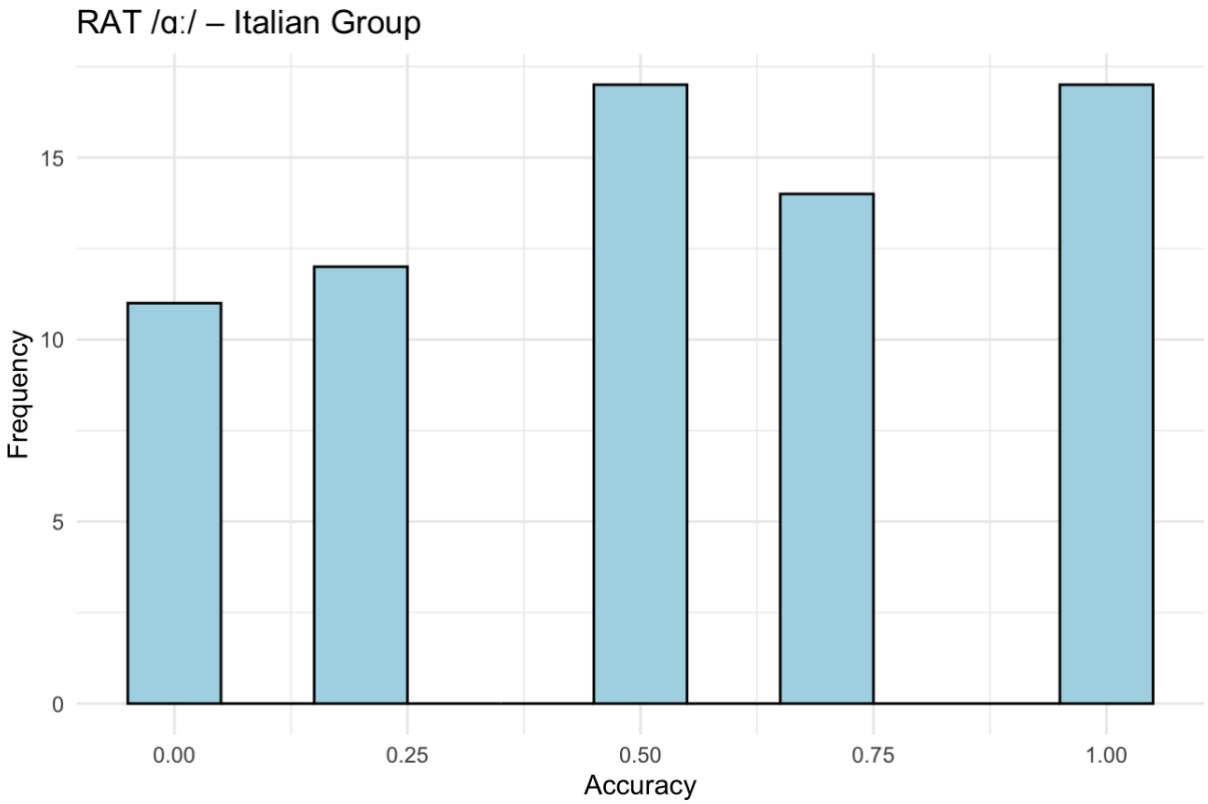

Histogram of Respondents' PT scores for /əʊ/
Frequency
50
40
30
20
10
0
0.00
0.25
0.50
0.75
1.00
Scores

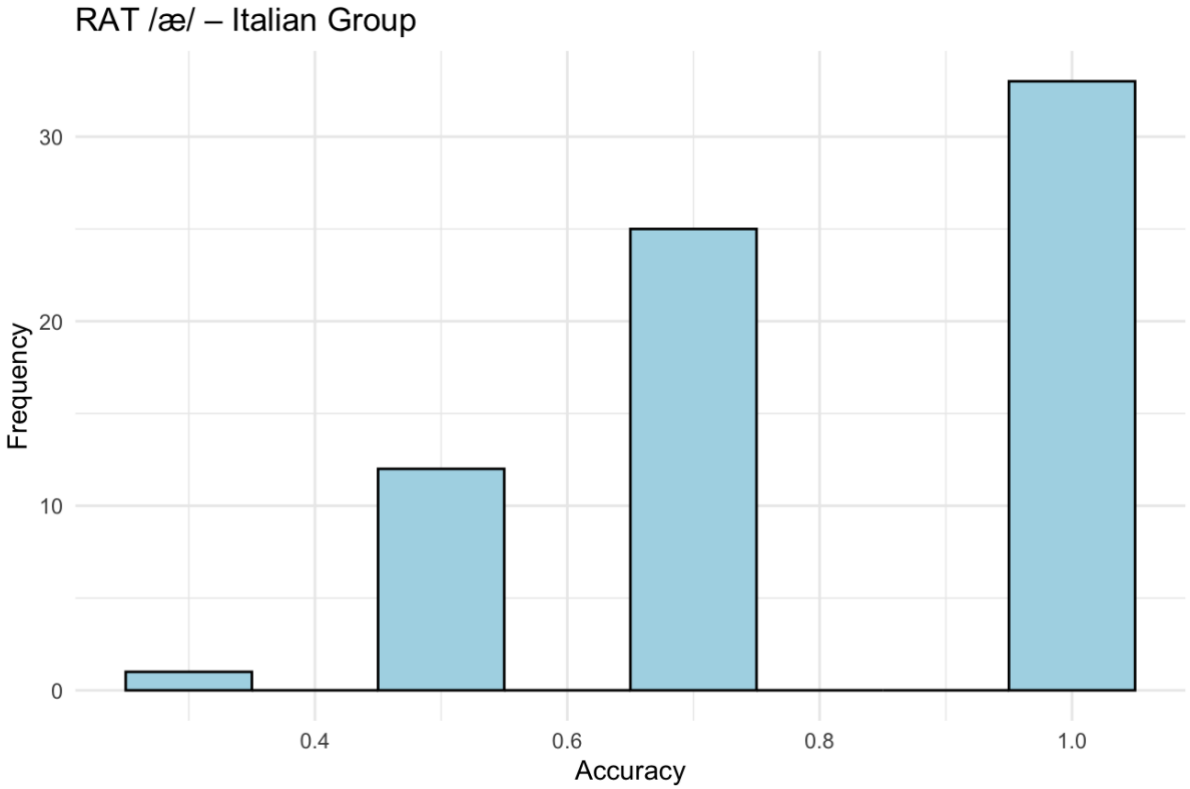

Perception Task: Accuracy per Vowel
Accuracy Score
1.00
0.75
0.50
0.25
0.00
ɑː
æ
ʌ
iː
ɪ
ɒ
əʊ
Vowel

Appendix M: Distribution of Read-Aloud Task accuracy scores per vowel in the Italian Group

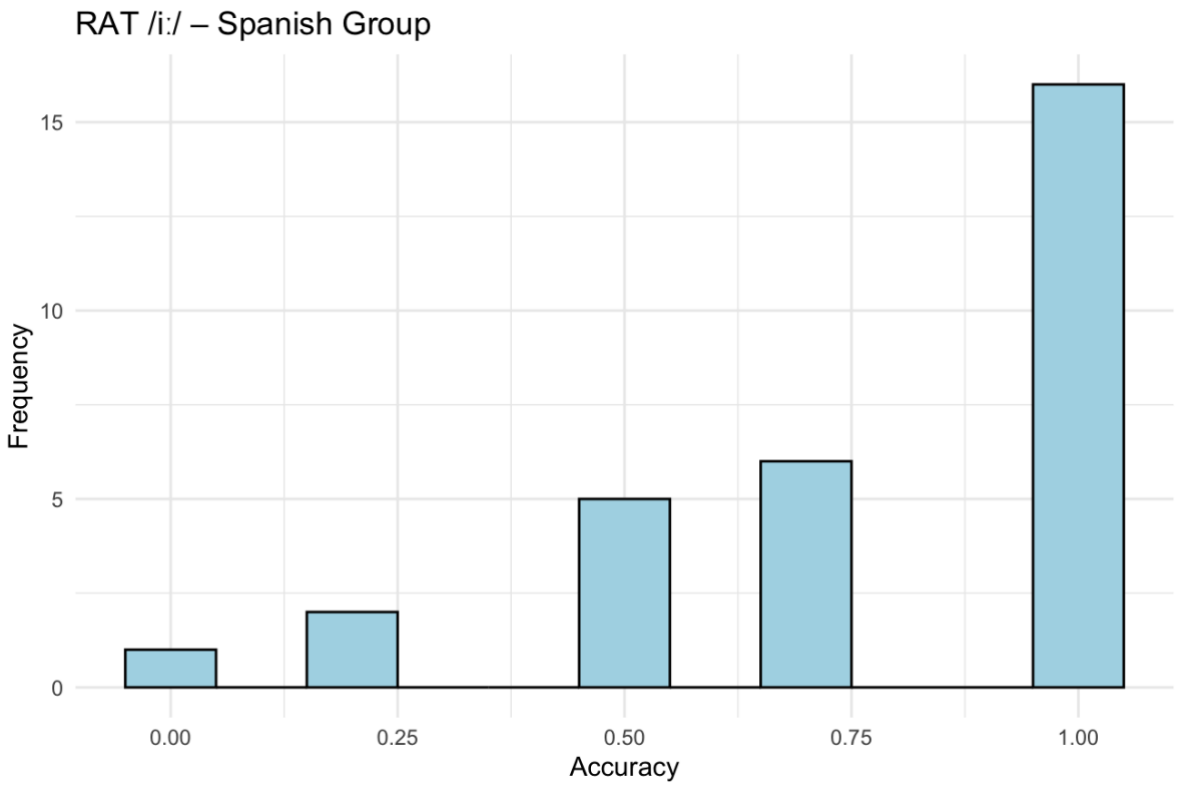


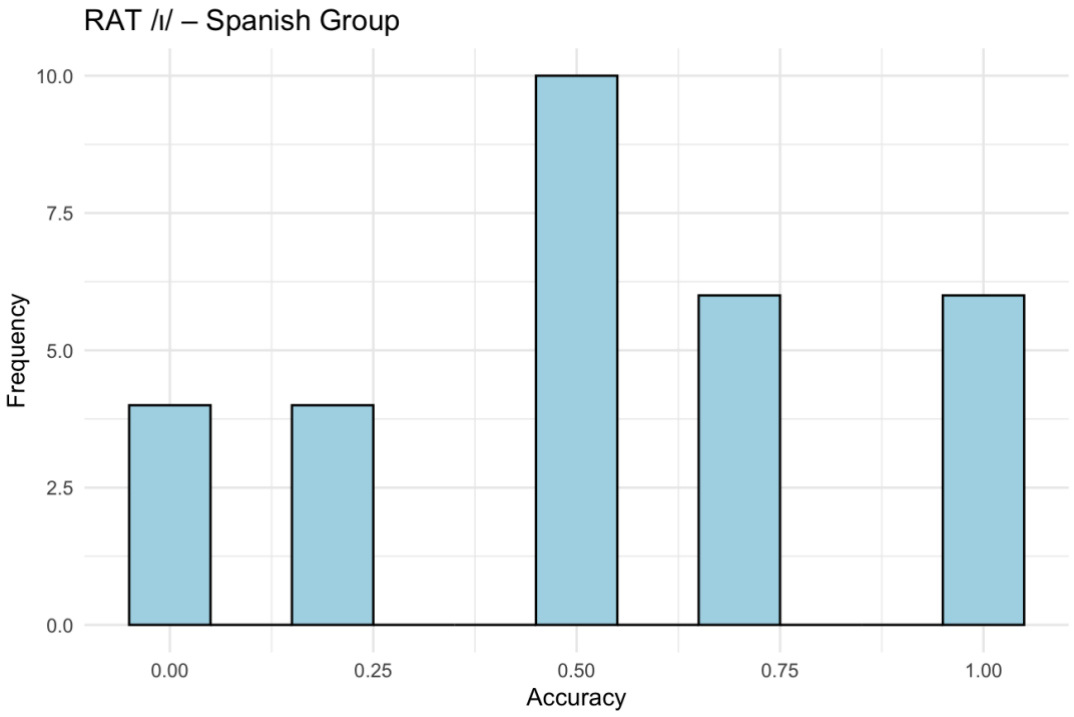


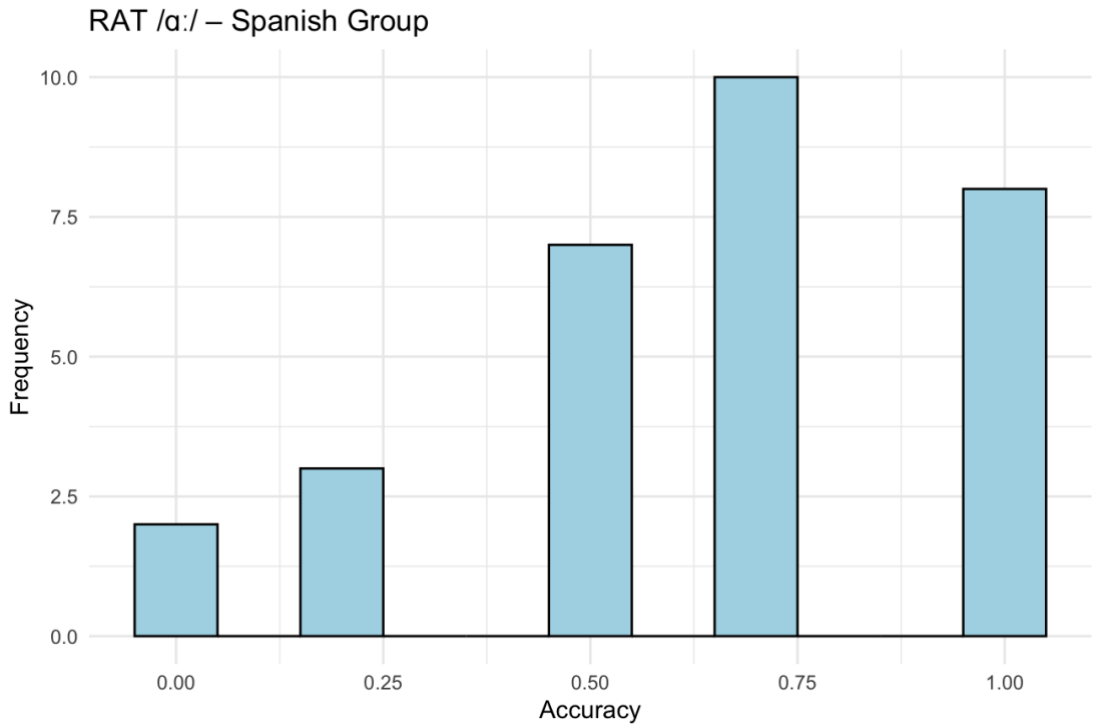

RAT /ɑː/ – Italian Group
Frequency
15
10
5
0
0.00
0.25
0.50
0.75
1.00
Accuracy

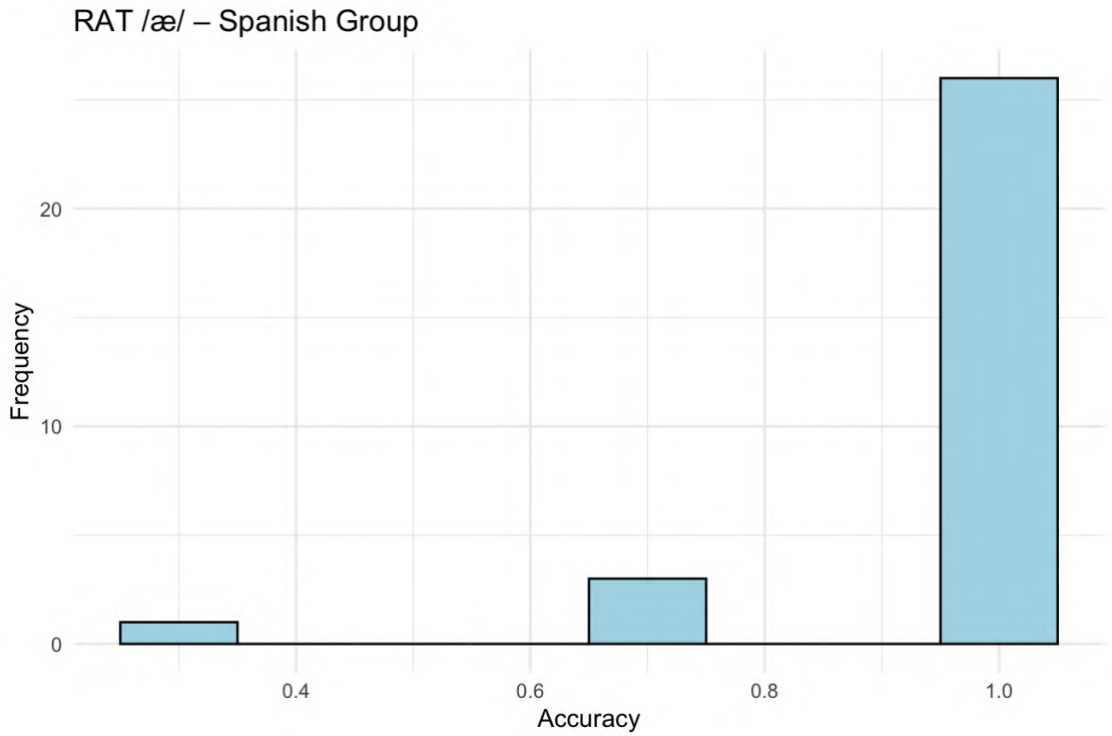

RAT /æ/ – Italian Group
Frequency
30
20
10
0
0.4
0.6
0.8
1.0
Accuracy

Appendix N: Distribution of Read-Aloud Task accuracy scores per vowel in the Spanish Group

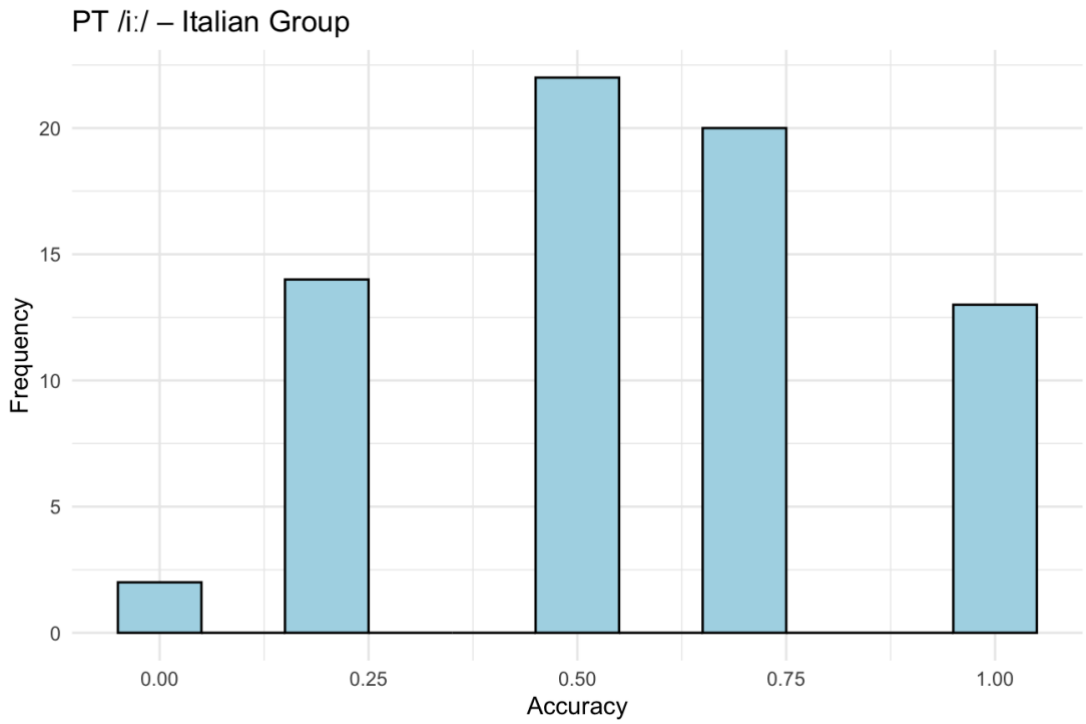


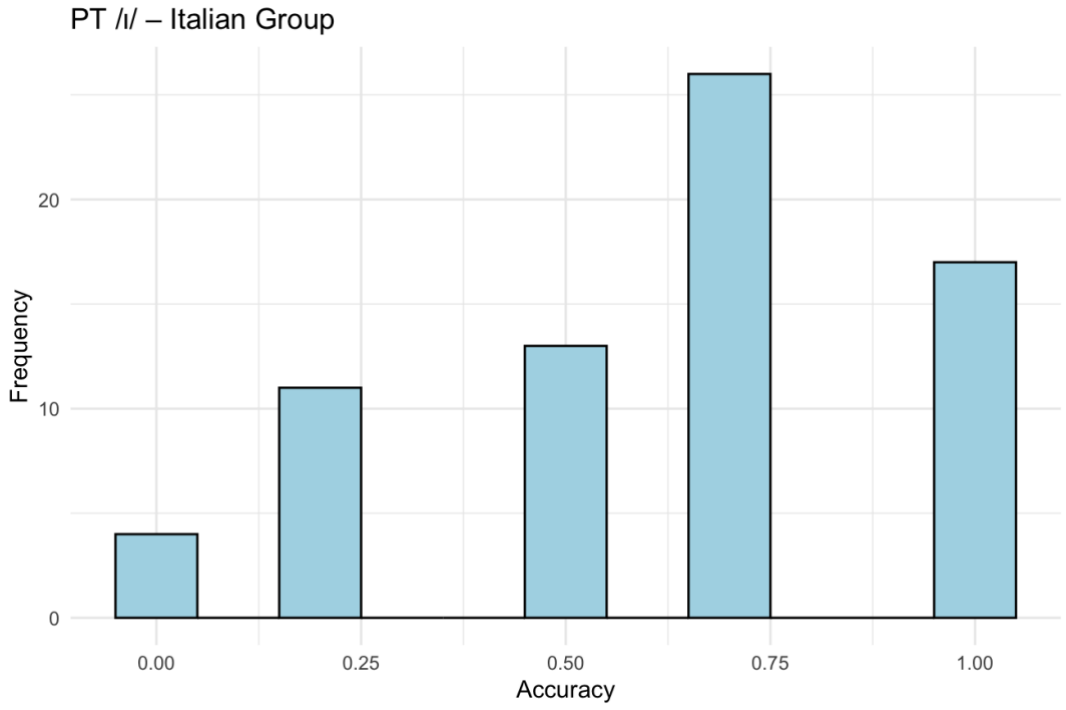


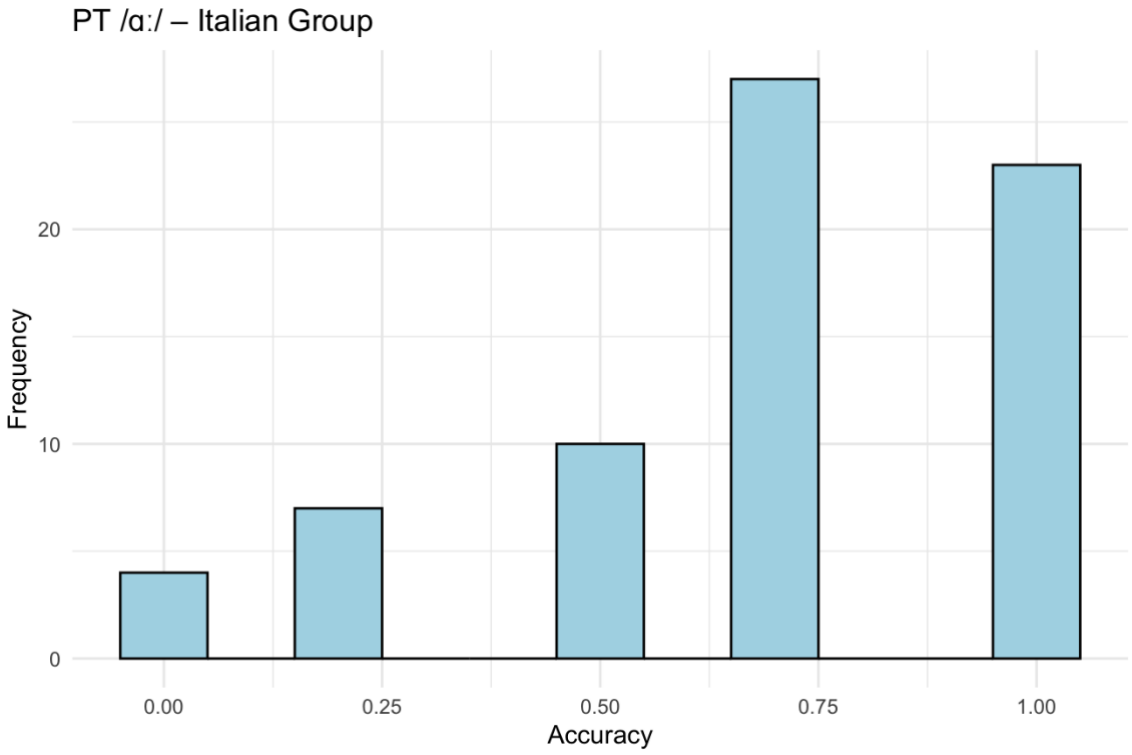

RAT /ɑː/ – Spanish Group
10.0
7.5
5.0
2.5
0.0
Frequency
0.00
0.25
0.50
0.75
1.00
Accuracy

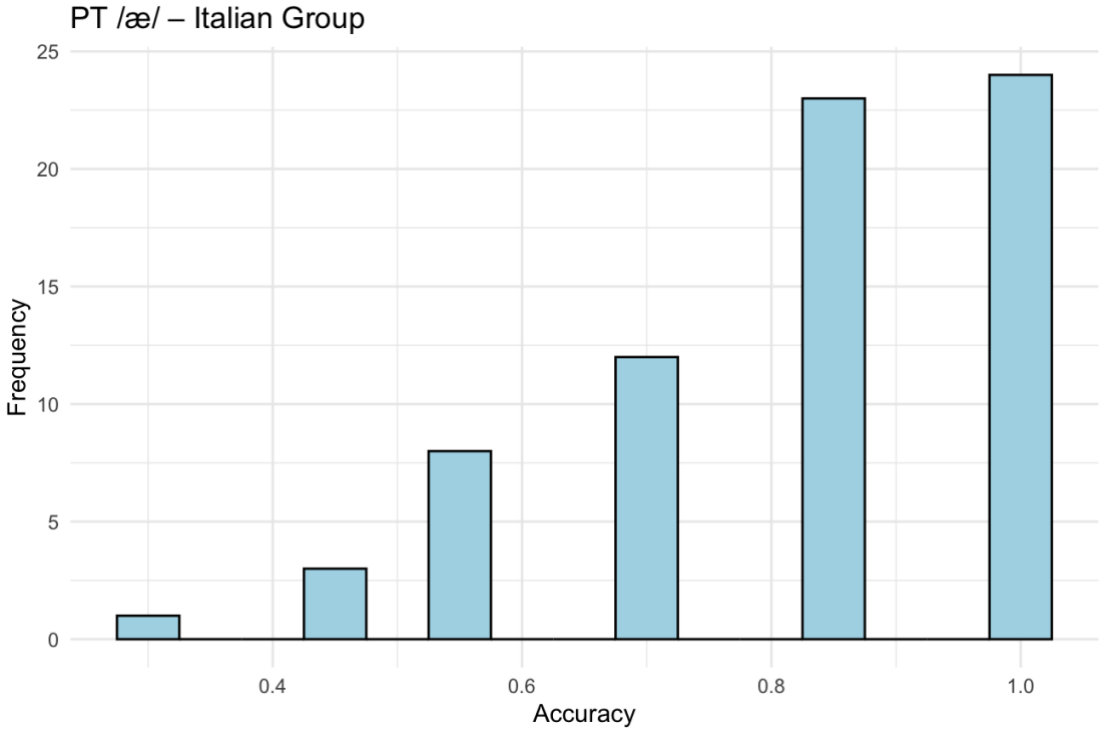

RAT /æ/ – Spanish Group
20
10
0
Frequency
0.4
0.6
0.8
1.0
Accuracy

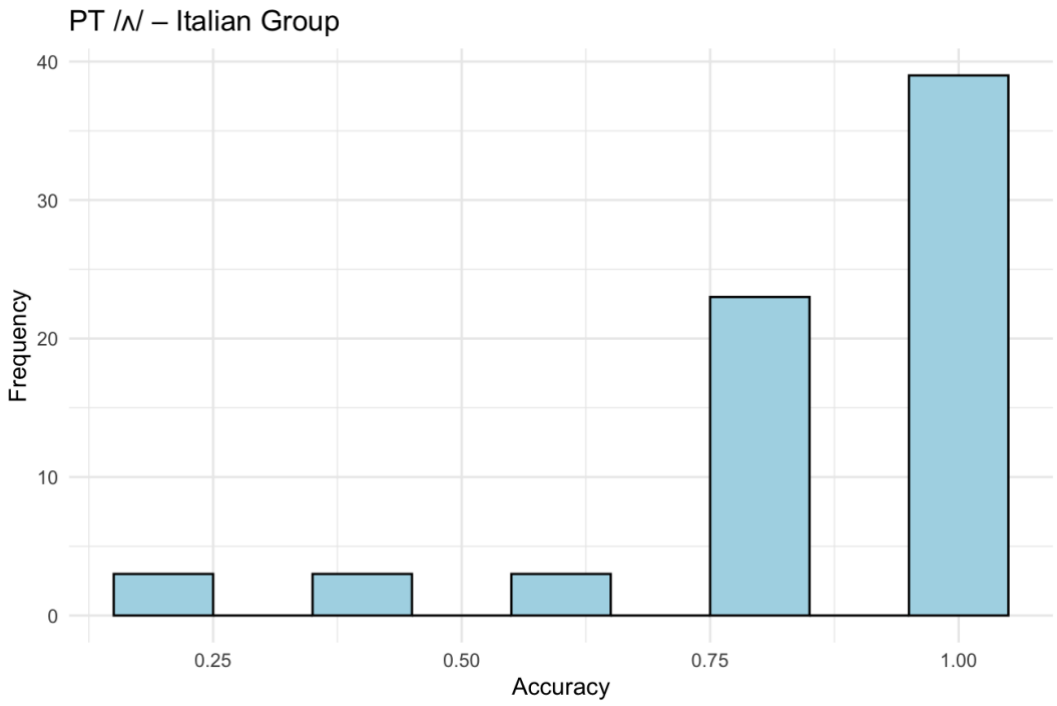
Appendix O: Distribution of Perception Task accuracy scores per vowel in the Italian Group

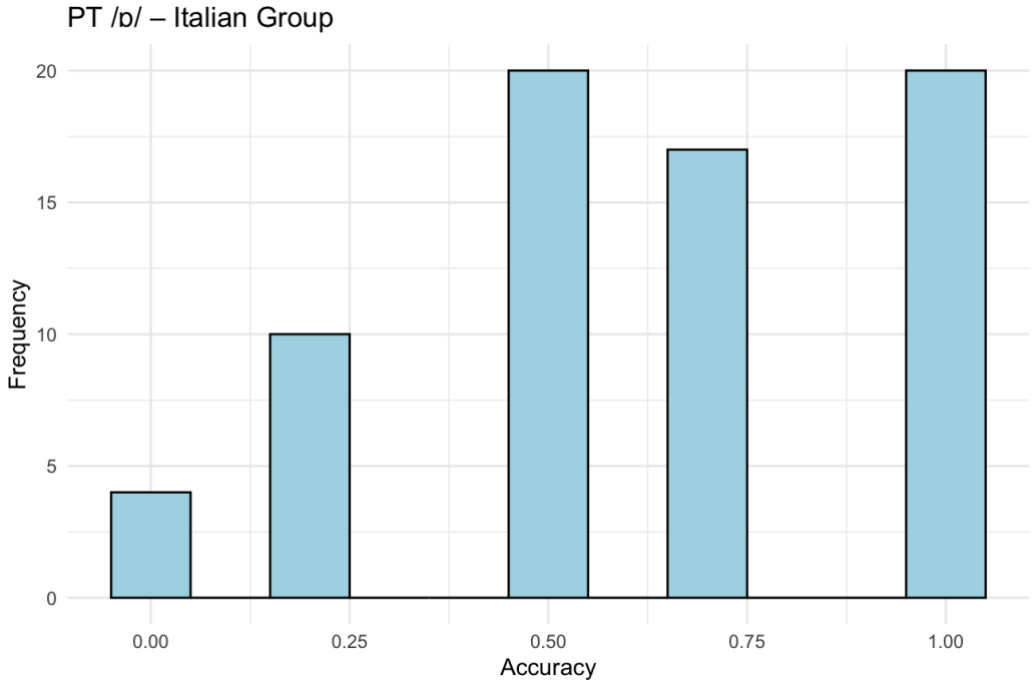


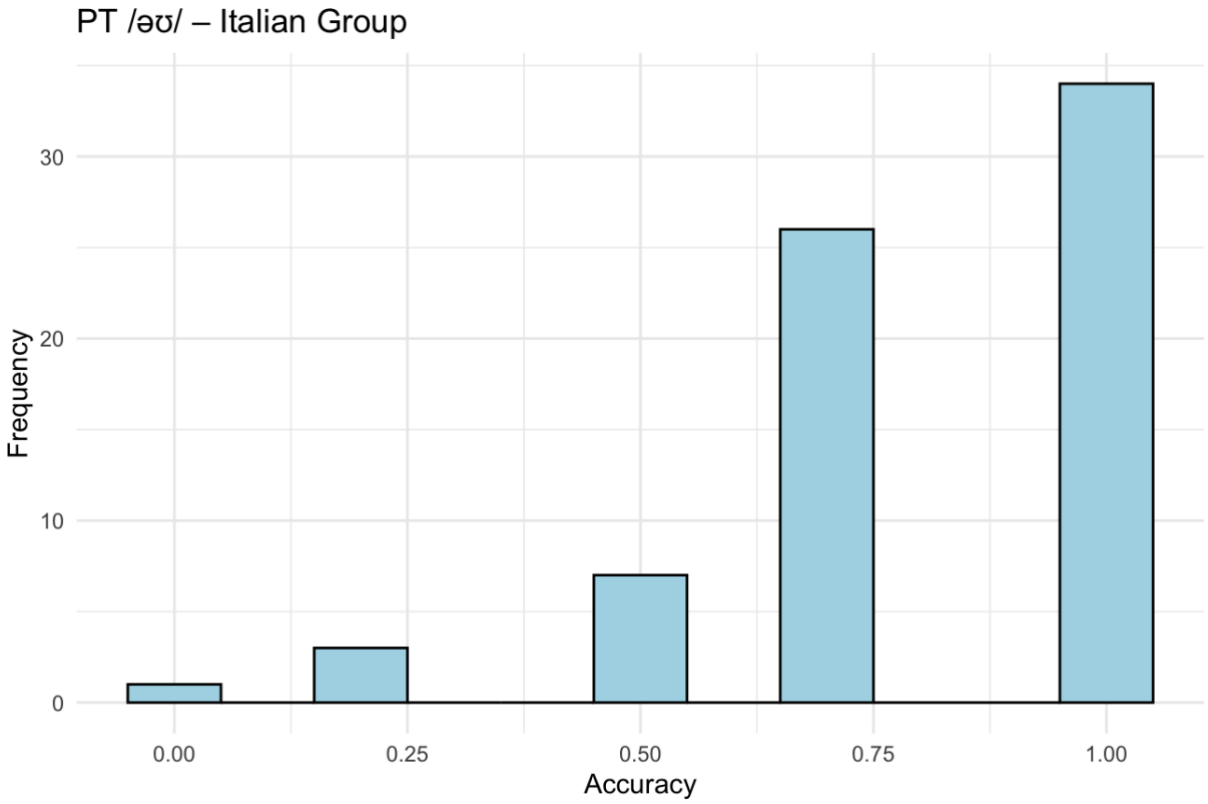



PT /ɑː/ – Italian Group
Frequency
20
10
0
0.00
0.25
0.50
0.75
1.00
Accuracy

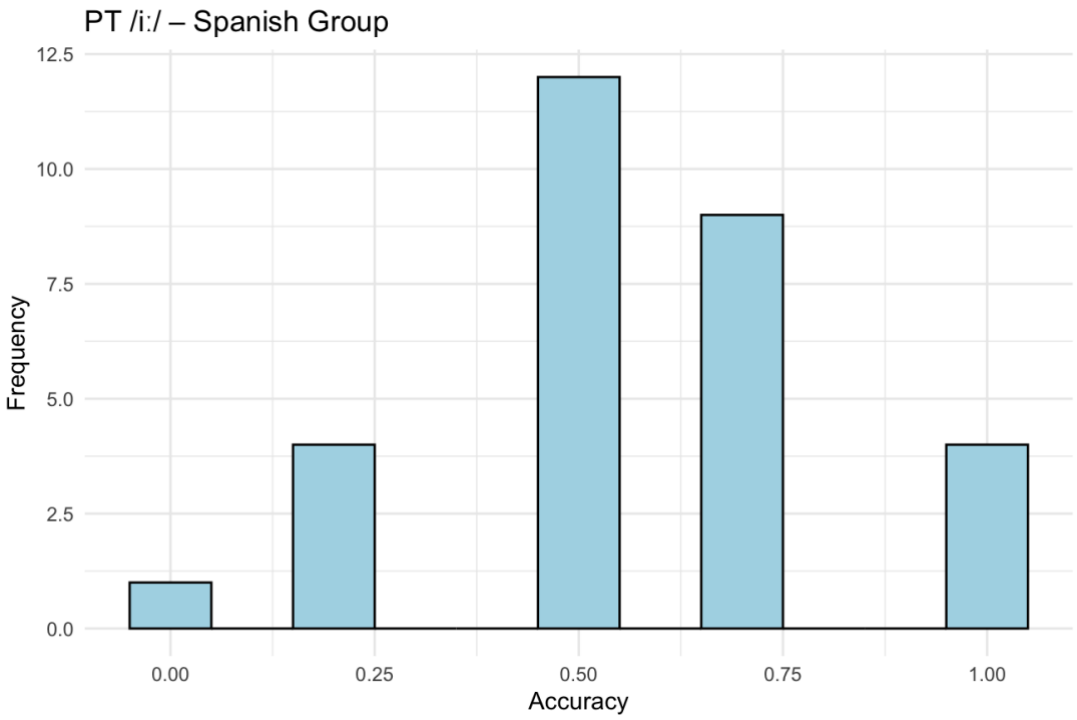

PT /æ/ – Italian Group
Frequency
25
20
15
10
5
0
0.4
0.6
0.8
1.0
Accuracy

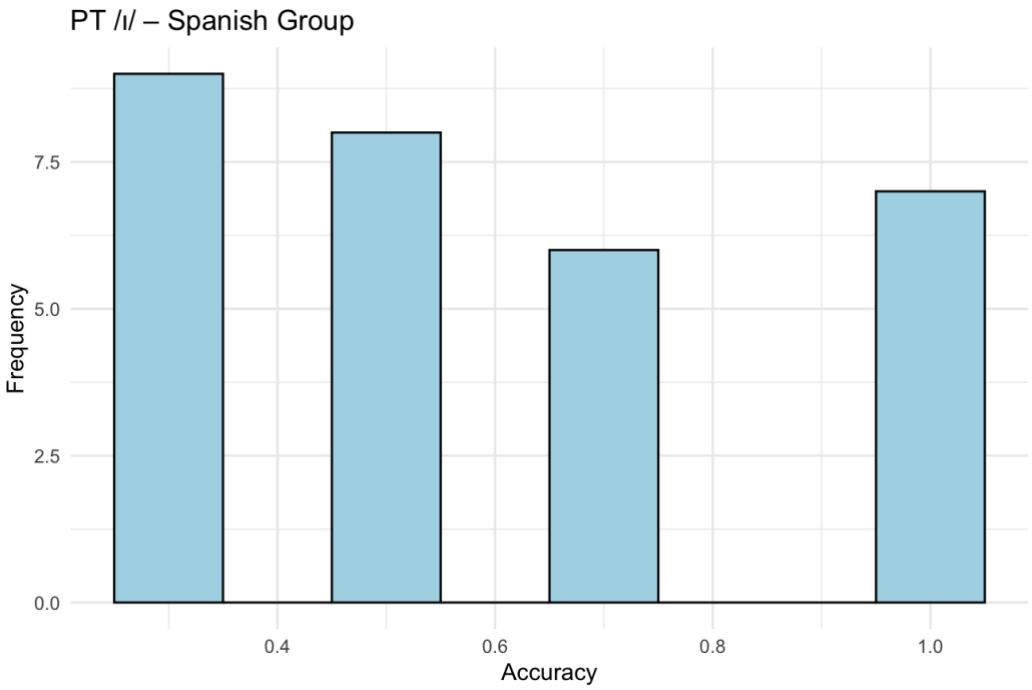

PT /ʌ/ – Italian Group
Frequency
40
30
20
10
0
0.25
0.50
0.75
1.00
Accuracy

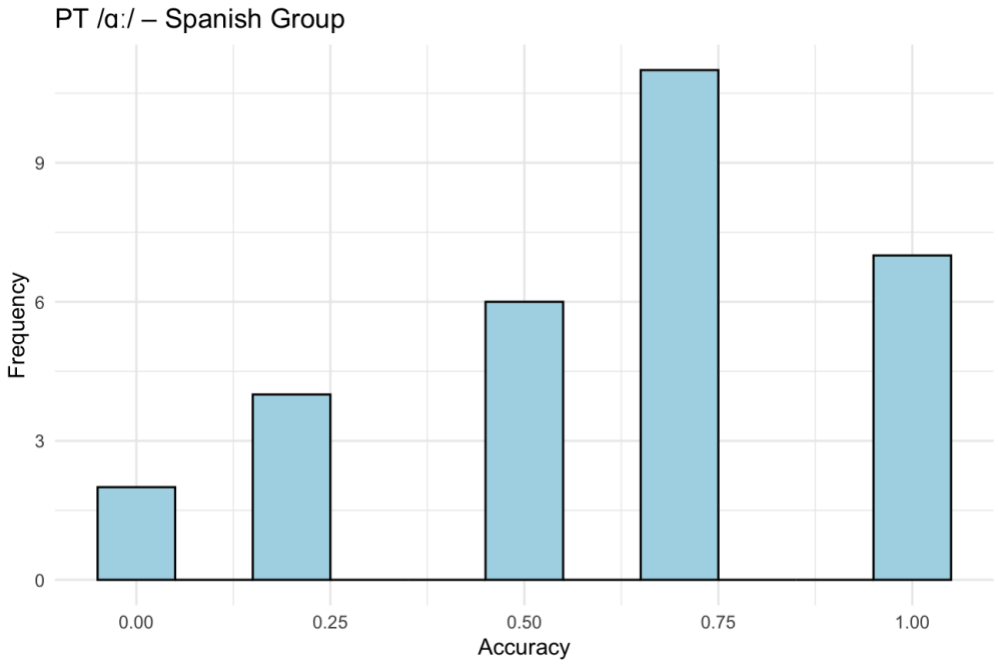

PT /ɒ/ – Italian Group
Frequency
20
15
10
5
0
0.00
0.25
0.50
0.75
1.00
Accuracy

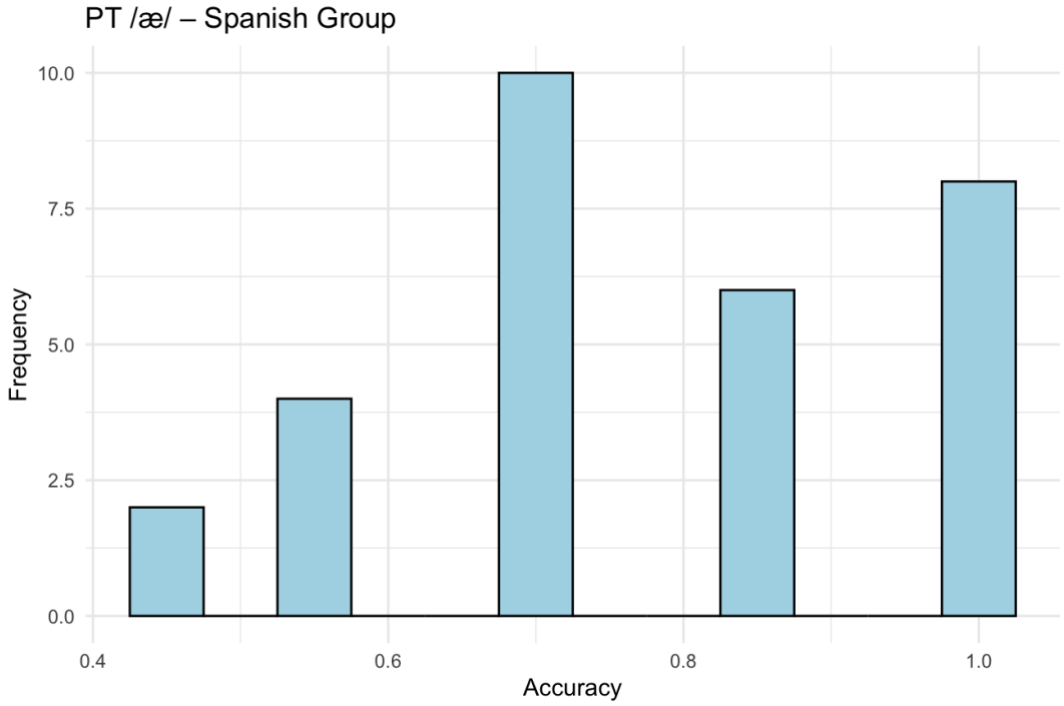

PT /əʊ/ – Italian Group
Frequency
30
20
10
0
0.00
0.25
0.50
0.75
1.00
Accuracy

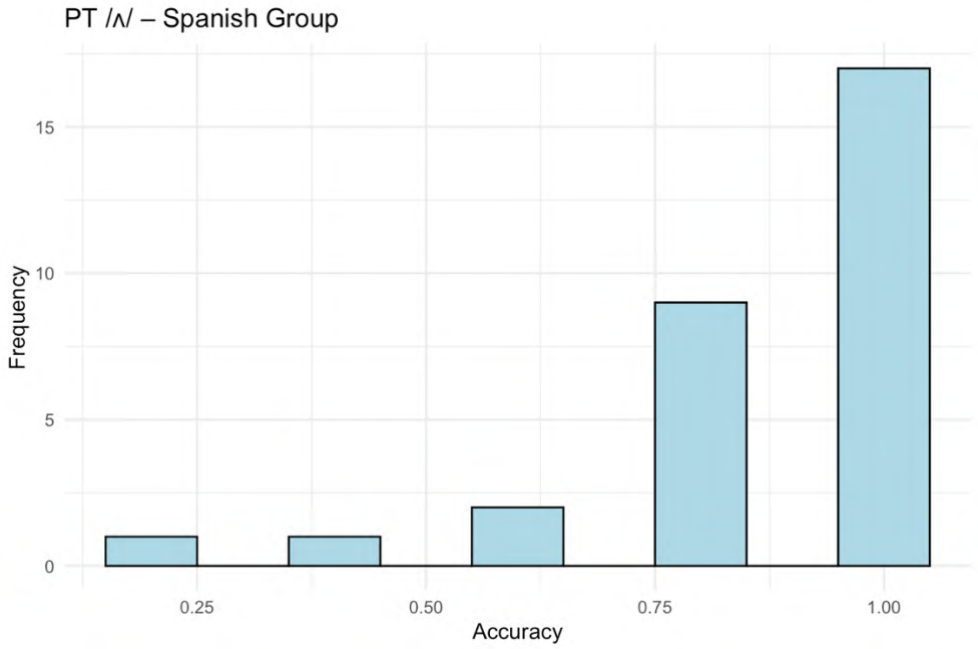
Appendix P: Distribution of Perception Task accuracy scores per vowel in the Spanish Group

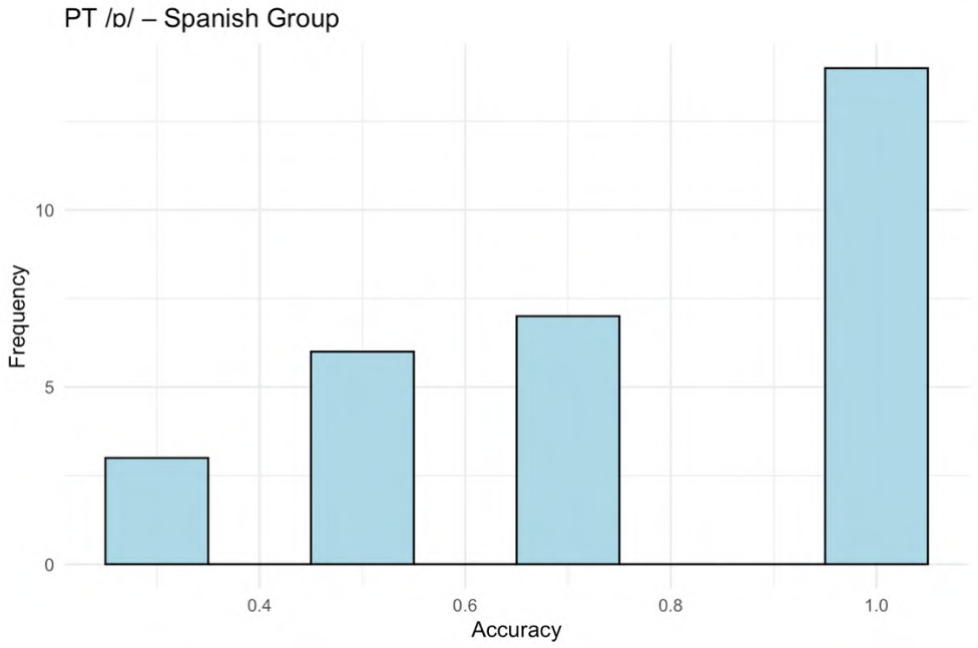


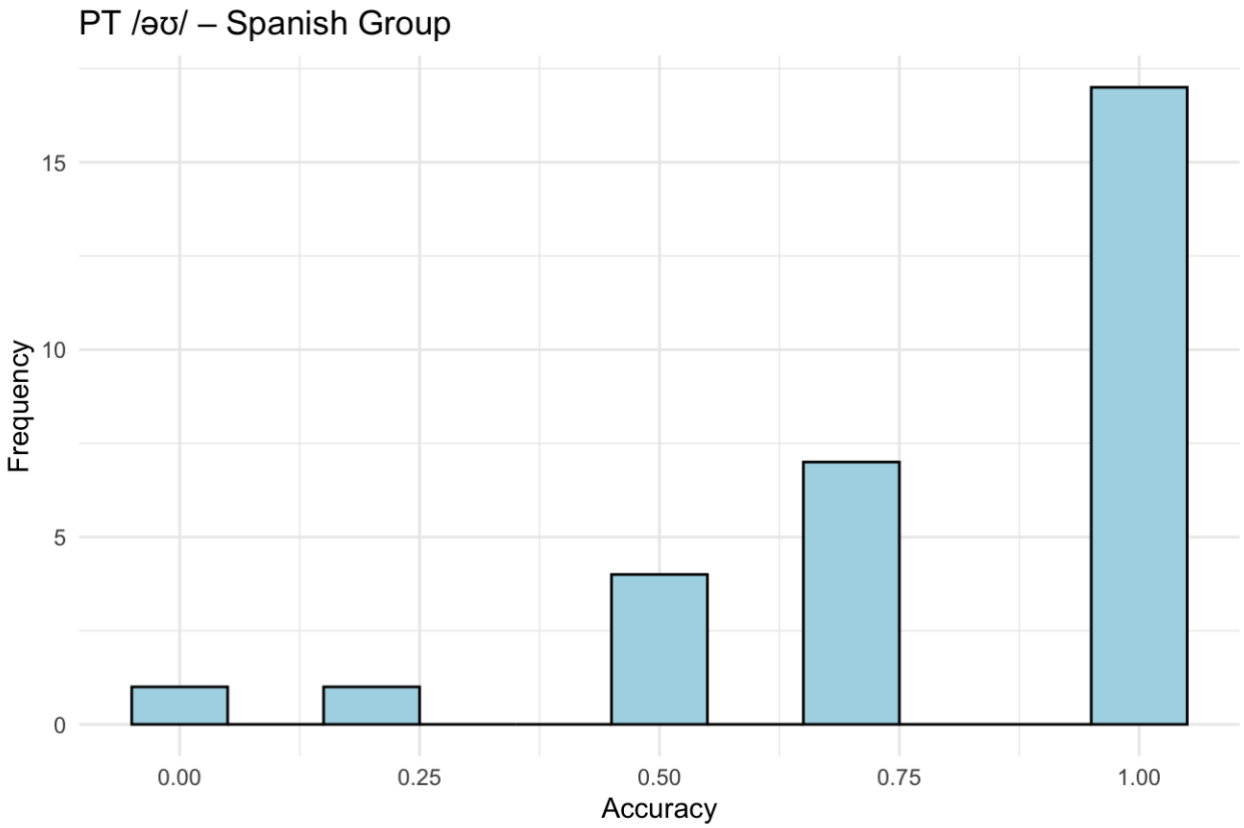



PT /ɑː/ – Spanish Group
Frequency
9
6
3
0
0.00
0.25
0.50
0.75
1.00
Accuracy

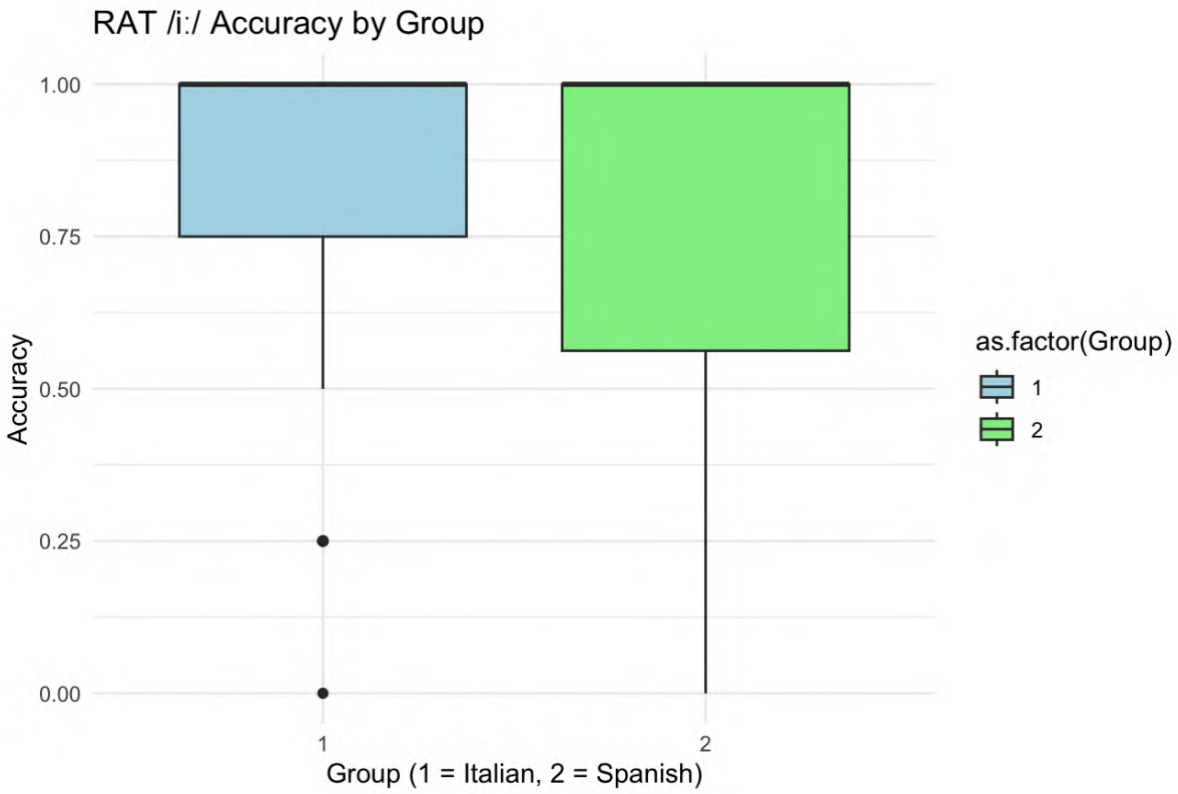

PT /æ/ – Spanish Group
Frequency
10.0
7.5
5.0
2.5
0.0
0.4
0.6
0.8
1.0
Accuracy

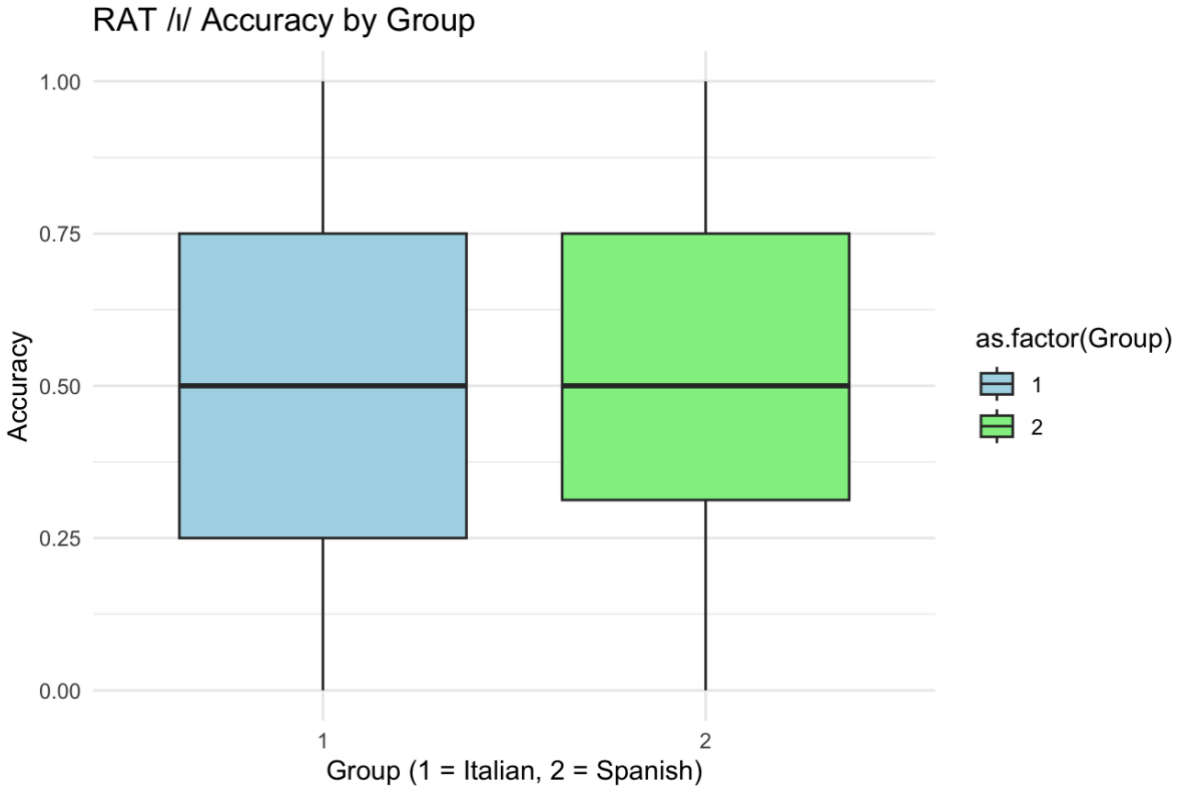

PT /ʌ/ – Spanish Group
Frequency
15
10
5
0
0.25
0.50
0.75
1.00
Accuracy

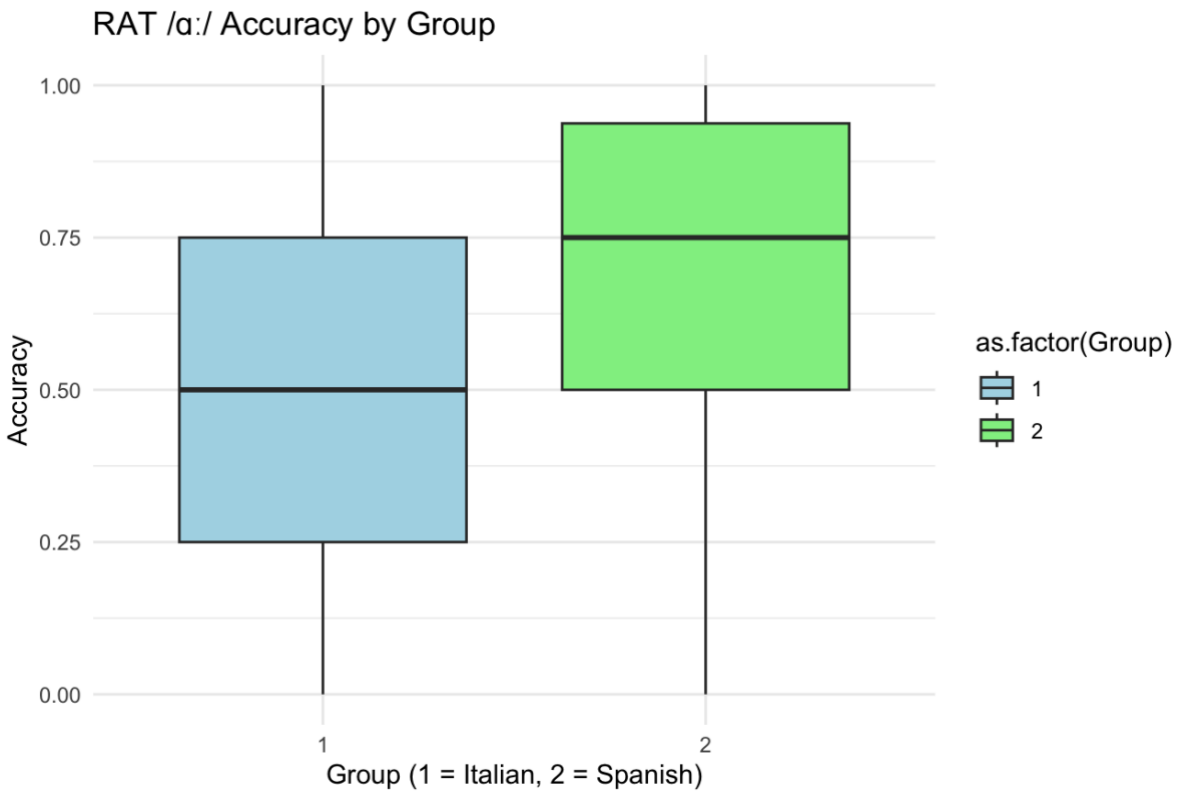

PT /ɒ/ – Spanish Group
Frequency
10
5
0
0.4
0.6
0.8
1.0
Accuracy

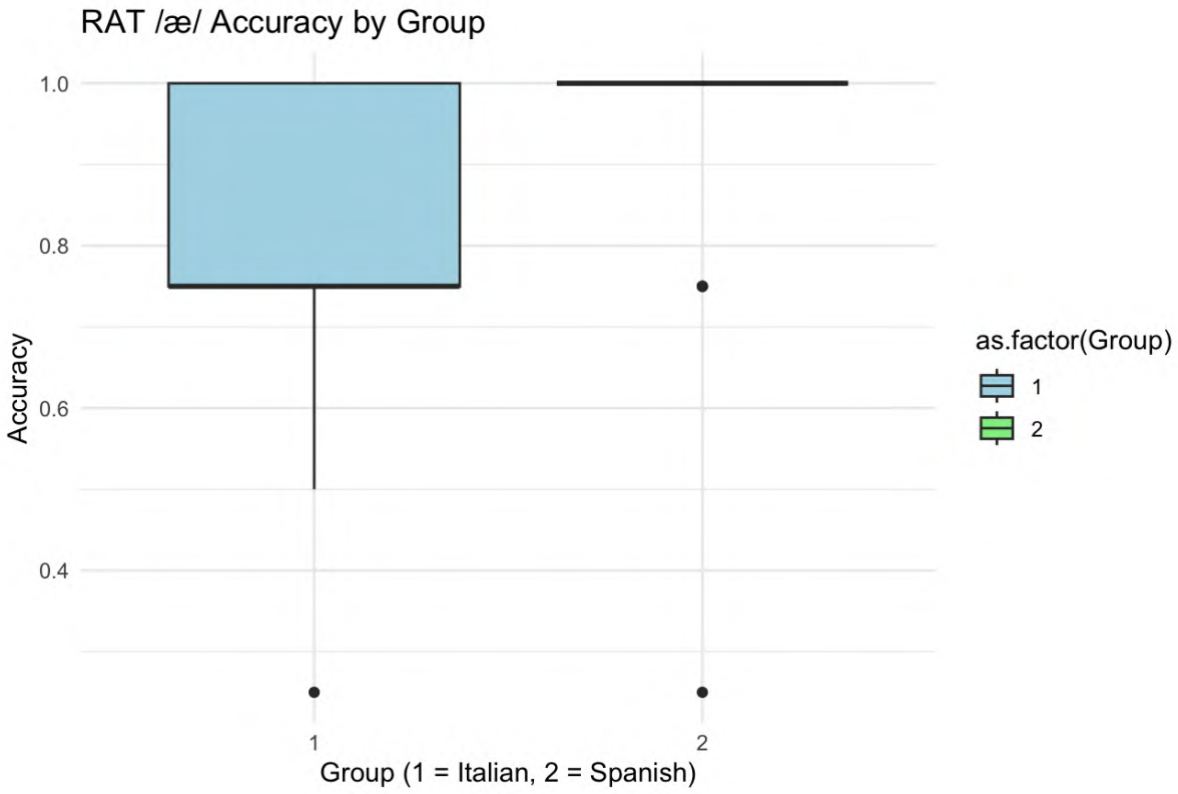

PT /əʊ/ – Spanish Group
Frequency
15
10
5
0
0.00
0.25
0.50
0.75
1.00
Accuracy

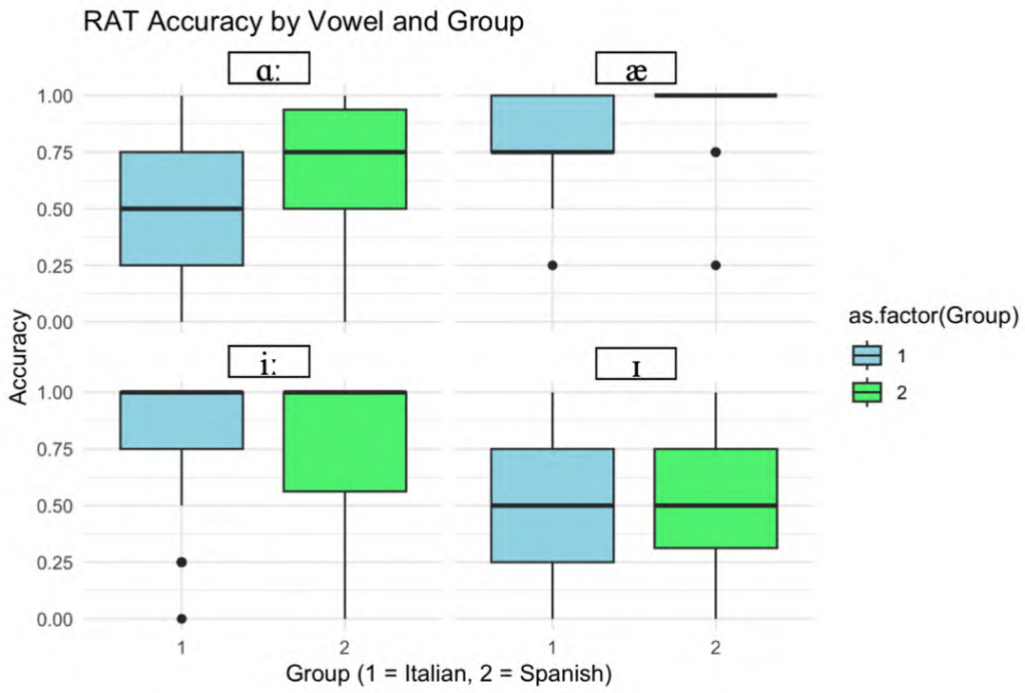
Appendix Q: Boxplots of Read-Aloud Task Accuracy by Vowel and Group



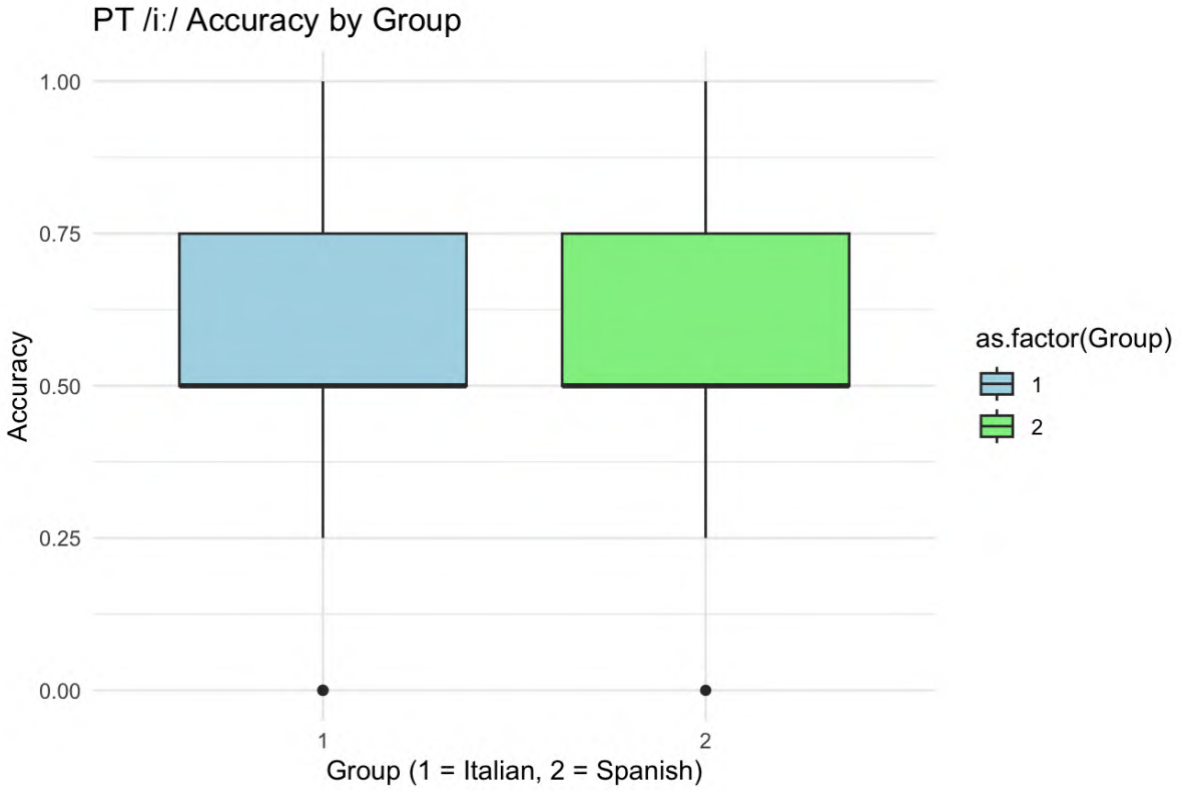


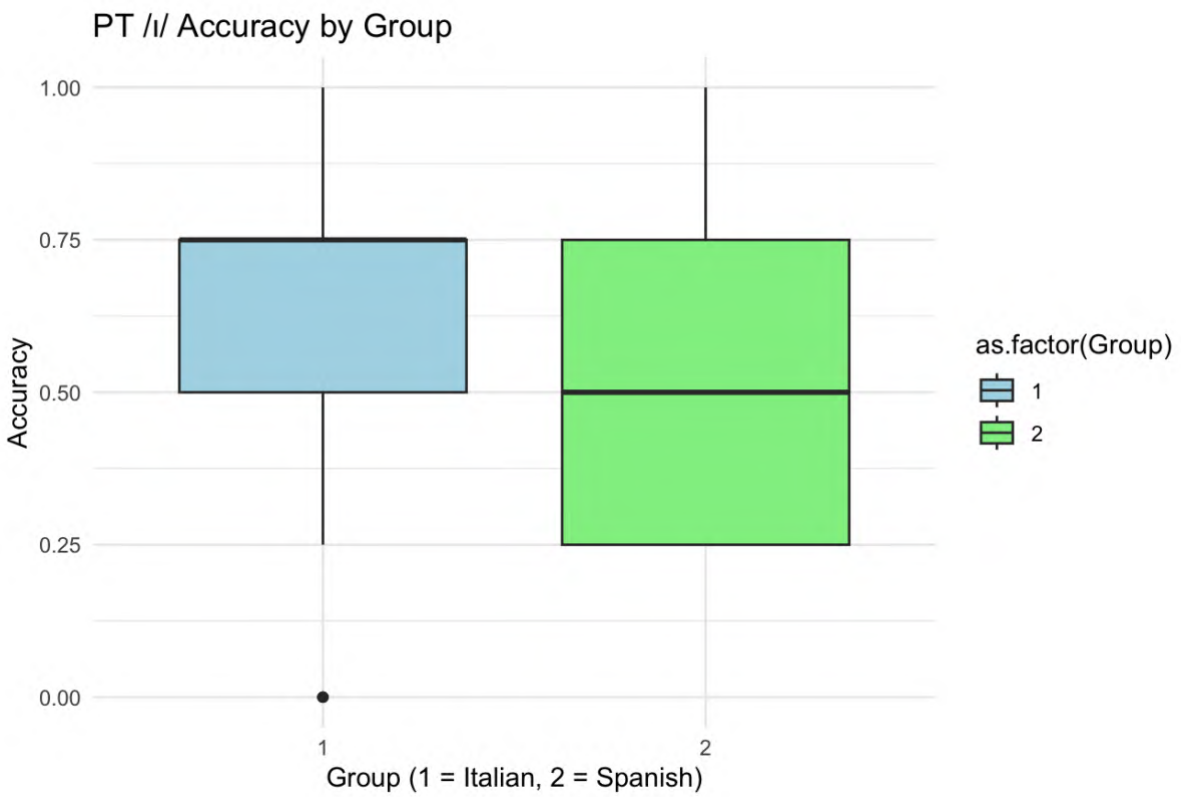

RAT /ɑː/ Accuracy by Group
1.00
0.75
0.50
0.25
0.00
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

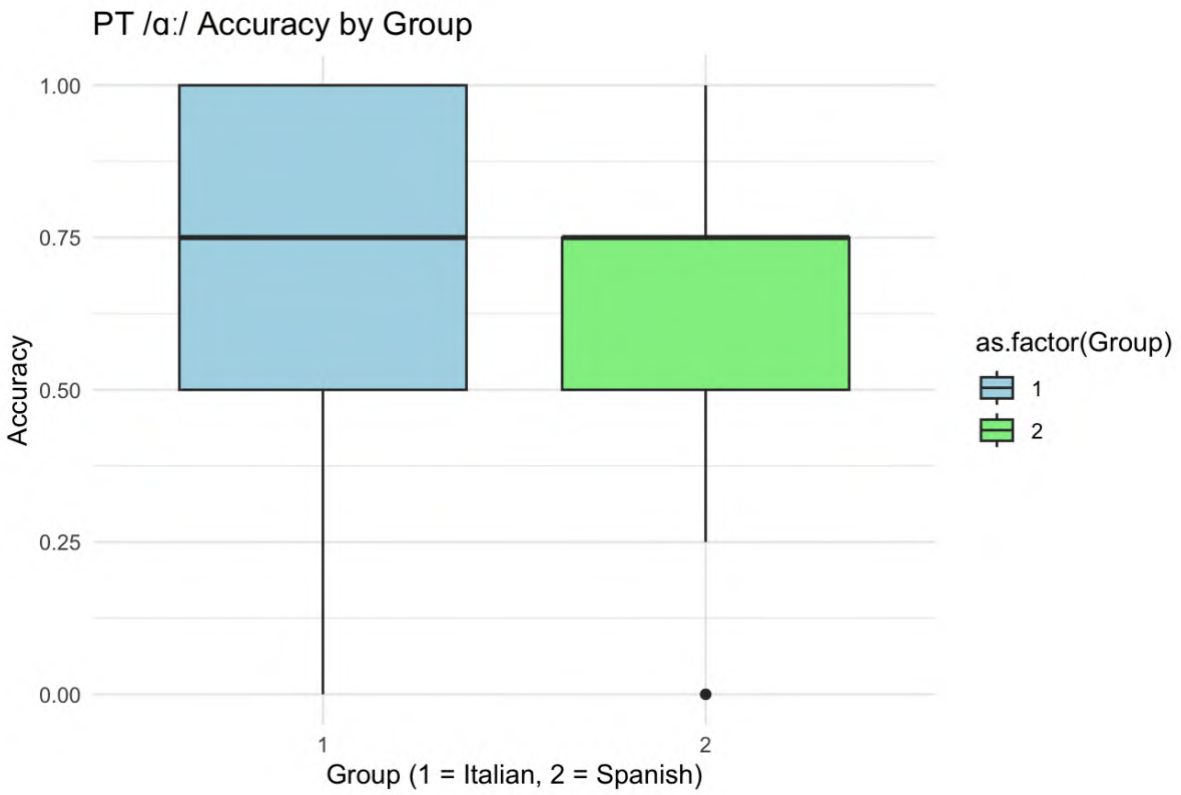

RAT /æ/ Accuracy by Group
1.0
0.8
0.6
0.4
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

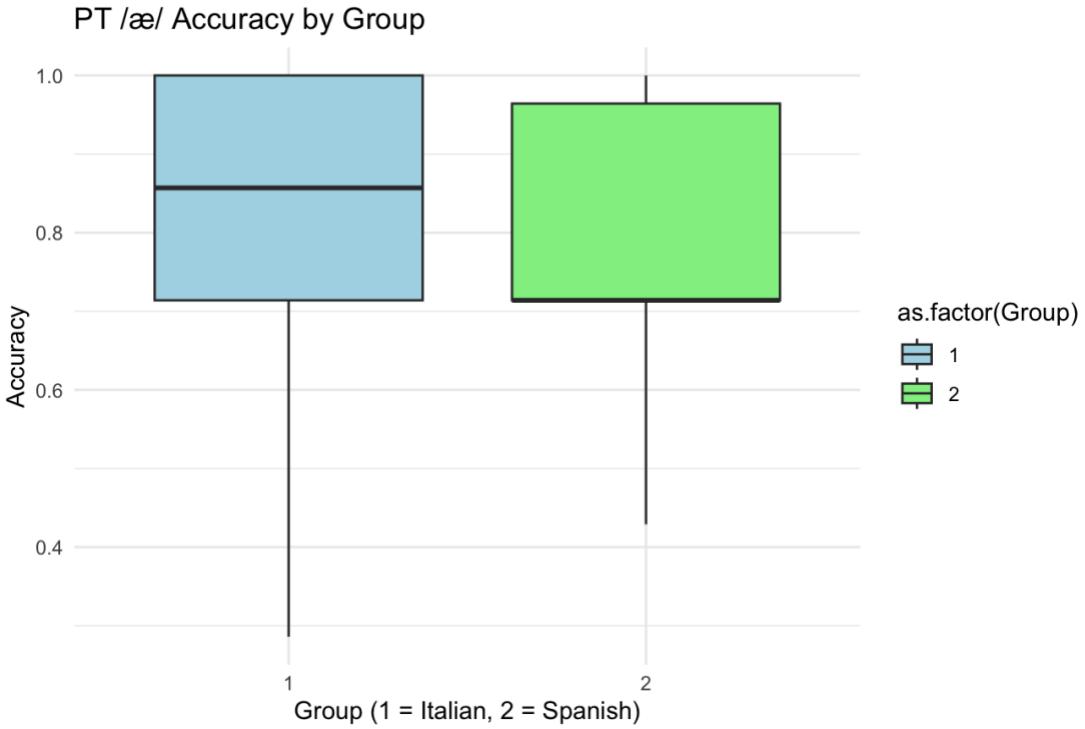

RAT Accuracy by Vowel and Group
ɑː
æ
iː
ɪ
Accuracy
1.00
0.75
0.50
0.25
0.00
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

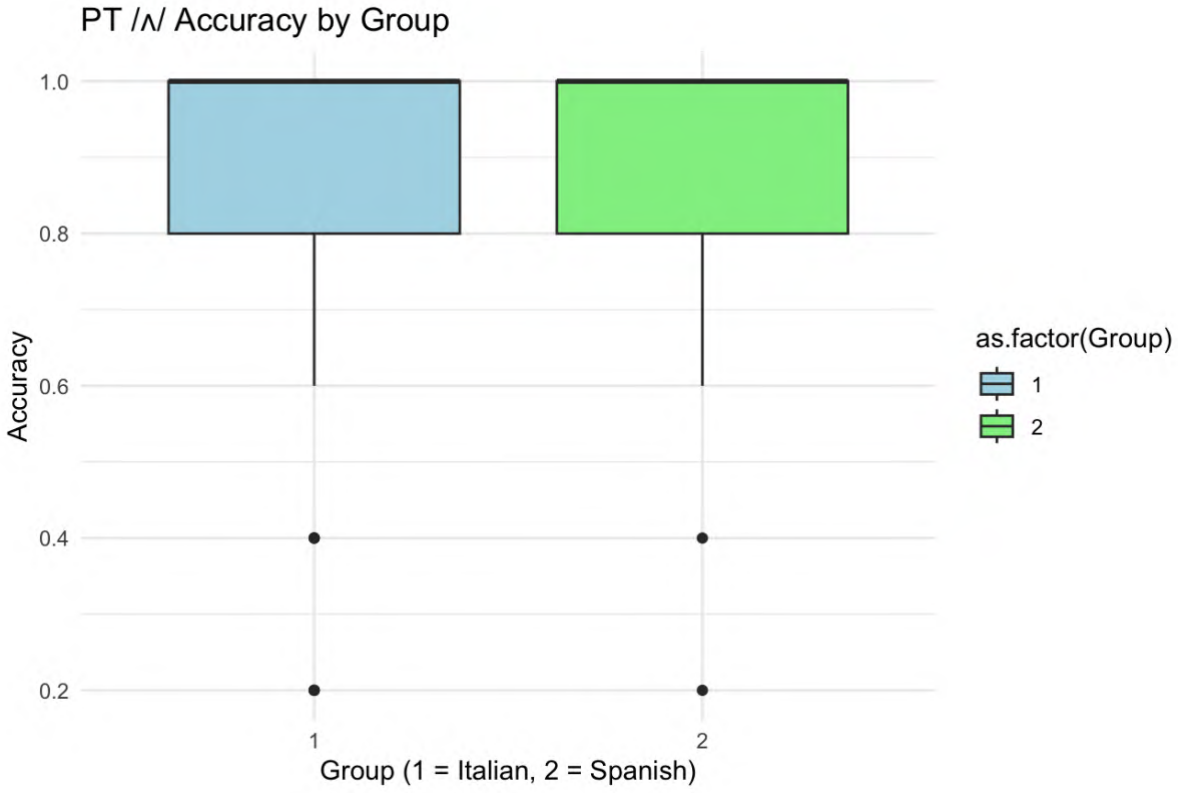
Appendix R: Boxplots of Perception Task Accuracy by Vowel and Group

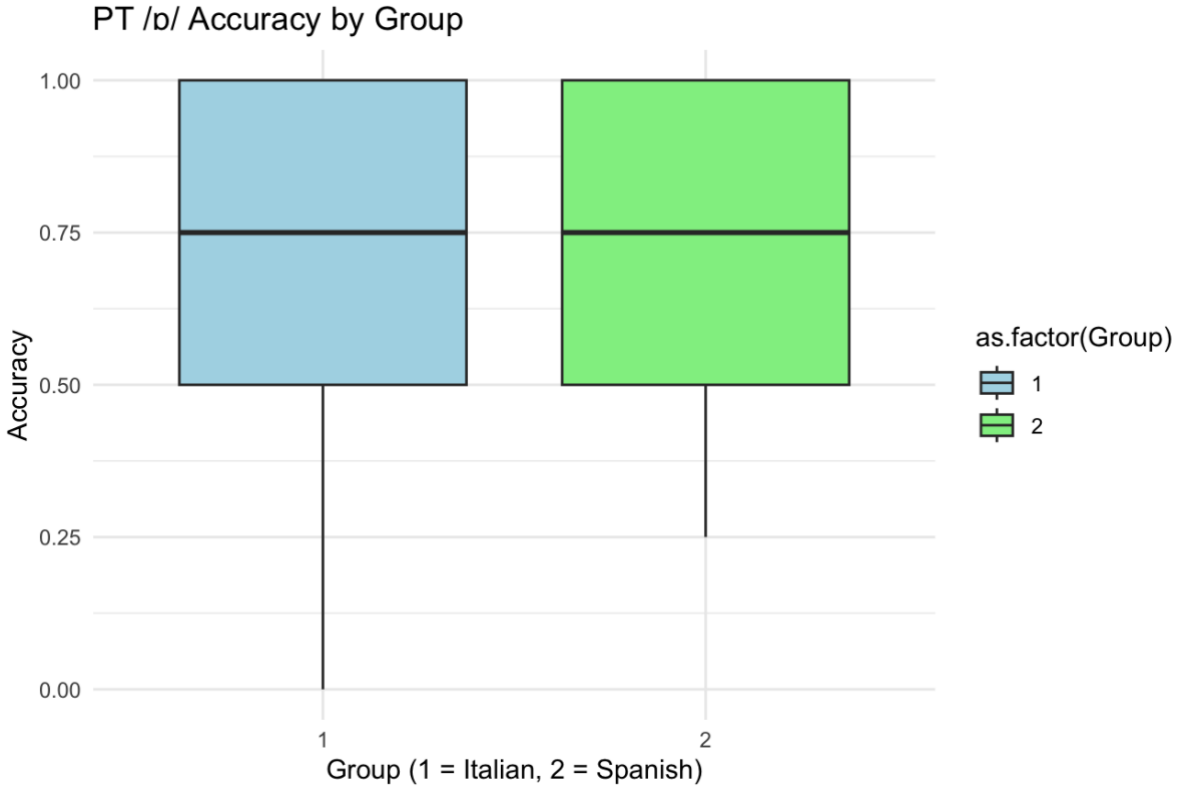


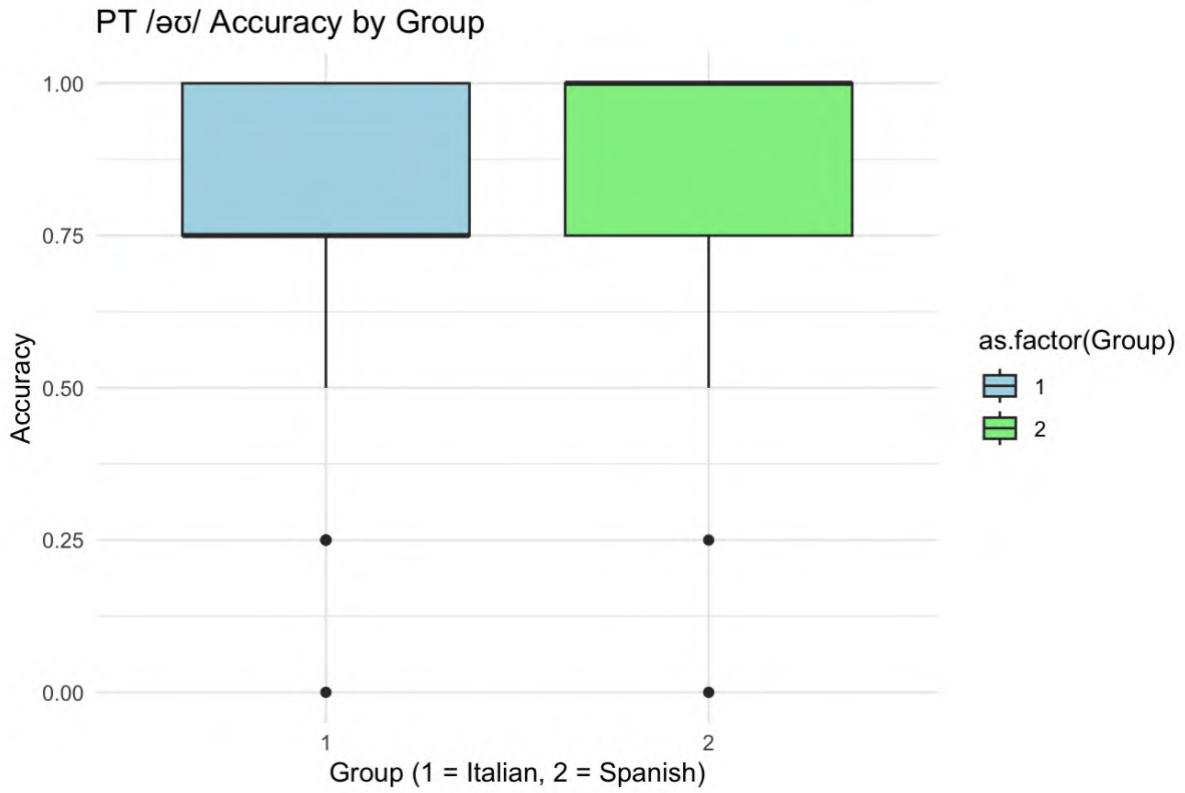


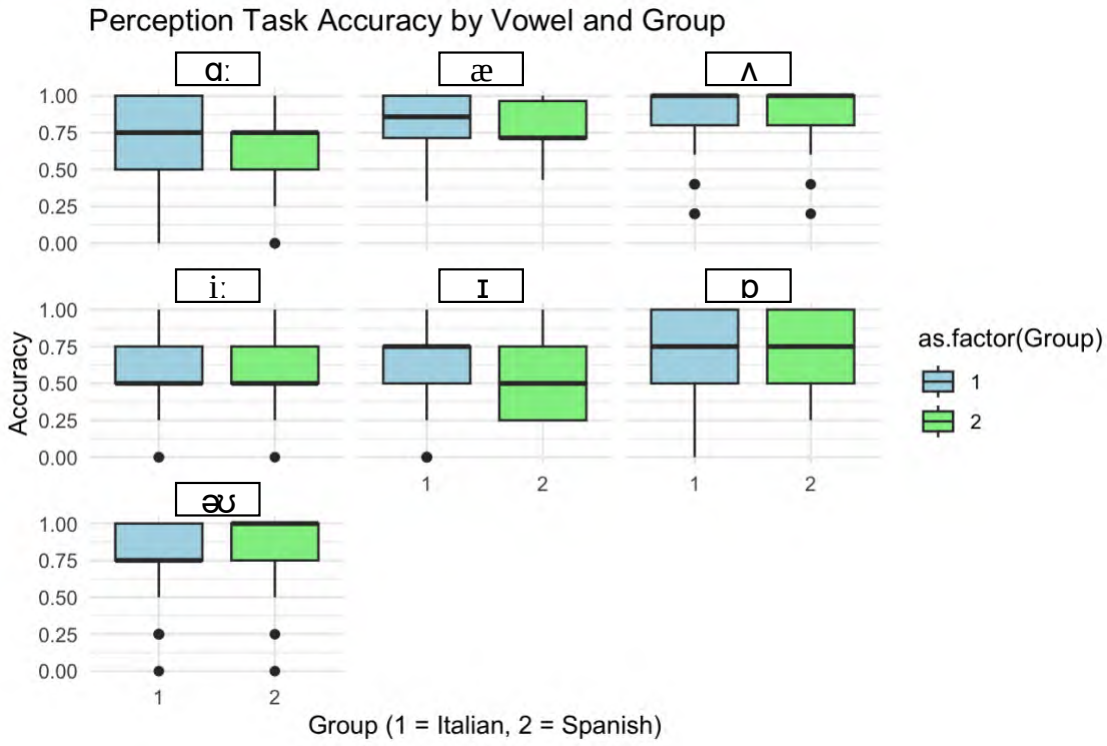

PT /ɑː/ Accuracy by Group
1.00
0.75
0.50
0.25
0.00
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2


PT /æ/ Accuracy by Group
1.0
0.8
0.6
0.4
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

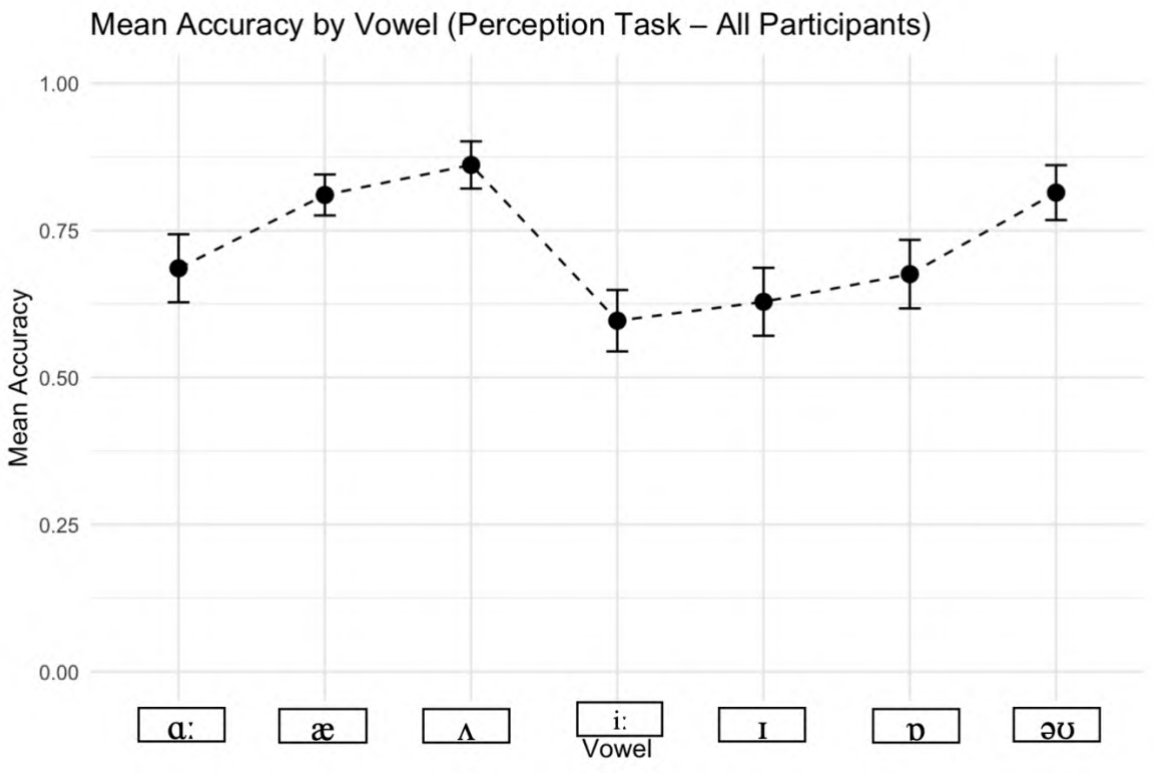

PT /ʌ/ Accuracy by Group
1.0
0.8
0.6
0.4
0.2
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2


PT /ɒ/ Accuracy by Group
1.00
0.75
0.50
0.25
0.00
Accuracy
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

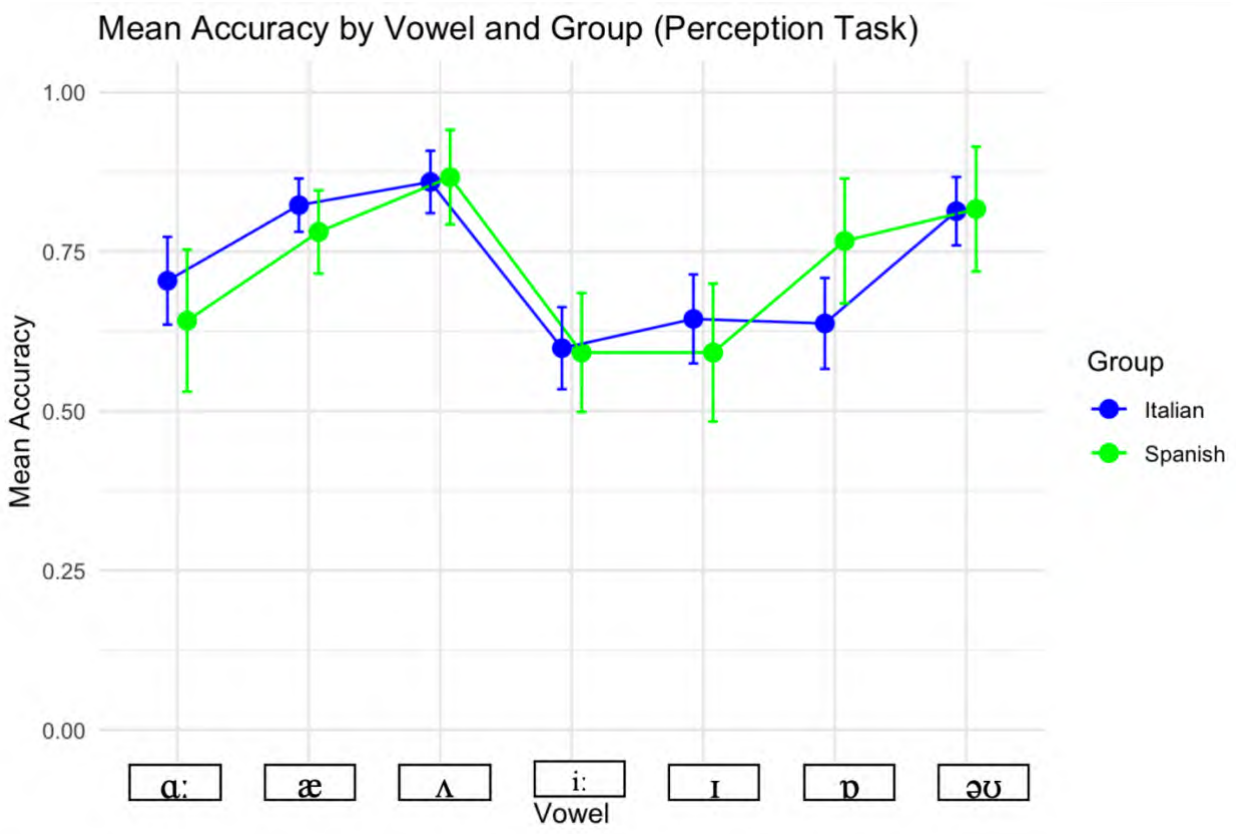

PT /əʊ/ Accuracy by Group
Accuracy
1.00
0.75
0.50
0.25
0.00
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2


Perception Task Accuracy by Vowel and Group
ɑː
æ
ʌ
iː
ɪ
ɒ
əʊ
Accuracy
1.00
0.75
0.50
0.25
0.00
1
2
Group (1 = Italian, 2 = Spanish)
as.factor(Group)
1
2

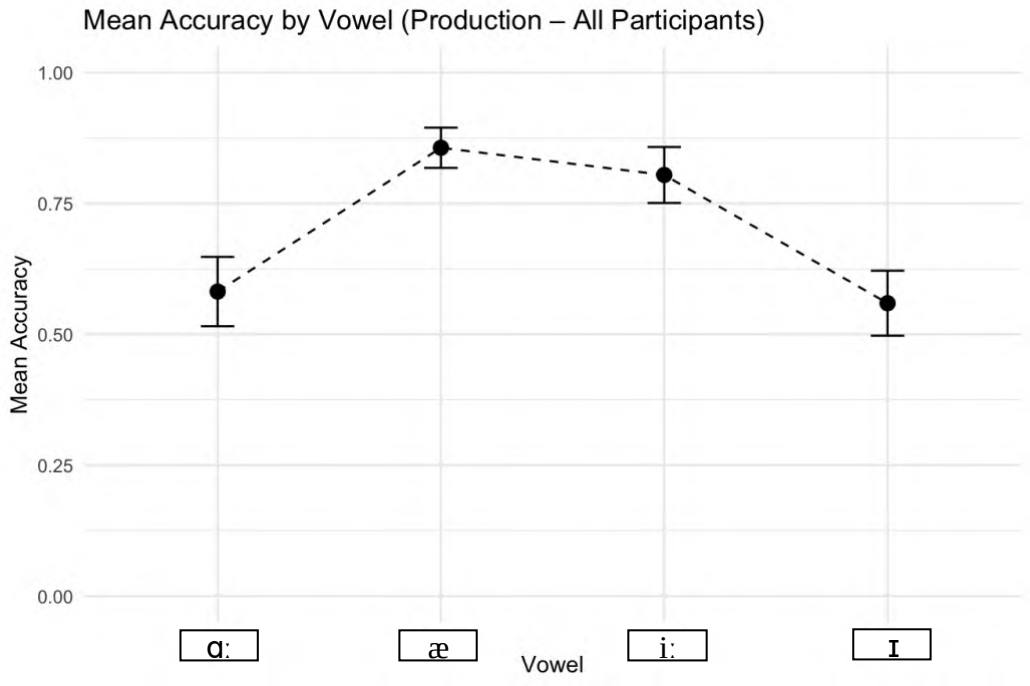
Appendix S: Point-and-Line Plot of Mean Perception Accuracy by Vowel (Whole Sample)



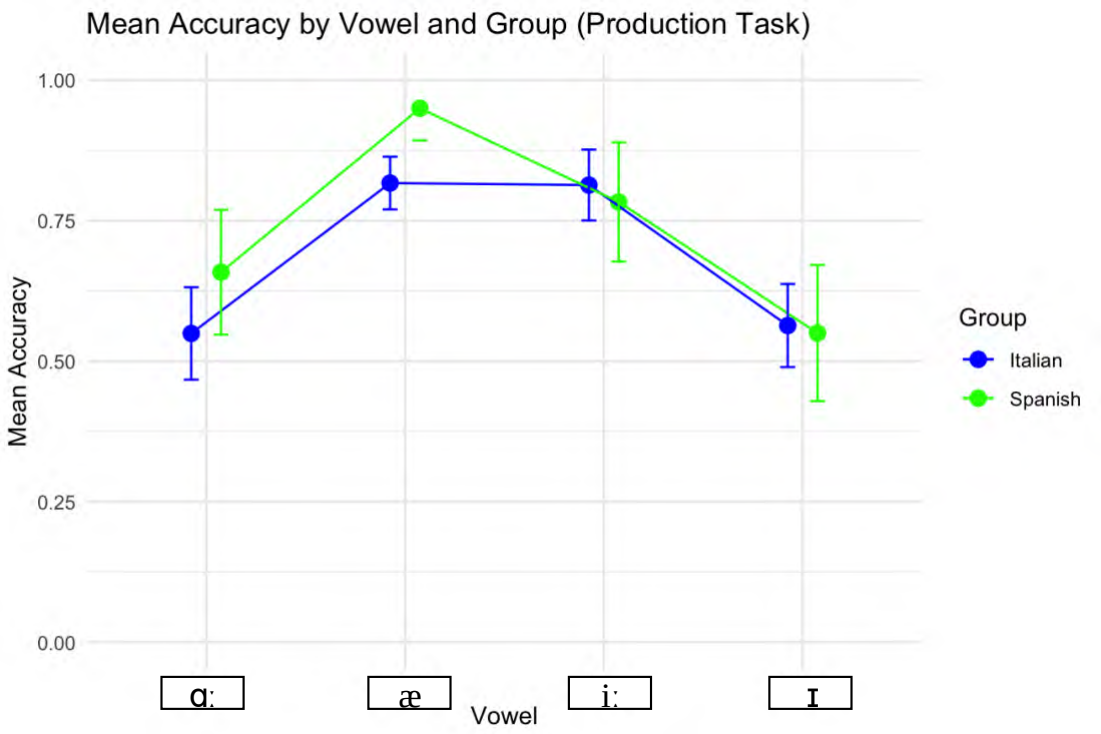
Appendix T: Point-and-Line Plot of Mean Perception Accuracy by Vowel and Group



Appendix U: Point-and-Line Plot of Mean Production Accuracy by Vowel (Whole Sample)



Appendix V: Point-and-Line Plot of Mean Production Accuracy by Vowel and Group

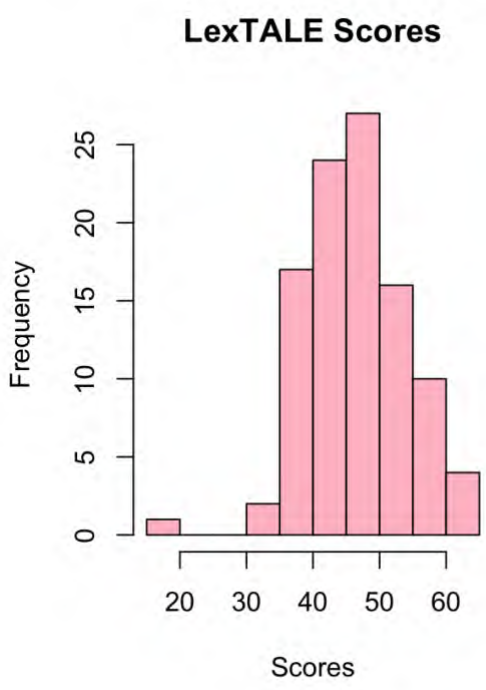


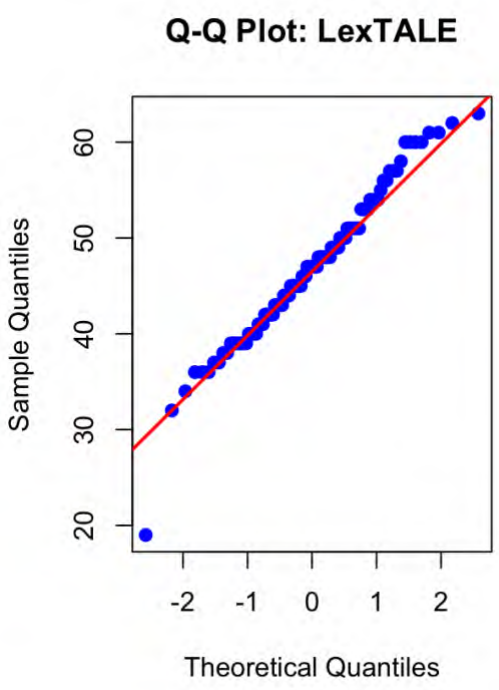
Appendix W: Normality checks for LexTALE and vowel accuracy scores in the Perception and Production Tasks

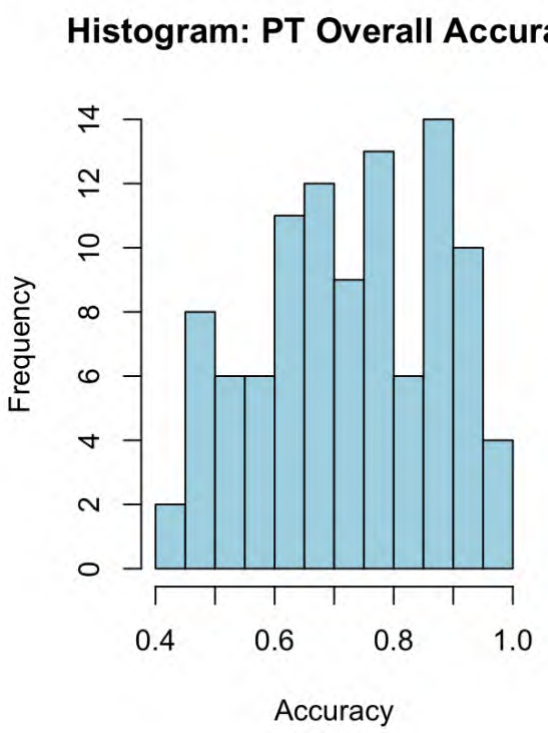
To determine whether the variables used in the correlation and regression analyses met the assumption of normality, both visual and statistical methods were employed. For each variable (LexTALE scores, overall accuracy scores, and individual vowel scores), histograms and Q–Q plots were generated and examined. In addition, Shapiro–Wilk tests were conducted to statistically assess normality.

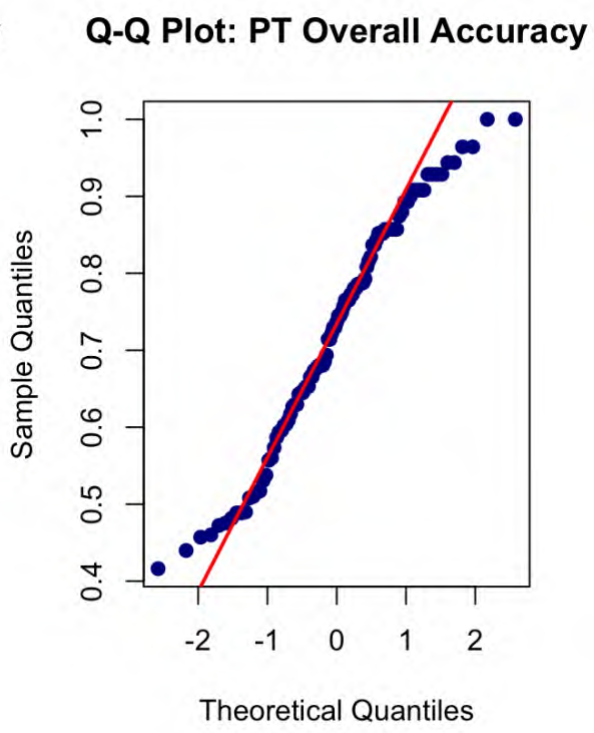
All histograms and Q–Q plots are provided in this appendix, followed by the full table of Shapiro–Wilk test results. A significance threshold of $p < .05$ was used to determine whether a variable deviated significantly from normality.

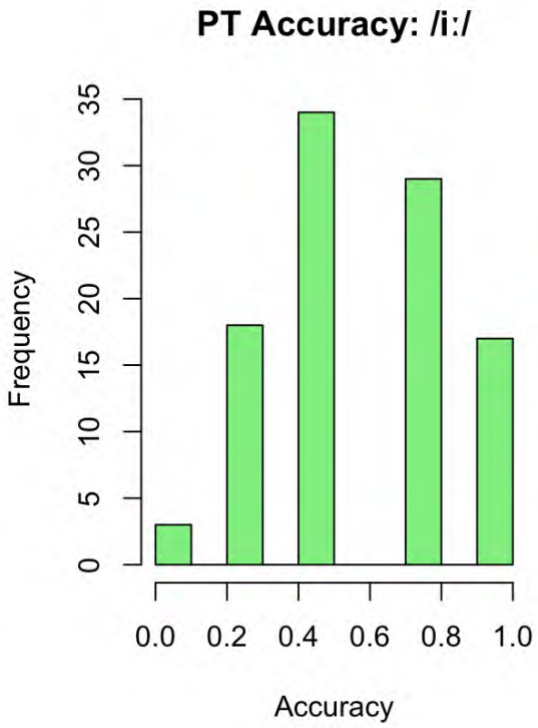
W1. Visual Checks: Histograms and Q-Q Plots

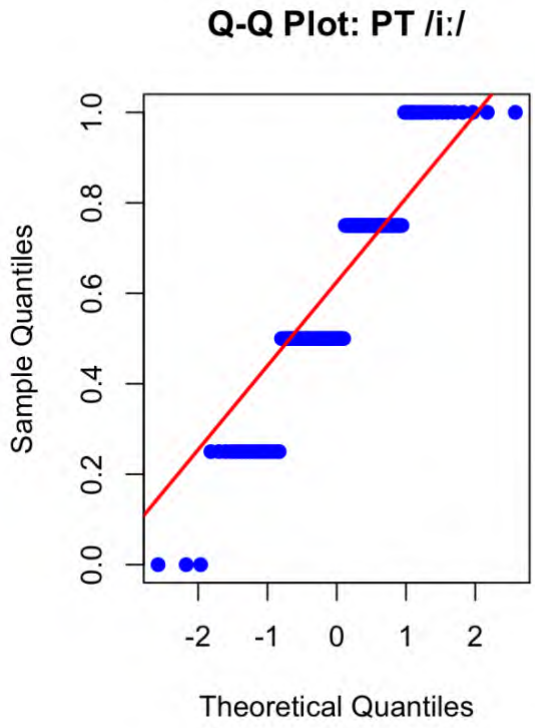


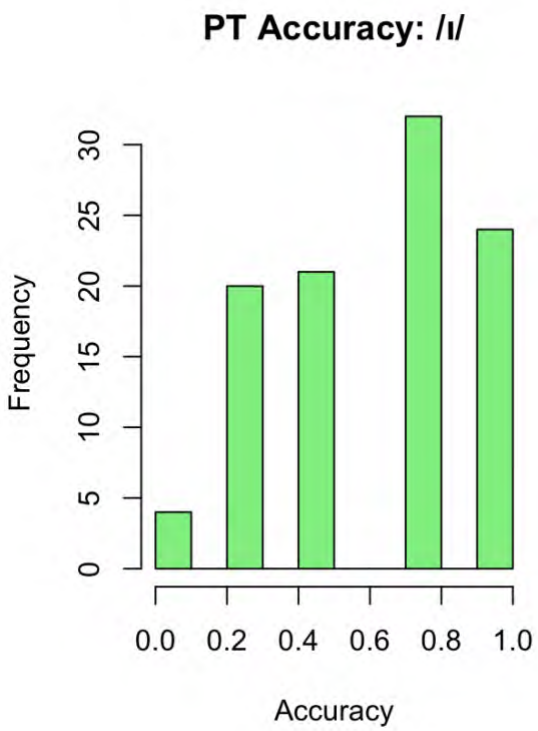


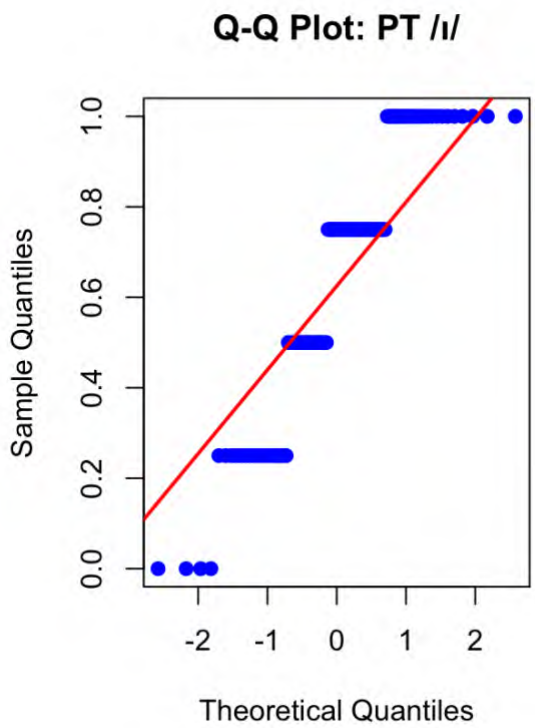

Histogram: PT Overall Accuracy
Frequency
0
2
4
6
8
10
12
14
0.4
0.6
0.8
1.0
Accuracy

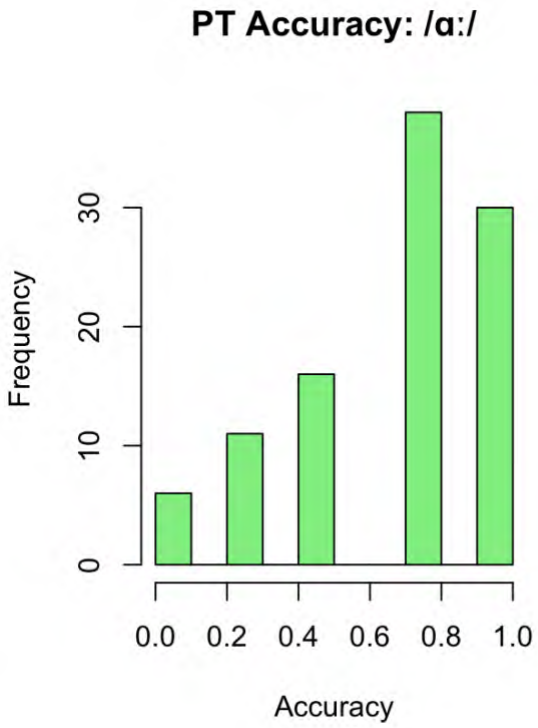

Q-Q Plot: PT Overall Accuracy
Sample Quantiles
0.4
0.5
0.6
0.7
0.8
0.9
1.0
-2
-1
0
1
2
Theoretical Quantiles

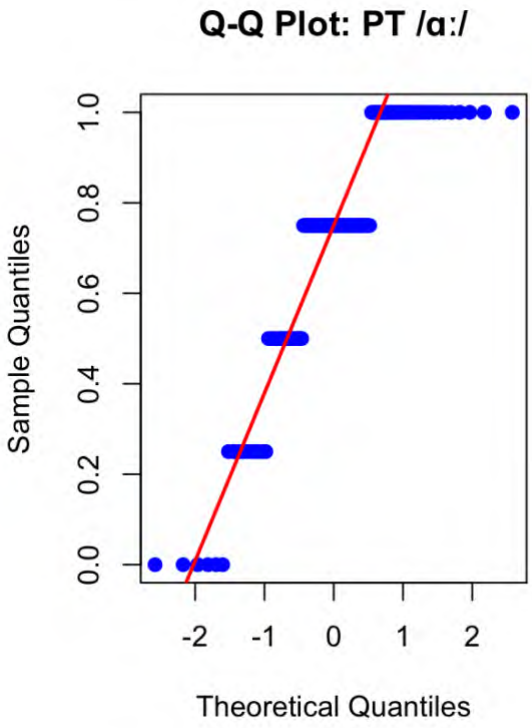

PT Accuracy: /i:/
Frequency
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

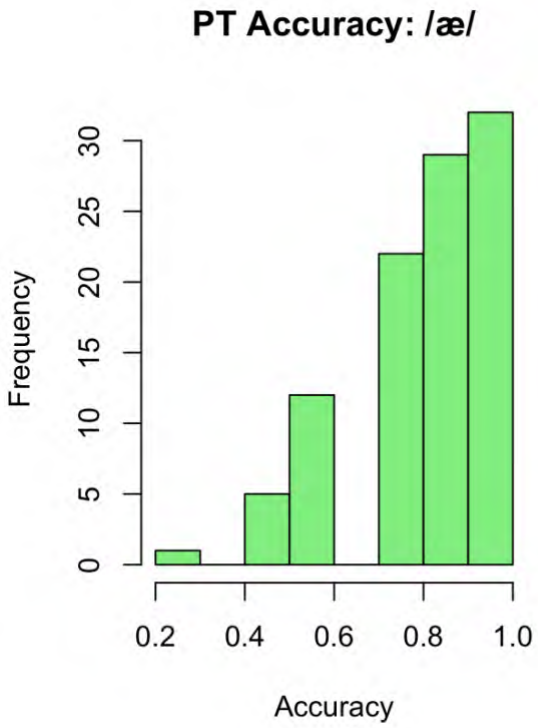

Q-Q Plot: PT /i:/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

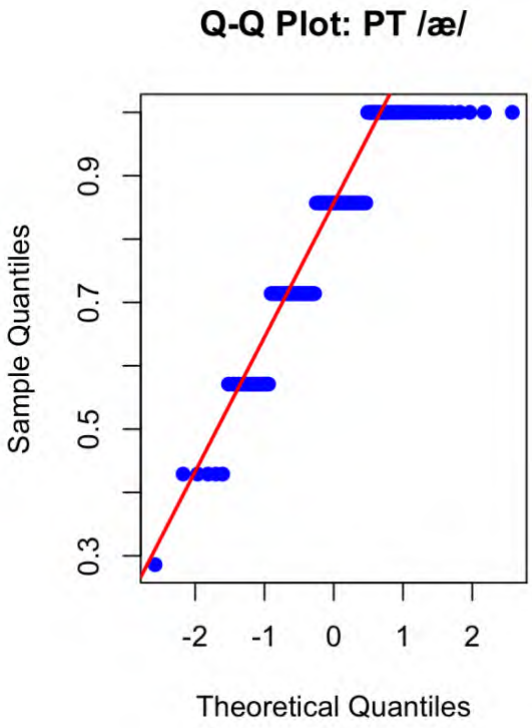

PT Accuracy: /ɪ/
Frequency
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

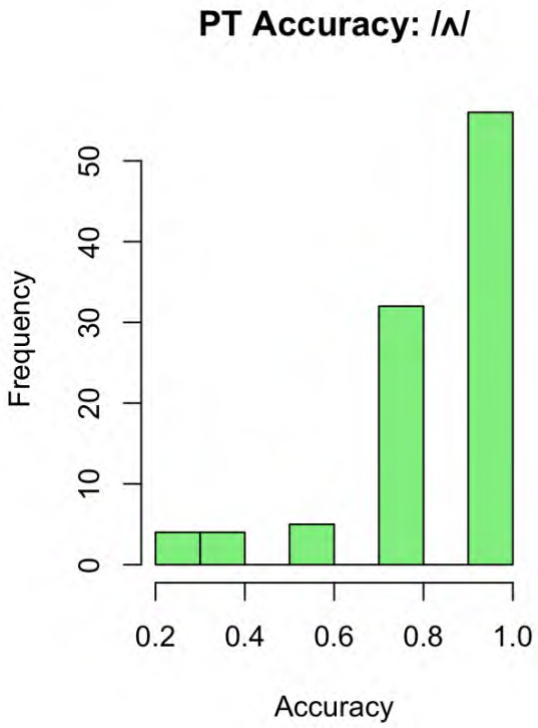

Q-Q Plot: PT /ɪ/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

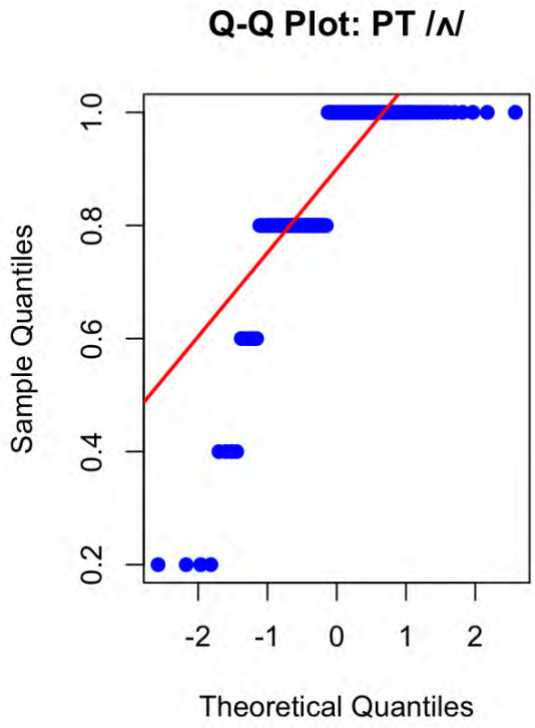

PT Accuracy: /ɑː/
Frequency
0
10
20
30
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

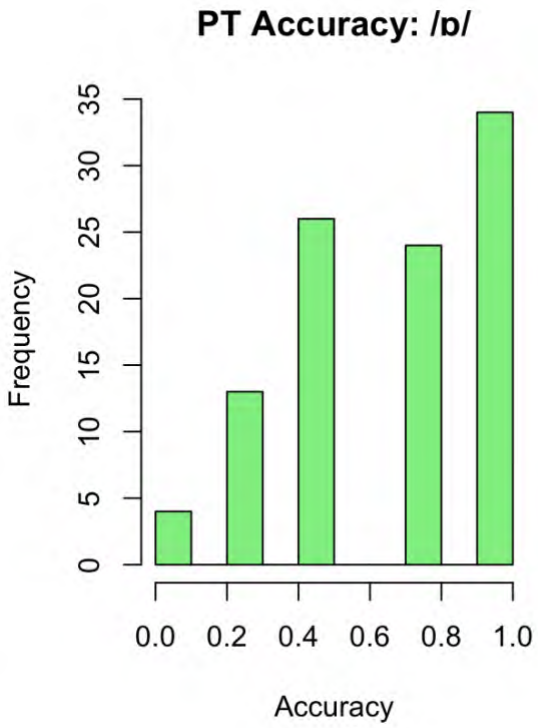

Q-Q Plot: PT /ɑː/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

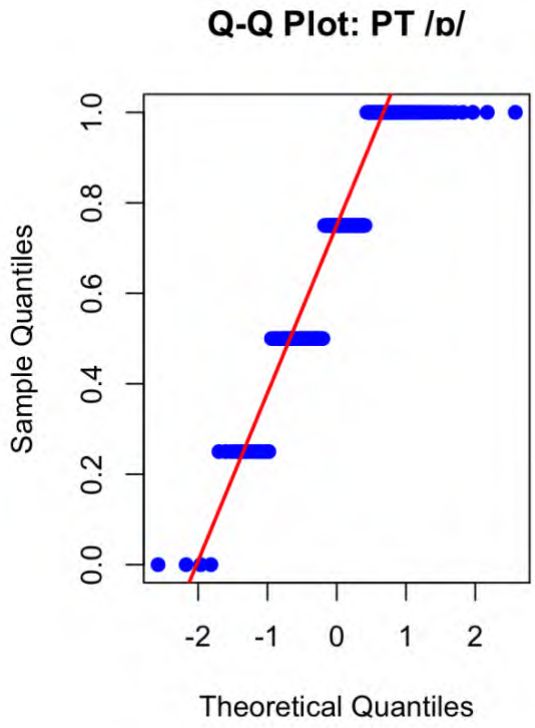

PT Accuracy: /æ/
Frequency
0
5
10
15
20
25
30
0.2
0.4
0.6
0.8
1.0
Accuracy

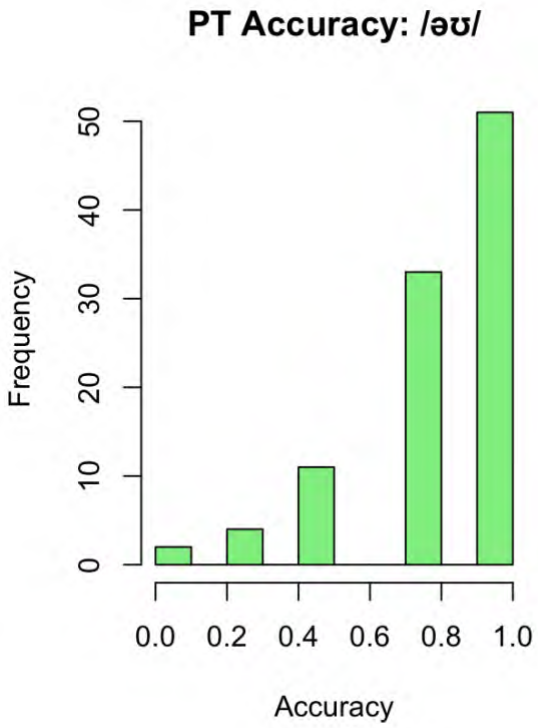

Q-Q Plot: PT /æ/
Sample Quantiles
0.3
0.5
0.7
0.9
-2
-1
0
1
2
Theoretical Quantiles

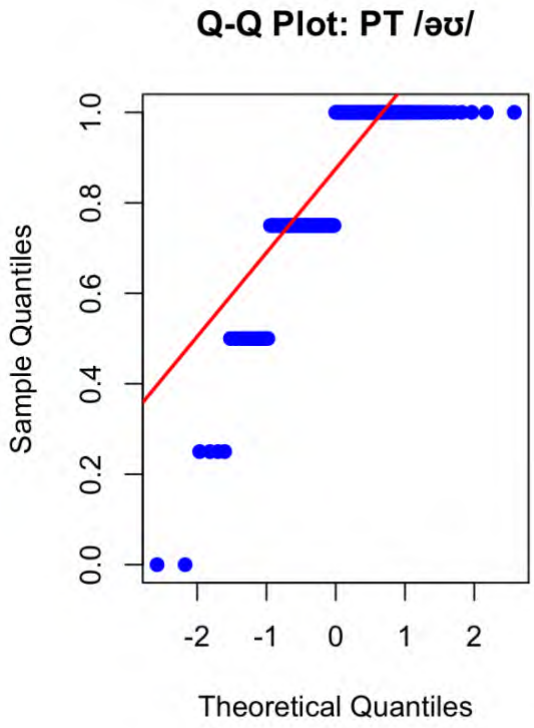

PT Accuracy: /ʌ/
Frequency
0
10
20
30
40
50
0.2
0.4
0.6
0.8
1.0
Accuracy

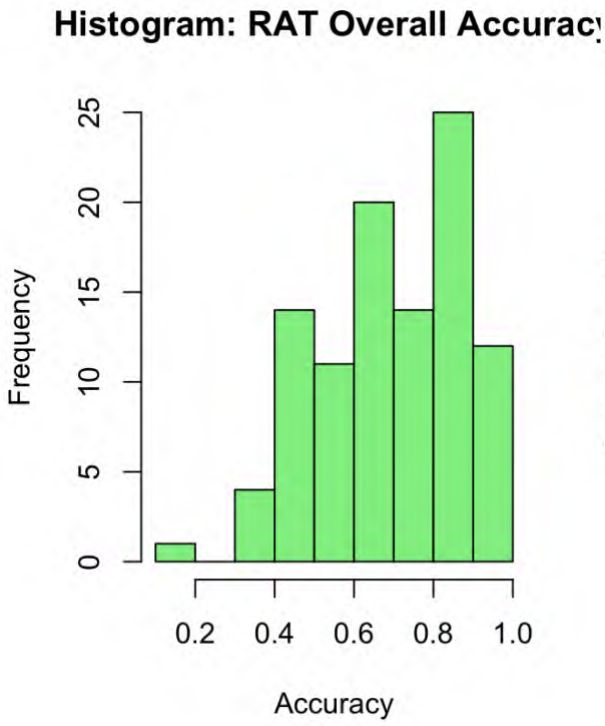

Q-Q Plot: PT /ʌ/
Sample Quantiles
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

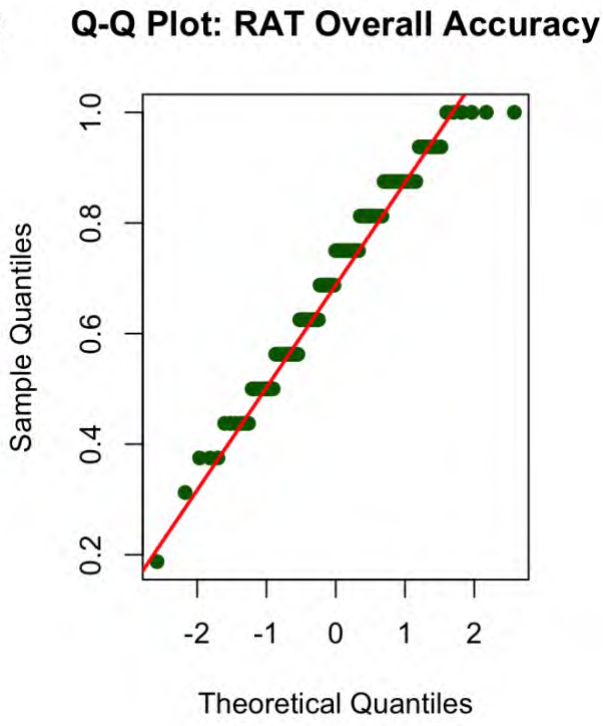

PT Accuracy: /ɒ/
Frequency
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

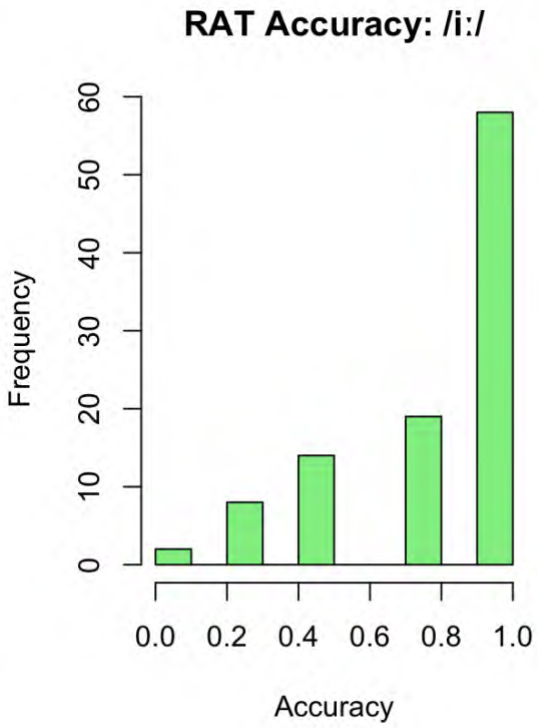

Q-Q Plot: PT /ɒ/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

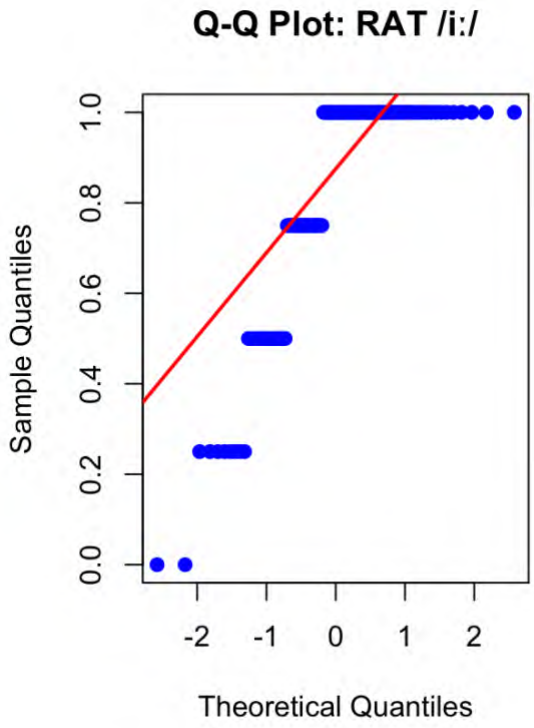

PT Accuracy: /əʊ/
Frequency
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

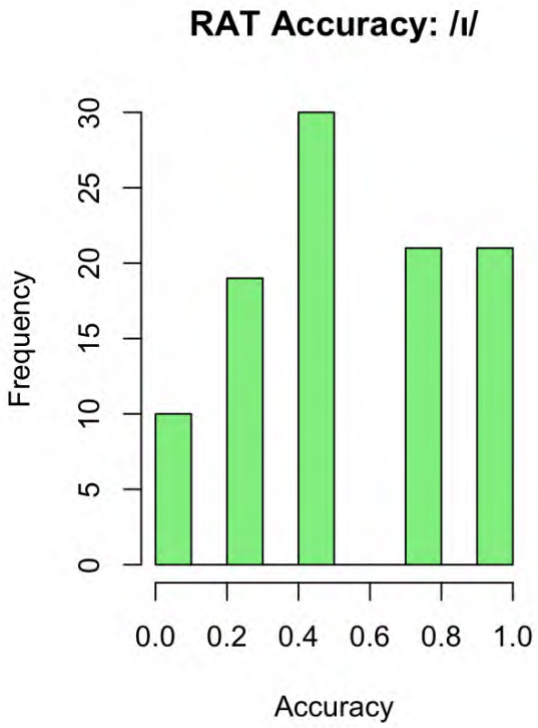

Q-Q Plot: PT /əʊ/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

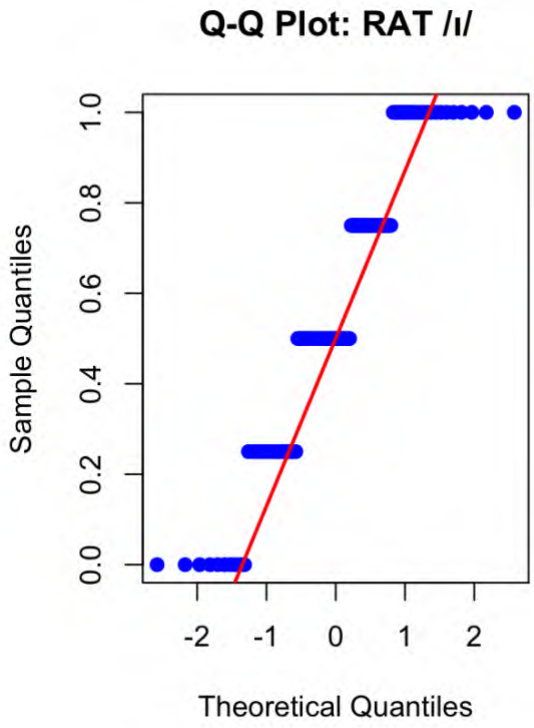

Histogram: RAT Overall Accuracy
Frequency
0
5
10
15
20
25
0.2
0.4
0.6
0.8
1.0
Accuracy

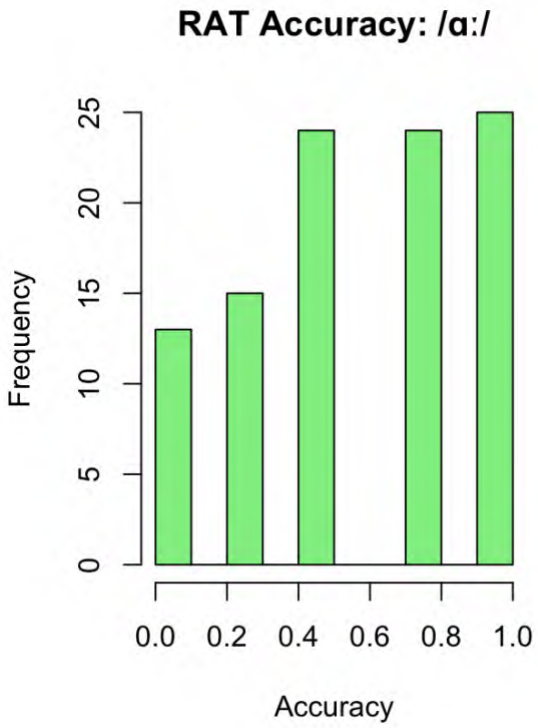

Q-Q Plot: RAT Overall Accuracy
Sample Quantiles
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

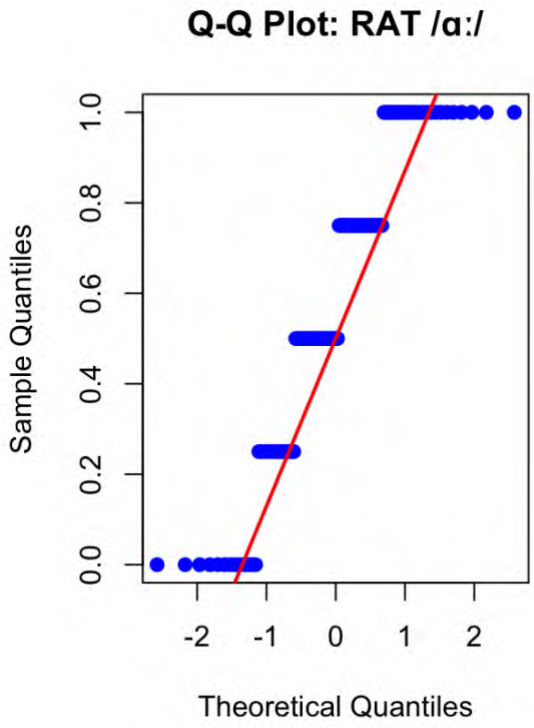

RAT Accuracy: /i:/
Frequency
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy

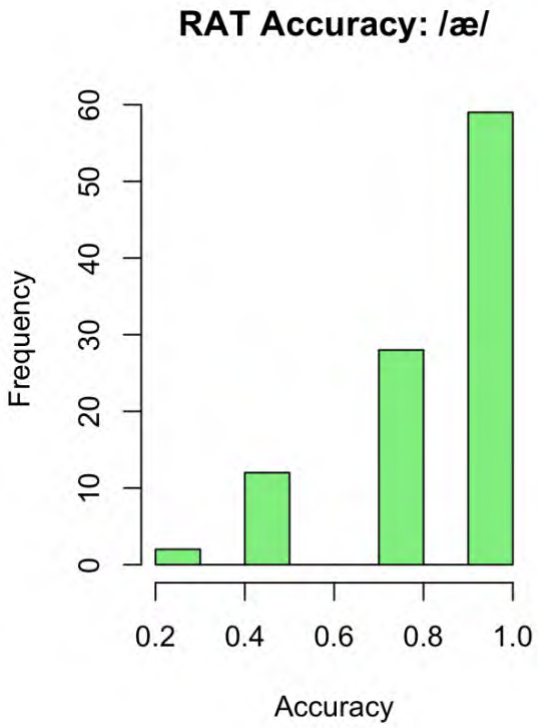

Q-Q Plot: RAT /i:/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

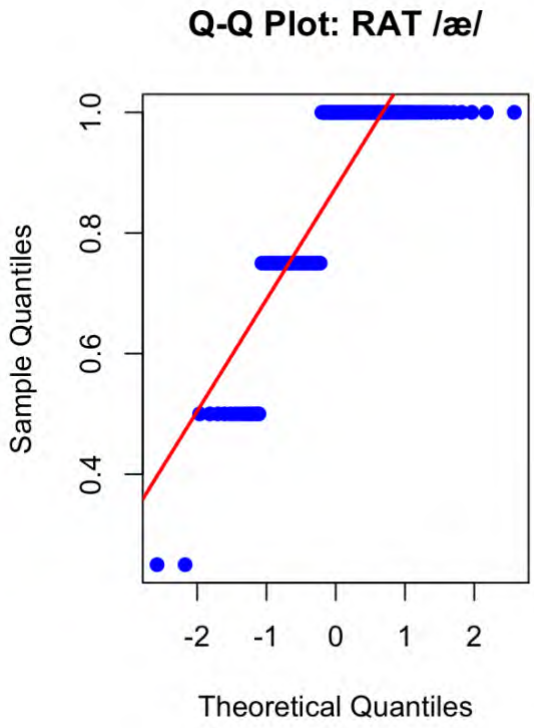

RAT Accuracy: /ɪ/
Frequency
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy


Q-Q Plot: RAT /ɪ/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles


RAT Accuracy: /ɑ:/
Frequency
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy


Q-Q Plot: RAT /ɑ:/
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles


RAT Accuracy: /æ/
Frequency
0
10
20
30
40
50
60
0.2
0.4
0.6
0.8
1.0
Accuracy


Q-Q Plot: RAT /æ/
Sample Quantiles
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theoretical Quantiles

W2. Shapiro-Wilk Test: LexTALE Scores

Variable	W	*p*-value	Interpretation
LexTALE	0.97702	0.07491	Normally distributed

W3. Shapiro-Wilk Test: Overall Accuracy Scores

Variable	W	*p*-value	Interpretation
Picture-Selection Task (SPT)	0.97018	0.02176	Not normally distributed
Read-Aloud Task (RAT)	0.96977	0.02026	Not normally distributed

---

W4. Shapiro-Wilk Test: Perception Task – Individual Vowels

Vowel	W	*p*-value	Interpretation
/iː/	0.90617	< .00001	Not normally distributed
/ɪ/	0.88926	< .00001	Not normally distributed
/ɑː/	0.85236	< .00001	Not normally distributed
/æ/	0.87371	< .00001	Not normally distributed
/ʌ/	0.68654	< .00001	Not normally distributed
/ɒ/	0.86922	< .00001	Not normally distributed
/əʊ/	0.75535	< .00001	Not normally distributed

W5. Shapiro-Wilk Test: Production Task – Individual Vowels

Vowel	**W**	***p*-value**	**Interpretation**
/iː/	0.73564	< .00001	Not normally distributed
/ɪ/	0.90554	0.000002353	Not normally distributed
/ɑː/	0.88923	0.0000004089	Not normally distributed
/æ/	0.72410	< .00001	Not normally distributed

Appendix X: Correlation Analyses Scatterplots for Perception and Production Tasks

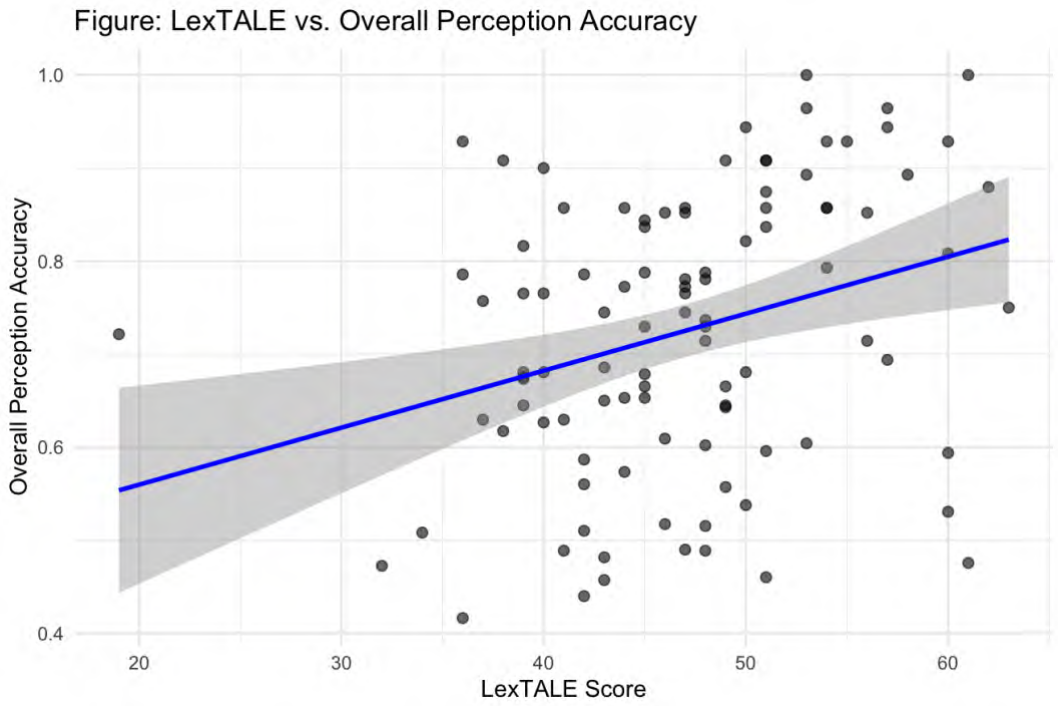


Figure: LexTALE vs. /iː/ Perception Accuracy

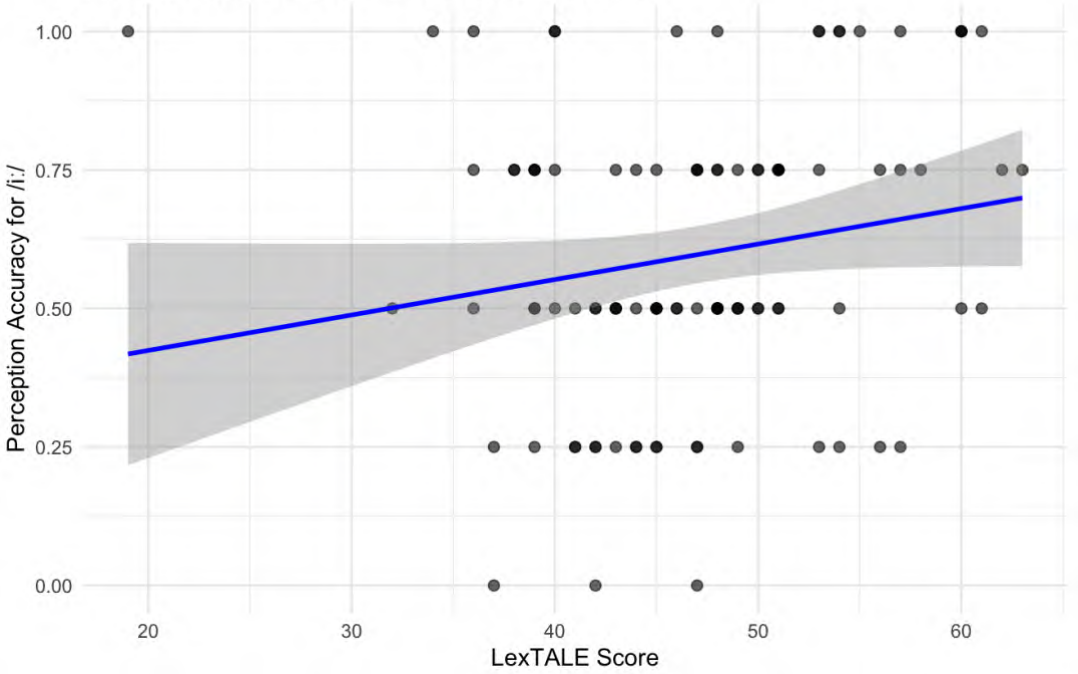


Figure: LexTALE vs. /ɪ/ Perception Accuracy

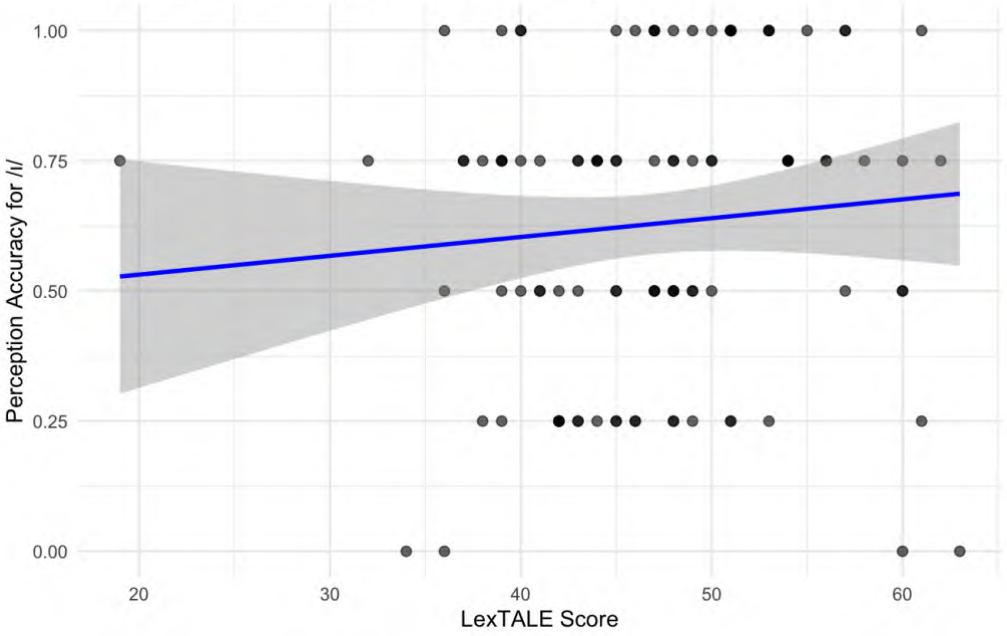


Figure: LexTALE vs. /ɑ:/ Perception Accuracy

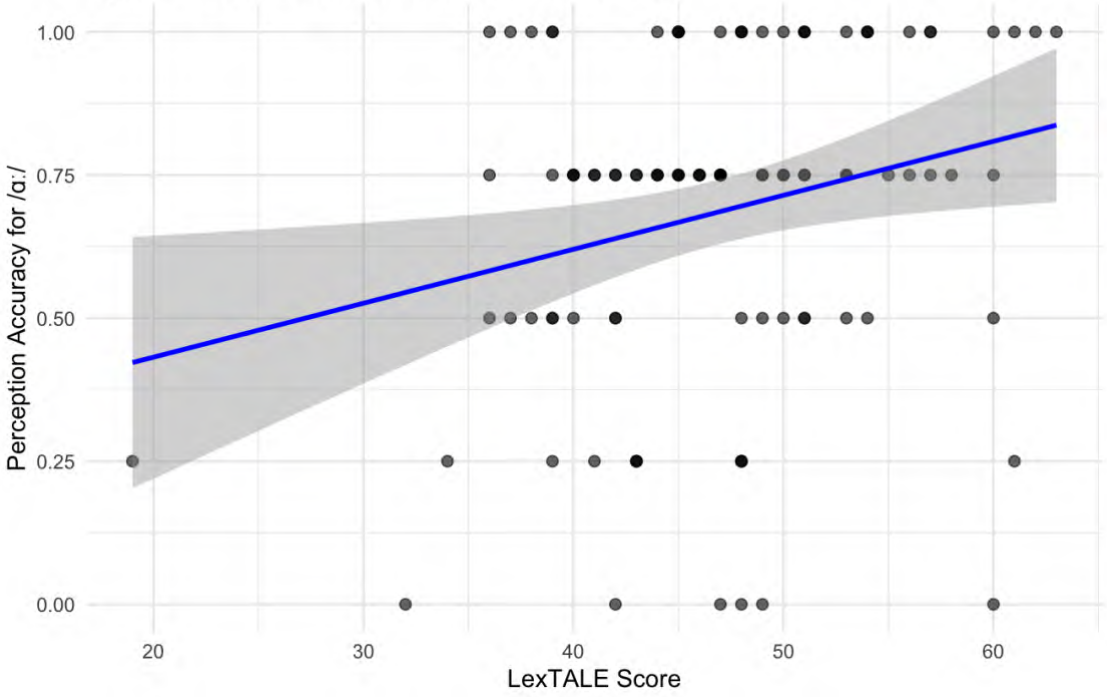


Figure: LexTALE vs. /æ/ Perception Accuracy

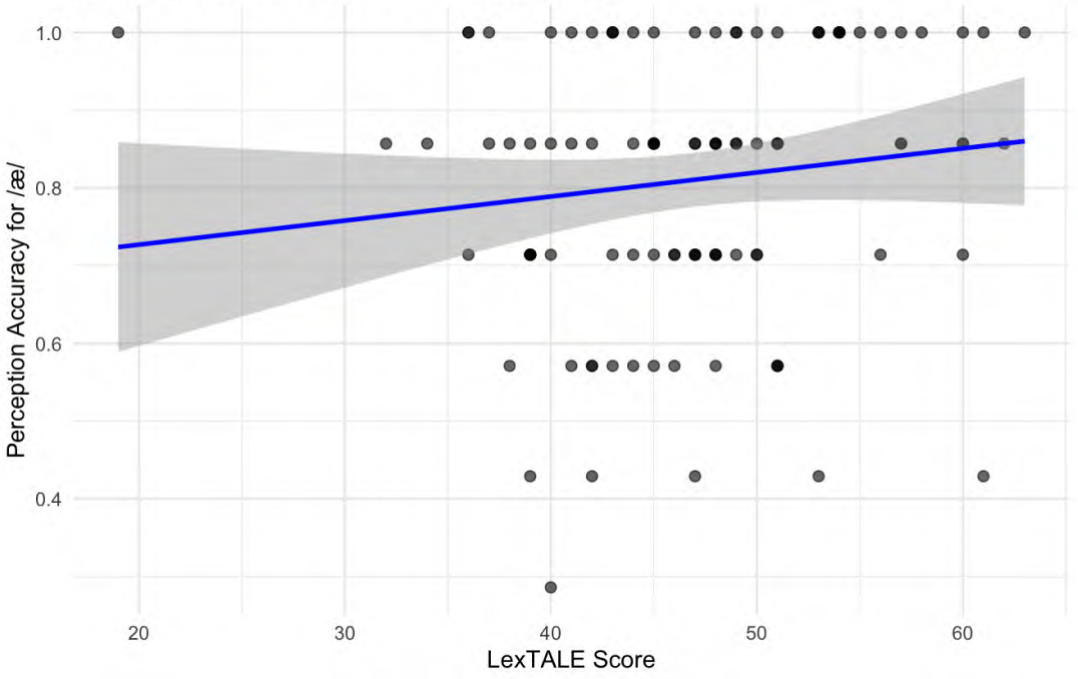


Figure: LexTALE vs. /ʌ/ Perception Accuracy

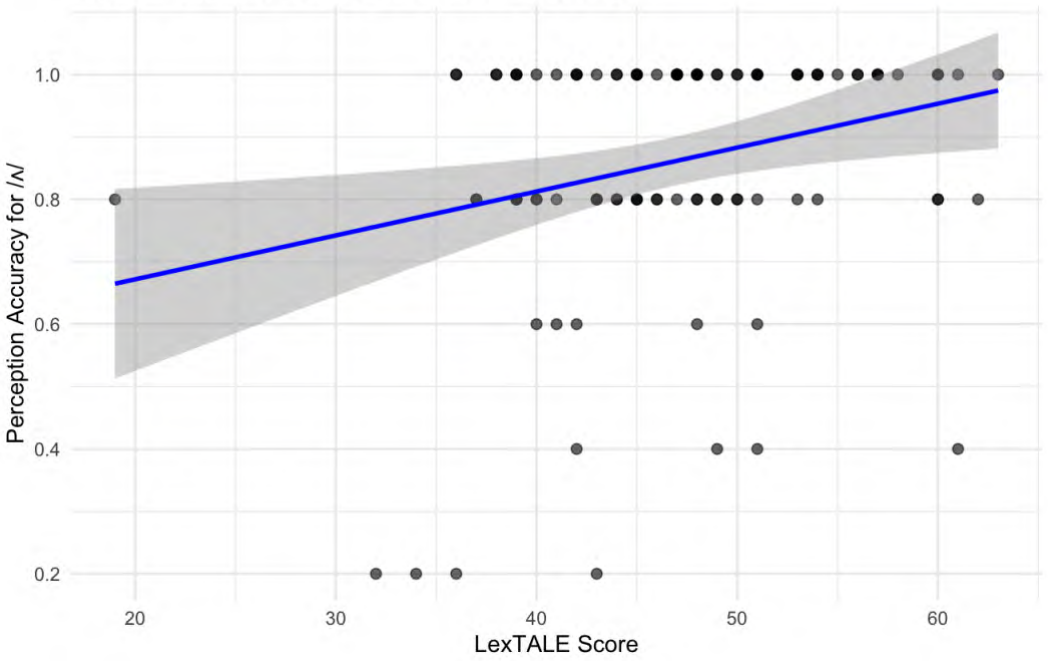


Figure: LexTALE vs. /ɒ/ Perception Accuracy

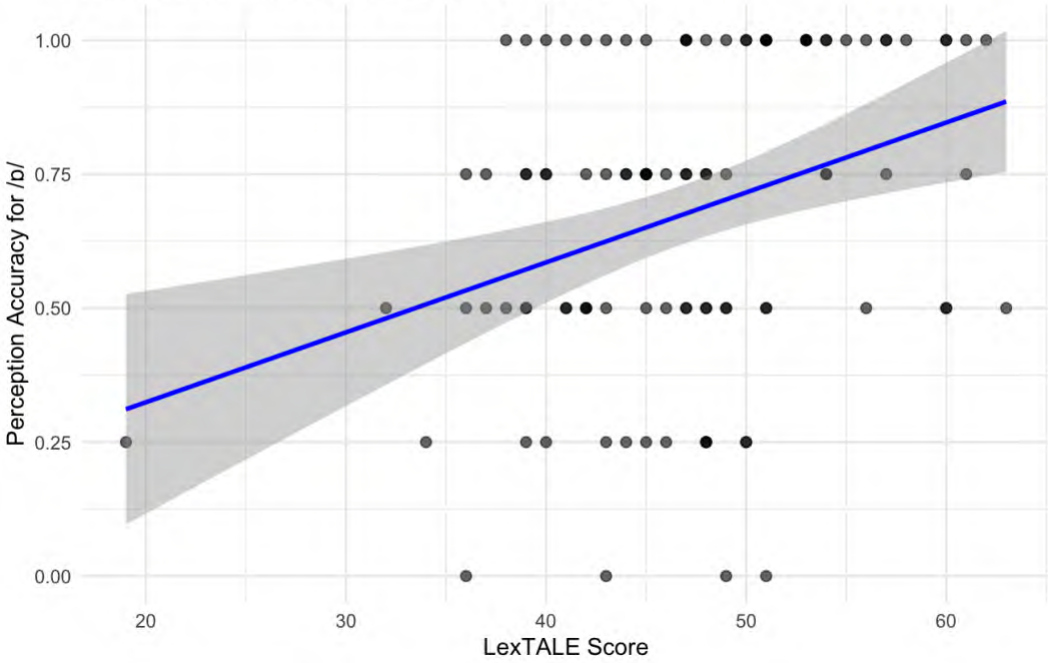


Figure: LexTALE vs. /əʊ/ Perception Accuracy

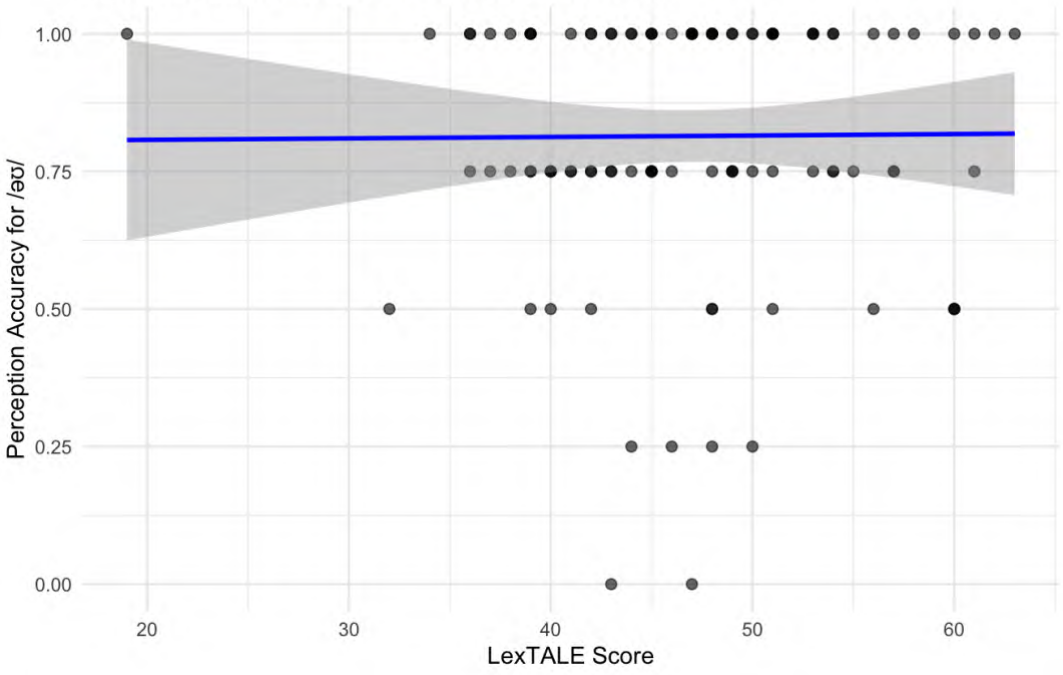


Figure: LexTALE vs. Overall Production Accuracy

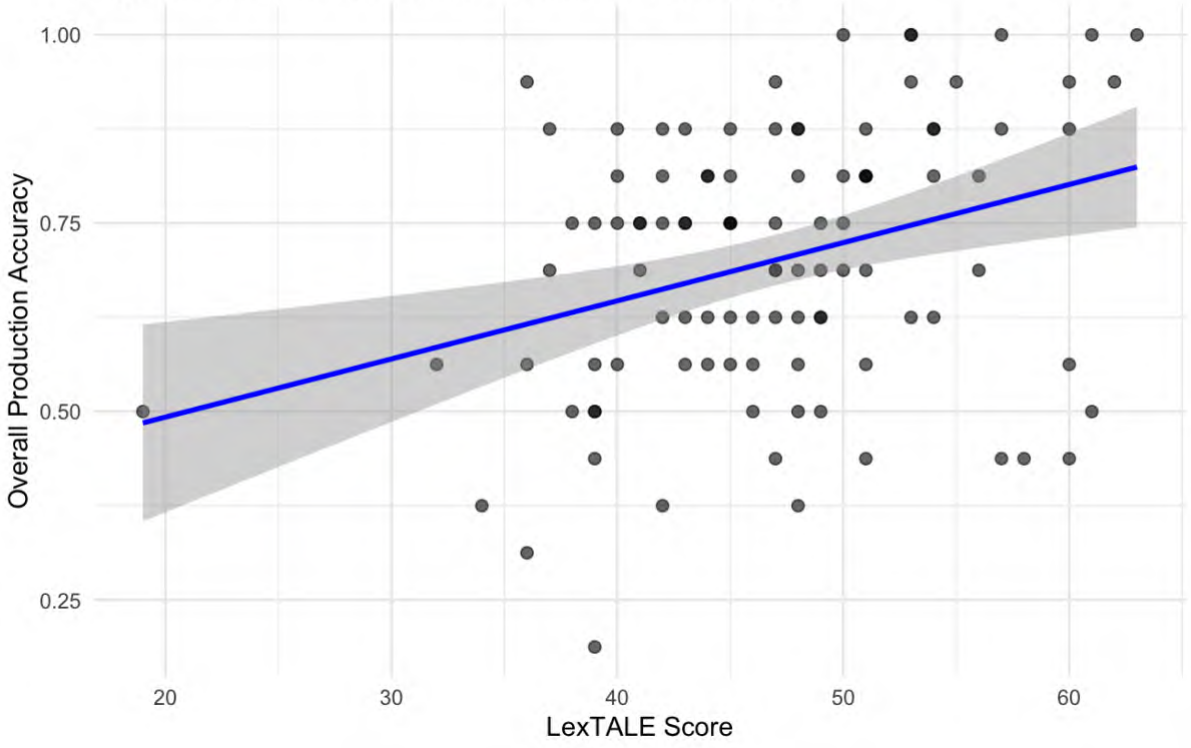


Figure: LexTALE vs. /iː/ Production Accuracy

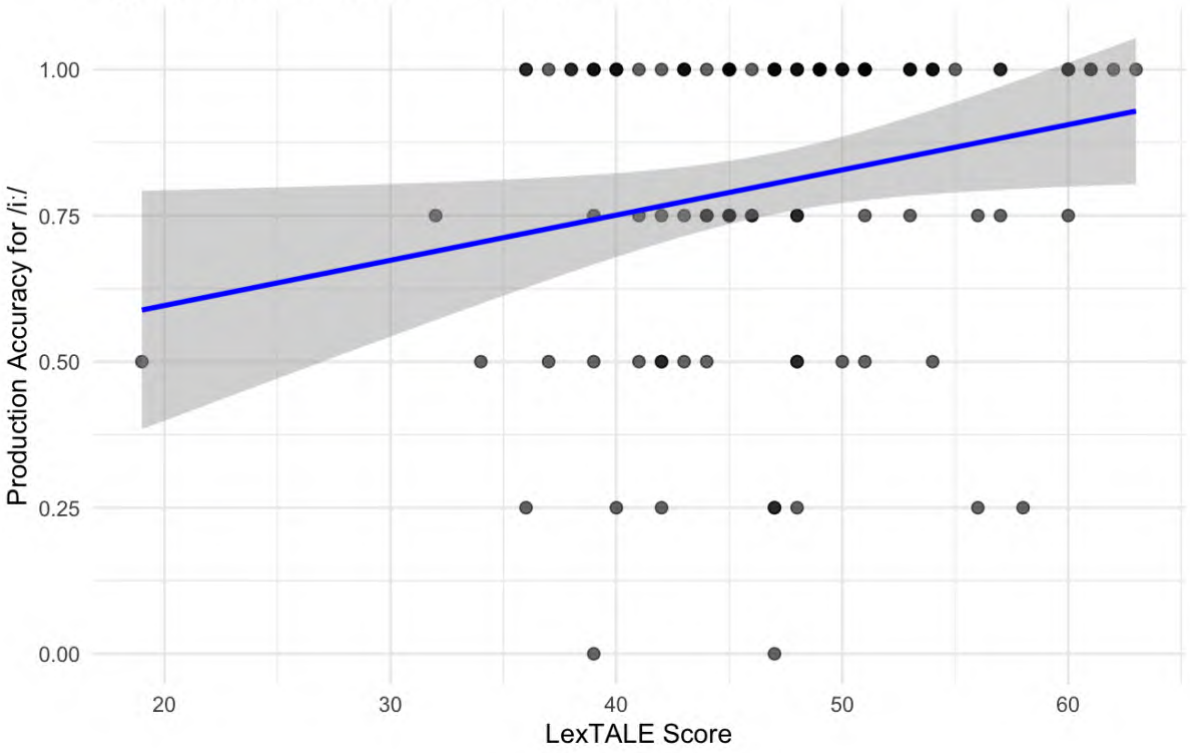


Figure: LexTALE vs. /ɪ/ Production Accuracy

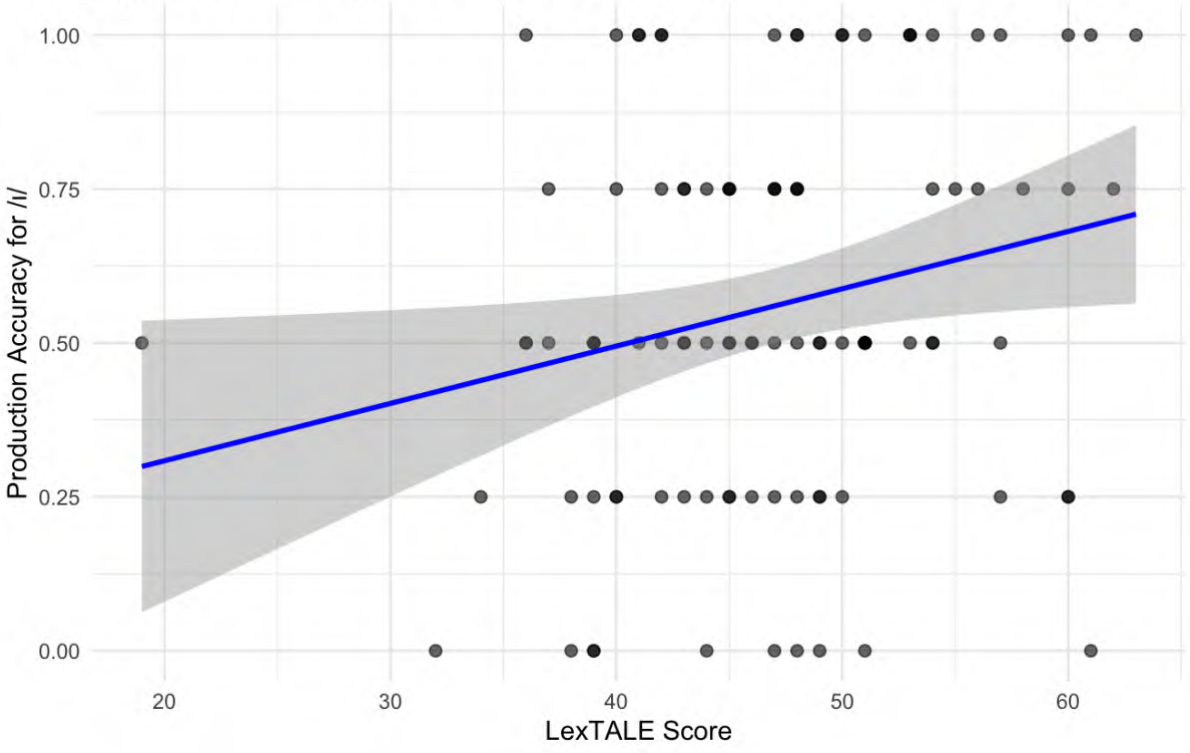


Figure: LexTALE vs. /ɑː/ Production Accuracy

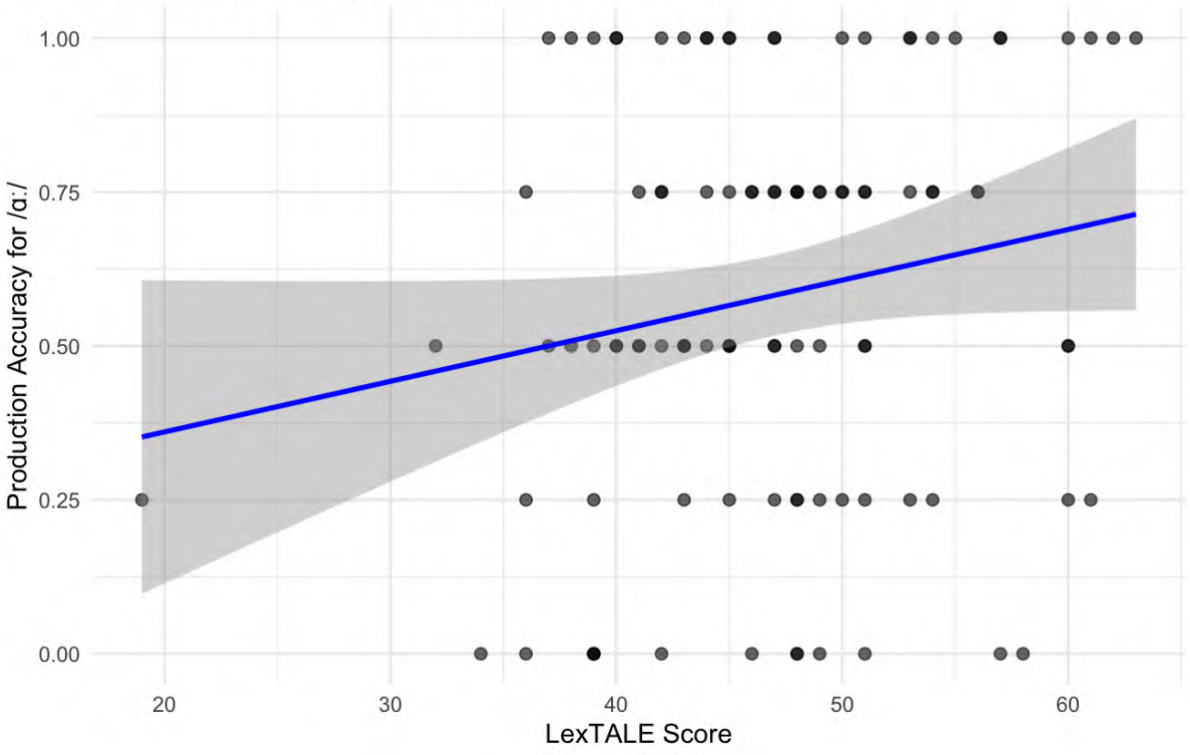
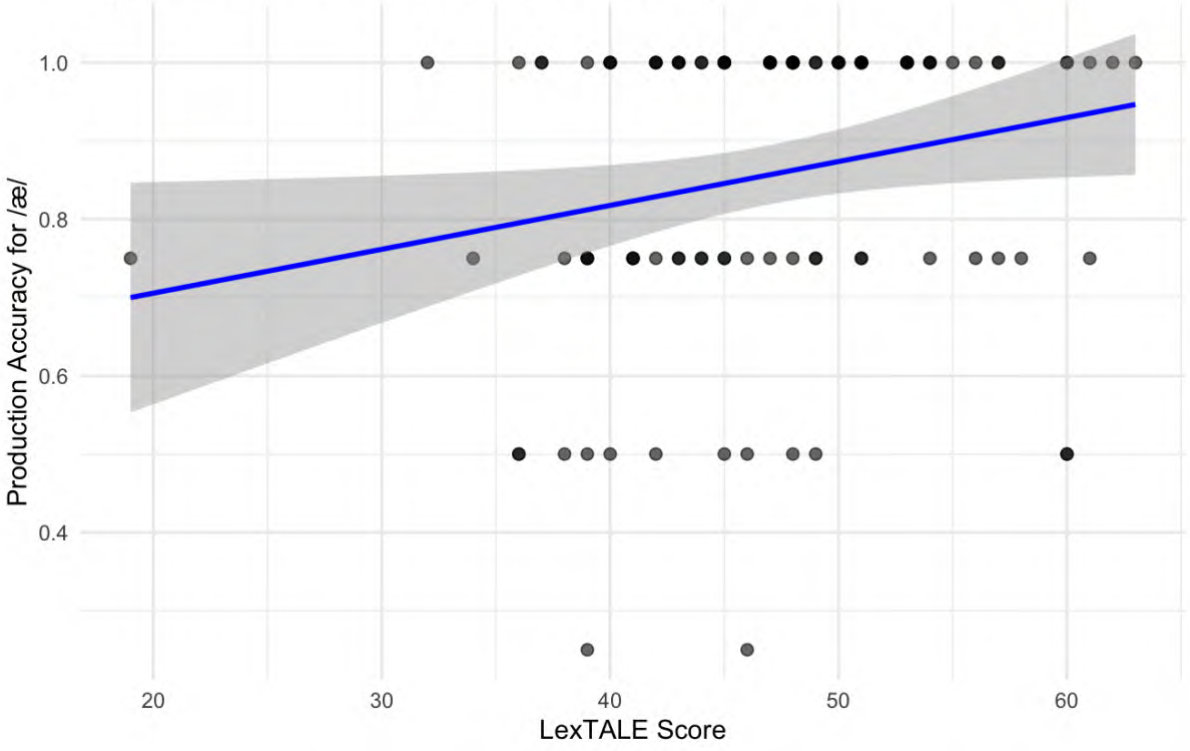
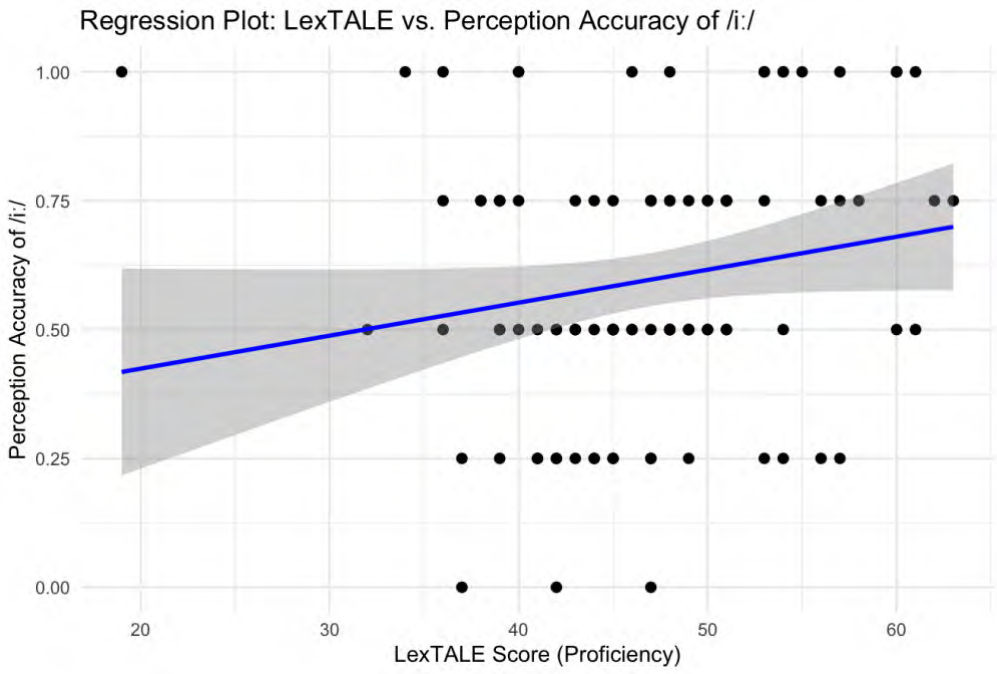


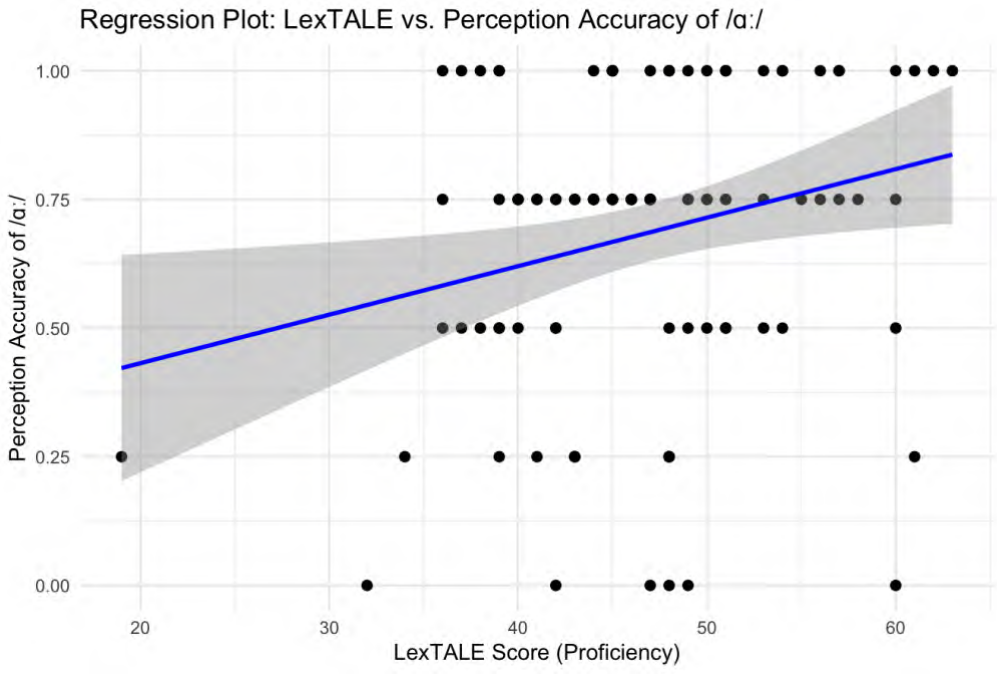
Figure: LexTALE vs. /æ/ Production Accuracy

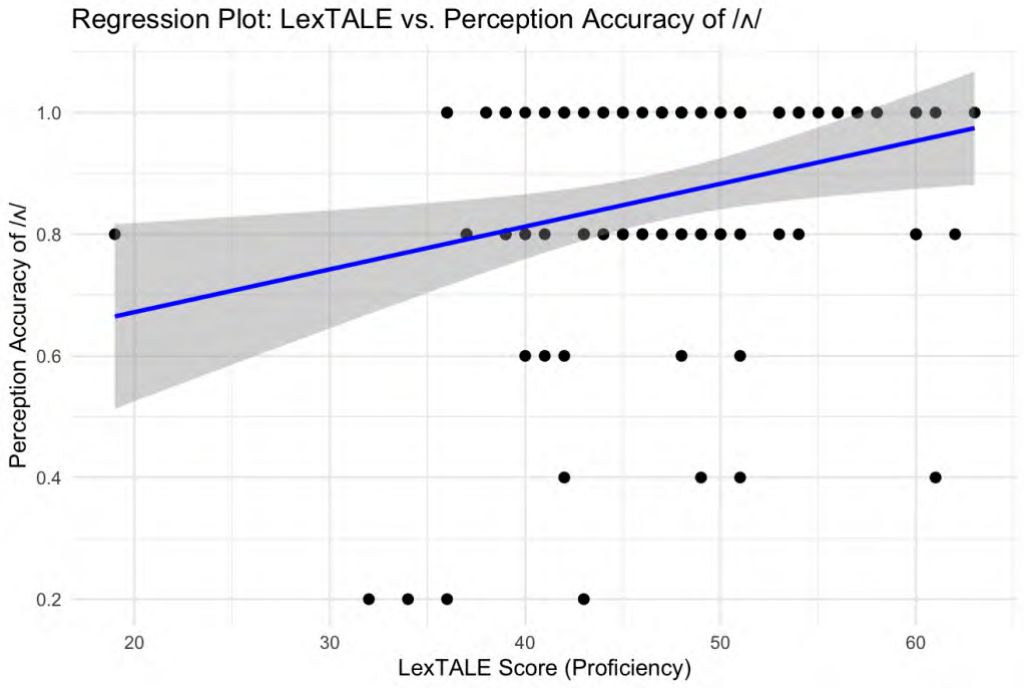


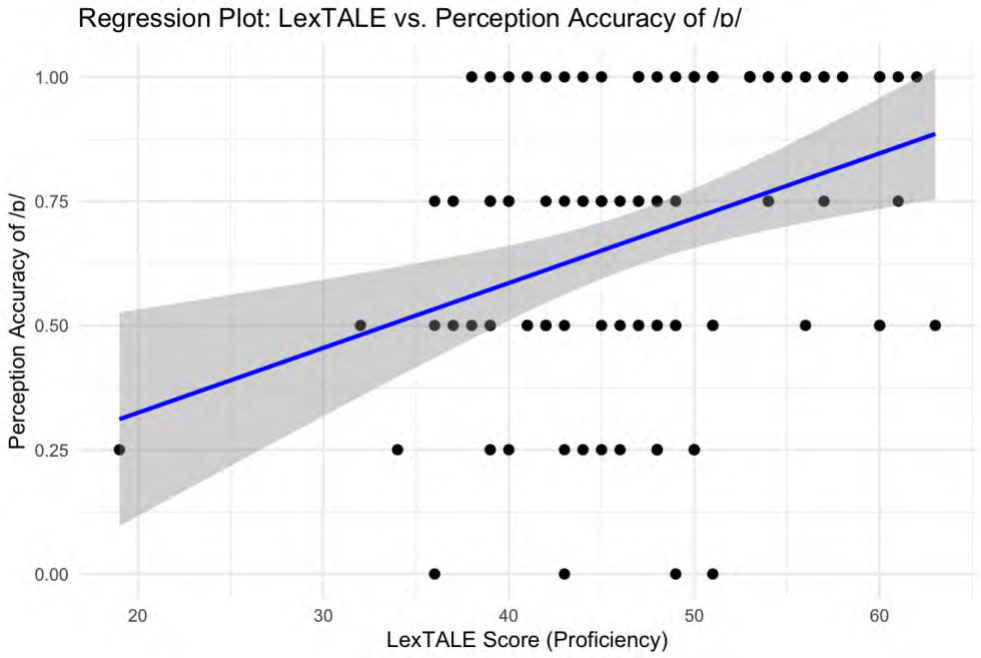
Appendix Y: Regression Scatterplots for Perception and Production Tasks

Y1. Perception Task

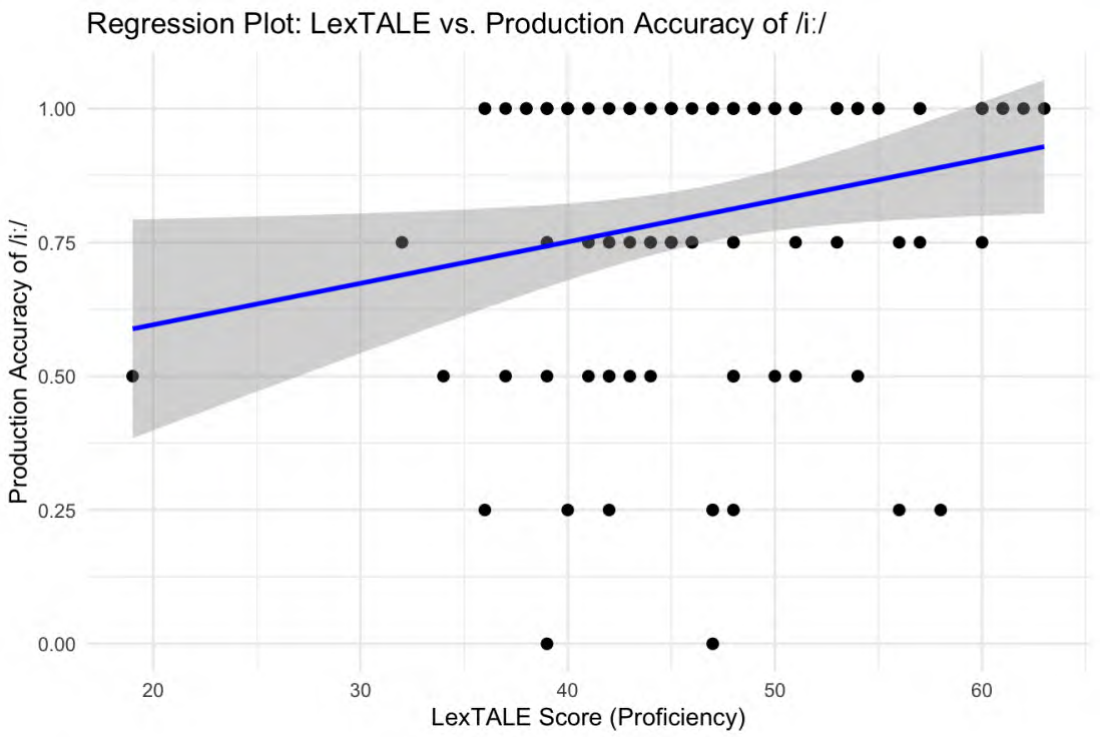


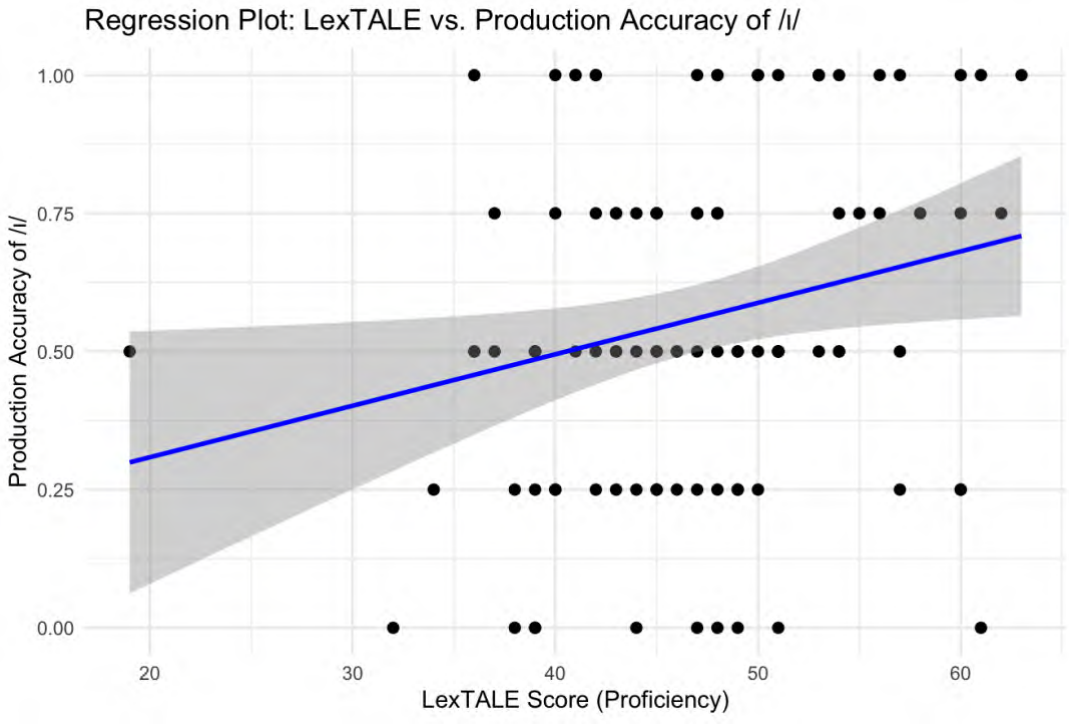


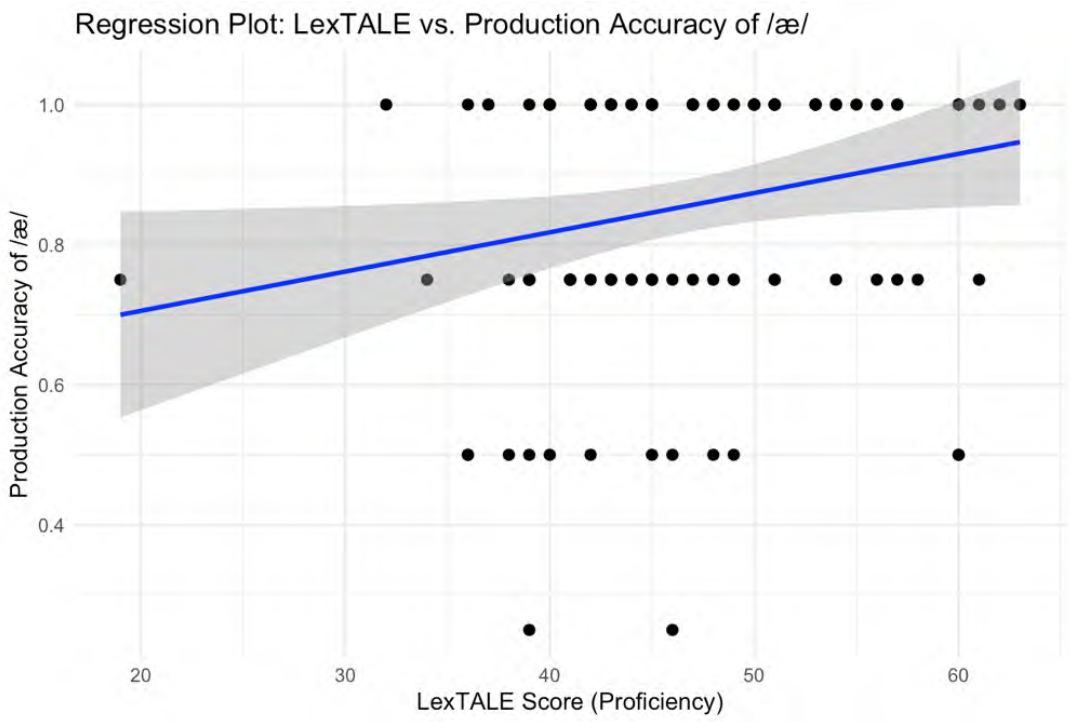

Regression Plot: LexTALE vs. Perception Accuracy of /ʌ/
Perception Accuracy of /ʌ/
1.0
0.8
0.6
0.4
0.2
20
30
40
50
60
LexTALE Score (Proficiency)


Regression Plot: LexTALE vs. Perception Accuracy of /ɒ/
Perception Accuracy of /ɒ/
1.00
0.75
0.50
0.25
0.00
20
30
40
50
60
LexTALE Score (Proficiency)

Y2. Production Task






Regression Plot: LexTALE vs. Production Accuracy of /æ/
Production Accuracy of /æ/
1.0
0.8
0.6
0.4
20
30
40
50
60
LexTALE Score (Proficiency)

Appendix Z: Q-Q Plot Assessing the Normality of the Difference Scores Between Perception and Production Accuracy

